aws

开发人员指南

Amazon Kendra

Amazon Kendra: 开发人员指南

Table of Contents

什么是 Amazon Kendra？

Amazon Kendra 是一项智能搜索服务，它使用自然语言处理和高级机器学习算法，从您的数据中返回搜索问题的特定答案。

与传统的基于关键字的搜索不同，Amazon Kendra 使用语义和上下文理解功能来决定文档是否与搜索查询相关。它会返回问题的具体答案，为用户提供接近与人类专家互动的体验。

> **Note**
>
> 您还可以使用 Amazon Kendra 语义搜索功能来对其他搜索服务的结果进行重新排名。有关更多详细信息，请参阅 Amazon Kendra 智能排名。

借助 Amazon Kendra，您可以通过将多个数据存储库连接到索引，以及提取和爬取文档来创建统一的搜索体验。您可以使用文档元数据为用户创建功能丰富的自定义搜索体验，帮助他们高效地找到正确的查询答案。

什么是 Amazon Kendra？

查询 Amazon Kendra

您可以向 Amazon Kendra 提出以下类型的查询：

事实类问题 - 有关是谁、是什么、在何时或何地的简单问题，例如离西雅图最近的服务中心在哪里？事实类问题具有基于事实的答案，可使用单个单词或短语返回答案。答案来自常见问题解答或您的索引文档。

描述性问题 - 答案可能是句子、段落或整个文档的问题。例如，如何将 Echo Plus 连接到我的网络？或者，低收入家庭如何获得税收优惠？

关键字和自然语言问题 - 包含复杂对话内容的问题，其含义可能不明确。例如，主题演讲。当 Amazon Kendra 遇到像“address”这种具有多种上下文含义的单词时，它会正确推断出搜索查询背后的含义并返回相关信息。

Amazon Kendra 的优势

Amazon Kendra 高度可扩展，能够满足性能需求，与 Amazon S3 和 Amazon Lex 等其他 AWS 服务紧密集成，并且具有企业级安全性。使用 Amazon Kendra 的一些好处包括：

简单性 - Amazon Kendra 提供了管理搜索文档的控制台和 API。您可以使用简单的搜索 API 来将 Amazon Kendra 集成到您的客户端应用程序中，例如，网站或移动应用程序。

连接性 - Amazon Kendra 可以连接到第三方数据存储库或数据来源，例如，Microsoft SharePoint。您可以使用数据来源轻松地为文档编制索引和进行搜索。

准确性 - 与使用关键字搜索的传统搜索服务不同，Amazon Kendra 会尝试了解问题的上下文并返回与您的查询最相关的单词、片段或文档。Amazon Kendra 使用机器学习来改善搜索结果。

安全性 - Amazon Kendra 提供高度安全的企业搜索体验。您的搜索结果反映组织的安全模型，可以根据用户或组对文档的访问权限进行筛选。客户负责对用户进行身份验证和授权。

Amazon Kendra 版本

Amazon Kendra 有两个版本：开发人员版本和企业版本。下表概述了这两个版本的功能以及相互之间的差异。

Amazon Kendra 开发人员版本	Amazon Kendra 企业版本
Amazon Kendra 开发人员版本以较低的成本提供 Amazon Kendra 的所有功能。 理想用例 • 探索 Amazon Kendra 如何为文档编制索引 • 试用功能 • 开发使用 Amazon Kendra 的应用程序 功能 • 包含 750 小时使用时间的免费套餐 • 最多编制 5 个索引，每个索引最多包含 5 个数据来源	Amazon Kendra 企业版本提供 Amazon Kendra 的所有功能，专为生产环境而设计。 理想用例 • 为整个企业文档库编制索引 • 在生产环境中部署应用程序 功能 • 最多编制 5 个索引，每个索引最多包含 50 个数据来源 • 100,000 个文档或 30 GB 的提取文本 • 每天大约 8,000 次查询或每秒 0.1 次查询

Amazon Kendra 开发人员版本	Amazon Kendra 企业版本
• 10,000 个文档或 3 GB 的提取文本 • 每天大约 4,000 次查询或每秒 0.05 次查询 • 在 1 个可用区 (AZ) 中运行 - 参阅可用区（AWS 区域中的数据中心） 限制 • 不适用于生产应用程序 • 无法保证延迟或可用性	• 在 3 个可用区 (AZ) 中运行 - 参阅可用区（AWS 区域中的数据中心） Note 您可以使用服务配额控制台来增加此配额。 限制 • 无

Note

有关 Amazon Kendra 支持的区域、端点和服务限额列表，请参阅 Amazon Kendra 的端点和限额。

Amazon Kendra 定价

您可以免费开始使用 Amazon Kendra 开发人员版本，该版本在前 30 天内提供长达 750 小时的使用时间。

试用期到期后，您需要为所有预配置的 Amazon Kendra 索引付费，即使这些索引为空且未运行任何查询。试用期结束后，使用 Amazon Kendra 数据来源扫描和同步文档需要支付额外费用。

有关费用和价格的完整列表，请参阅 Amazon Kendra 的价格。

您是 Amazon Kendra 新用户吗？

如果您是首次接触 Amazon Kendra 的用户，我们建议您按顺序阅读以下内容：

1	2	3	4	5	6
Amazon Kendra 的工作原理	开始使用	创建索引	通过批量上传将文档直接添加到索引中	创建数据来源连接器	搜索索引
介绍 Amazon Kendra 组件并描述如何使用它们来创建搜索解决方案。	介绍如何设置账户和测试 Amazon Kendra 搜索 API。	介绍如何使用 Amazon Kendra 来创建搜索索引和添加数据来源以同步文档。	介绍如何将文档直接添加到 Amazon Kendra 索引中。	介绍如何将数据存储库中的文档添加到 Amazon Kendra 索引中。	介绍如何使用 Amazon Kendra 搜索 API 来搜索索引。

Amazon Kendra 的工作原理

Amazon Kendra 为您的应用程序提供搜索功能。它可以直接为您的文档编制索引，也可以从第三方文档存储库编制索引，并智能地向用户提供相关信息。您可以使用 Amazon Kendra 为各种类型的文档创建可更新的索引。有关支持的文档类型的列表， Amazon Kendra 请参阅文档类型。

Amazon Kendra 与其他服务集成。例如，您可以为Amazon Lex 聊天机器人提供 Amazon Kendra 搜索功能，为用户的问题提供有用的答案。您可以使用Amazon Simple Storage Service 存储桶作为数据源， Amazon Kendra 以连接您的文档并为其编制索引。而且，您可以使用 AWS Identity and Access Management 设置资源的访问策略或权限。

Amazon Kendra 包含以下组件：

- 用于存储您的文档并使其可搜索的索引。
- 用于存储您的文档并将 Amazon Kendra 连接到的数据来源。您可以自动将数据源与 Amazon Kendra 索引同步，以便您的索引与源存储库保持同步。
- 一个将文档直接添加到索引的文档添加 API。

您可以 Amazon Kendra 通过控制台或 API 使用。您可以创建、更新和删除索引。删除索引会删除其所有数据源连接器，并从中永久删除您的所有文档信息 Amazon Kendra。

主题

索引

索引保存文档的内容，其结构使文档可搜索。向索引中添加文档的方式取决于您存储文档的方式。

- 如果您将文档存储在某种存储库中，例如存储 Amazon S3 桶或 Microsoft SharePoint 站点，则使用数据源连接器将存储库中的文档编入索引。
- 如果您不将文档存储在存储库中，则可以使用 BatchPutDocumentAPI 直接索引您的文档。

- 对于必须存储在 Amazon Kendra （ Amazon S3 ）存储桶中的常见问题和答案，您可以从存储桶上传

您可以使用 Amazon Kendra 控制台 AWS CLI、或 AWS SDK 创建索引。有关可以编制索引的文档类型的信息，请参阅文档类型。

使用 Amazon Kendra 保留或常用文档字段

借助 UpdateIndex API，您可以使用`DocumentMetadataConfigurationUpdates`并指定要映射到等效文档属性/字段名称的 Amazon Kendra 保留索引字段名称来创建保留字段或常用字段。您还可以创建自定义字段。如果您使用数据源连接器，则大多数连接器都包含将数据源文档字段映射到 Amazon Kendra 索引字段的字段映射。如果您使用控制台，则要更新字段，方法是选择数据来源，选择编辑操作，然后在“字段映射”部分旁边继续配置数据来源。

您可以将 `Search` 对象配置为将字段设置为可显示、可分面、可搜索和可排序。您可以将 `Relevance` 对象配置为设置字段的排名顺序、提升持续时间或时间段，以应用于映射到特定字段值的提升、新鲜度、重要性值和重要性值。如果您使用控制台，则可以通过在导航菜单中选择 facet 选项来设置字段的搜索设置。要设置相关性调整，请在导航菜单中选择搜索索引的选项，输入查询，然后使用侧面板选项调整搜索相关性。创建字段后无法更改字段类型。

Amazon Kendra 有以下可供您使用的保留或常用文档字段：

- `_authors` - 负责文档内容的一位或多位作者名单。
- `_category` - 将文档置于特定组中的类别。
- `_created_at` - 以 ISO 8601 格式创建文档的日期和时间。例如，2012-03-25T12:30:10+01:00 是中部欧洲时间 2012 年 3 月 25 日中午 12:30（ 10 秒）的 ISO 8601 日期-时间格式。
- `_data_source_id` - 包含文档数据来源的标识符。
- `_document_body` - 文档的内容。
- `_document_id` - 文档的唯一标识符。
- `_document_title` - 文档标题。
- `_excerpt_page_number` - PDF 文件中显示文档摘录的页码。如果您的索引是在 2020 年 9 月 8 日之前创建的，则必须重新编制文档索引才能使用此属性。
- `_faq_id` - 如果这是问答类型文档（ FAQ ），则为常见问题解答的唯一标识符。
- `_file_type` - 文档的文件类型，例如 pdf 或 doc。
- `_last_updated_at` - 上次更新端点的日期和时间，采用 ISO 8601 格式。例如，2012-03-25T12:30:10+01:00 是中部欧洲时间 2012 年 3 月 25 日中午 12:30（ 10 秒）的 ISO 8601 日期-时间格式。

- _source_uri - 文档可用的 URI。例如，公司网站上的文档的 URI。
- _version - 文档特定版本的标识符。
- _view_count - 查看文档的次数。
- _language_code（字符串）-适用于文档的语言的代码。如果您未指定语言，默认为英语。有关支持的语言（包括其代码）的更多信息，请参阅添加非英语语言文档。

对于自定义字段，您可以将 DocumentMetadataConfigurationUpdates 与 UpdateIndex API 配合使用来创建这些字段，就像创建保留字段或公用字段时一样。您必须为自定义字段设置相应的数据类型。如果您使用控制台，则要更新字段，方法是选择数据来源，选择编辑操作，然后在“字段映射”部分旁边继续配置数据来源。某些数据来源不支持添加新字段或自定义字段。创建字段后无法更改字段类型。

以下是您可以为自定义字段设置的类型：

- Date
- 数字
- 字符串
- 字符串列表

如果您使用 BatchPutDocumentAPI 将文档添加到索引，则会Attributes列出文档的字段/属性，然后使用该DocumentAttribute对象创建字段。

对于从 Amazon S3 数据源编制索引的文档，您可以使用包含字段信息的 JSON 元数据文件创建字段。

如果您使用支持的数据库作为数据来源，则可以使用字段映射选项配置字段。

搜索索引

创建索引后，您可以开始搜索文档。有关更多信息，请参阅搜索索引。

文档

本节说明如何对其支持的多种文档格式以及文档的不同字段/属性进行 Amazon Kendra 索引。

主题

- 文档类型或格式
- 文档属性或字段

文档类型或格式

Amazon Kendra 支持常用的文档类型或格式，例如 PDF、HTML PowerPoint、Word 等。一个索引可以包含多种文档格式。

Amazon Kendra 提取文档内部的内容以使文档可搜索。解析文档的方式是为了优化对提取的文本和文档中任何表格内容（HTML 表格）的搜索。这意味着将文档结构化为用于搜索的字段或属性。文档元数据（例如上次修改日期）可能是有用的搜索字段。

可以将文档组织成行和列。例如，每个文档是一行，每个文档字段/属性（例如标题和正文内容）都是一列。例如，如果您使用数据库作为数据来源，则应将数据结构化或组织成行和列。

您可以通过以下方式将文档添加到索引中：

- BatchPutDocument API
- 数据来源连接器

如果要添加常见问题解答文件，可以使用 CreateFaqAPI 添加存储在存储 Amazon S3 桶中的文件。您可以在基本 CSV 格式、在标题中包含自定义字段/属性的 CSV 格式以及包含自定义字段的 JSON 格式之间进行选择。默认文件格式为 CSV。

以下内容提供了有关每种支持的文档格式以及在为文档编制索引时， Amazon Kendra 如何处理每种格式的信息。

文档格式	视为	如何处理文档	原始结构
可移植文档格式（PDF）	HTML	转换为 HTML，然后提取内容。	非结构化
HyperText 标记语言 (HTML)	HTML	HTML 标签会被过滤掉以提取内容。内容必须介于主 HTML 起始标签和结束标签（<HTML>content</HTML> ）之间。	半结构化
可扩展标记语言 (XML)	XML	XML 标签会被过滤掉以提取内容。	半结构化

文档格式	视为	如何处理文档	原始结构
可扩展样式表语言转换（XSLT）	XSLT	标签会被过滤掉以提取内容。	半结构化
MarkDown （医学博士）	纯文本	提取内容时包含 MarkDown 语法。	半结构化
逗号分隔值（CSV）	CSV	从每个单元格中提取的内容，将单个文件视为单个文档结果。	结构化用于常见问题解答文件，否则为半结构化
Microsoft Excel（XLS 和 XLSX）	XLS 和 XLSX	从每个单元格中提取的内容，将单个文件视为单个文档结果。	半结构化
JavaScript 对象表示法 (JSON)	纯文本	内容是使用包含的 JSON 语法提取的。	半结构化
富文本格式（RTF）	RTF	RTF 语法会被过滤掉以提取内容。	半结构化
微软 PowerPoint (PPT)	PPT	仅从 PowerPoint 幻灯片中提取文本内容进行搜索。不会提取图像和其他内容。	非结构化
Microsoft Word（DOCX）	DOCX	仅从 Word 页面中提取文本内容进行搜索。不会提取图像和其他内容。	非结构化
纯文本（TXT）	TXT	提取文本文档中的所有文本。	非结构化

文档属性或字段

文档具有与之关联的属性或字段。文档的字段是文档的属性或文档结构中包含的内容。例如，您的每个文档都可能包含标题、正文和作者。您也可以为特定文档添加自定义字段。例如，如果您的索引搜索税务文件，则可以为税务文件类型指定自定义字段，例如 W-2、1099 等。

在查询中使用文档字段之前，必须将其映射到索引字段。例如，标题字段可以映射到字段 `_document_title`。有关更多信息，请参阅映射字段。要添加新字段，必须创建要将该字段映射到的索引字段。您可以使用控制台或 UpdateIndexAPI 创建索引字段。

您可以使用文档字段来筛选回复并生成分面搜索结果。例如，您可以筛选回复以仅返回文档的特定版本，也可以筛选搜索结果以仅返回与搜索词匹配的 1099 种税务文件。有关更多信息，请参阅筛选和分面搜索。

您也可以使用文档字段来手动调整查询响应。例如，在确定要在回复中返回哪些文档时，您可以选择提高标题字段的重要性以增加 Amazon Kendra 分配给该字段的权重。有关更多信息，请参阅调整搜索相关性。

如果要将文档直接添加到索引，则需要在 BatchPutDocumentAPI 的文档输入参数中指定字段。您可以在DocumentAttribute对象数组中指定自定义字段值。如果您使用的是数据来源，则用于添加文档字段的方法取决于数据来源。有关更多信息，请参阅映射数据来源字段。

使用 Amazon Kendra 保留或常用文档字段

借助 UpdateIndex API，您可以使用`DocumentMetadataConfigurationUpdates`并指定要映射到等效文档属性/字段名称的 Amazon Kendra 保留索引字段名称来创建保留字段或常用字段。您还可以创建自定义字段。如果您使用数据源连接器，则大多数连接器都包含将数据源文档字段映射到 Amazon Kendra 索引字段的字段映射。如果您使用控制台，则要更新字段，方法是选择数据来源，选择编辑操作，然后在“字段映射”部分旁边继续配置数据来源。

您可以将 `Search` 对象配置为将字段设置为可显示、可分面、可搜索和可排序。您可以将 `Relevance` 对象配置为设置字段的排名顺序、提升持续时间或时间段，以应用于映射到特定字段值的提升、新鲜度、重要性值和重要性值。如果您使用控制台，则可以通过在导航菜单中选择 facet 选项来设置字段的搜索设置。要设置相关性调整，请在导航菜单中选择搜索索引的选项，输入查询，然后使用侧面板选项调整搜索相关性。创建字段后无法更改字段类型。

Amazon Kendra 有以下可供您使用的保留或常用文档字段：

- `_authors` - 负责文档内容的一位或多位作者名单。

- _category - 将文档置于特定组中的类别。
- _created_at - 以 ISO 8601 格式创建文档的日期和时间。例如，2012-03-25T12:30:10+01:00 是中部欧洲时间 2012 年 3 月 25 日中午 12:30（10 秒）的 ISO 8601 日期-时间格式。
- _data_source_id - 包含文档数据来源的标识符。
- _document_body - 文档的内容。
- _document_id - 文档的唯一标识符。
- _document_title - 文档标题。
- _excerpt_page_number - PDF 文件中显示文档摘录的页码。如果您的索引是在 2020 年 9 月 8 日之前创建的，则必须重新编制文档索引才能使用此属性。
- _faq_id - 如果这是问答类型文档（FAQ），则为常见问题解答的唯一标识符。
- _file_type - 文档的文件类型，例如 pdf 或 doc。
- _last_updated_at - 上次更新端点的日期和时间，采用 ISO 8601 格式。例如，2012-03-25T12:30:10+01:00 是中部欧洲时间 2012 年 3 月 25 日中午 12:30（10 秒）的 ISO 8601 日期-时间格式。
- _source_uri - 文档可用的 URI。例如，公司网站上的文档的 URI。
- _version - 文档特定版本的标识符。
- _view_count - 查看文档的次数。
- _language_code（字符串）-适用于文档的语言的代码。如果您未指定语言，默认为英语。有关支持的语言（包括其代码）的更多信息，请参阅添加非英语语言文档。

对于自定义字段，您可以将 DocumentMetadataConfigurationUpdates 与 UpdateIndex API 配合使用来创建这些字段，就像创建保留字段或公用字段时一样。您必须为自定义字段设置相应的数据类型。如果您使用控制台，则要更新字段，方法是选择数据来源，选择编辑操作，然后在“字段映射”部分旁边继续配置数据来源。某些数据来源不支持添加新字段或自定义字段。创建字段后无法更改字段类型。

以下是您可以为自定义字段设置的类型：

- Date
- 数字
- 字符串
- 字符串列表

如果您使用 BatchPutDocumentAPI 将文档添加到索引，则会`Attributes`列出文档的字段/属性，然后使用该`DocumentAttribute`对象创建字段。

对于从 Amazon S3 数据源编制索引的文档，您可以使用包含字段信息的 JSON 元数据文件创建字段。

如果您使用支持的数据库作为数据来源，则可以使用字段映射选项配置字段。

数据来源

数据源是 Amazon Kendra 连接到您的文档或内容并为其编制索引的数据存储库或位置。例如，您可以配置为连接 Amazon Kendra 到 Microsoft SharePoint ，以便对存储在此源中的文档进行抓取和索引。您还可以通过提供 Amazon Kendra 要抓取的网址来索引网页。您可以自动将数据源与 Amazon Kendra 索引同步，这样数据源中添加、更新或删除的文档也可以在索引中添加、更新或删除。

支持的数据来源包括：

- Adobe Experience Manager
- Alfresco
- Aurora （ MySQL ）
- Aurora （ PostgreSQL ）
- Amazon FSx （ 视窗 ）
- Amazon FSx （ NetApp ONTAP ）
- 数据库数据来源
- Amazon RDS （ Microsoft SQL Server ）
- Amazon RDS (MySQL)
- Amazon RDS （ 甲骨文 ）
- Amazon RDS （ PostgreSQL ）
- Amazon S3 水桶
- Amazon Kendra 网络爬虫
- Amazon WorkDocs
- Box
- Confluence
- 自定义数据来源

- Dropbox
- Drupal
- GitHub
- Gmail
- Google Workspace Drives
- IBM DB2
- Jira
- Microsoft Exchange
- 微软 OneDrive
- 微软 SharePoint
- Microsoft Teams
- Microsoft SQL Server
- Microsoft Yammer
- MySQL
- Oracle Database
- PostgreSQL
- Quip
- Salesforce
- ServiceNow
- Slack
- Zendesk

有关支持的文档类型或格式的列表， Amazon Kendra 请参阅文档类型。在创建数据来源连接器之前，必须先创建索引，以便为数据来源中的文档编制索引。

> **Note**
>
> 要创建文档索引，无需使用数据来源。通过批量上传将文档直接添加到索引中。有关更多信息，请参阅将文档直接添加到索引中。

有关使用 Amazon Kendra 控制台、 AWS CLI 或 SDK 的演练，请参阅入门。

查询

要获得答案，用户需要查询索引。用户可以在查询中使用自然语言。该响应包含信息，例如标题、文本摘录以及提供最佳答案的文档在索引中的位置。

Amazon Kendra 使用您提供的有关文档的所有信息，而不仅仅是文档的内容，来确定文档是否与查询相关。例如，如果您的索引包含有关上次更新文档的时间的信息，则可以告诉您 Amazon Kendra 为最近更新的文档分配更高的相关性。

查询还可以包含如何筛选响应的标准，以便仅 Amazon Kendra 返回满足筛选条件的文档。例如，如果您创建了一个名为 department 的索引字段，则可以筛选响应，以便仅返回部门字段设置为 legal 的文档。有关更多信息，请参阅筛选搜索。

您可以通过调整索引中各个字段的相关性来影响查询结果。调整会改变字段在结果中的重要性。例如，如果您使用新类别提高文档的重要性，则该类别的文档更有可能包含在回复中。有关更多信息，请参阅调整搜索相关性。

有关使用查询的更多信息，请参阅搜索索引。

标签

通过分配标签或标签来管理您的索引、数据来源和常见问题解答。您可以使用标签以各种方式对 Amazon Kendra 资源进行分类。例如，按用途、所有者或应用程序进行分类，或按任意组合进行分类。每个标签都由键 和值组成，这两个参数都由您定义。

标签帮助您：

- 识别和整理您的 AWS 资源。许多 AWS 服务都支持标记，因此您可以为不同服务中的资源分配相同的标签，以表明这些资源是相关的。例如，您可以使用相同的标签标记索引和使用该索引的 Amazon Lex 机器人。
- 分配成本。您可以在 AWS Billing and Cost Management 控制面板上激活标签。 AWS 使用标签对您的成本进行分类，并向您提供每月成本分配报告。有关更多信息，请参阅《关于 AWS 账单和成本管理》中的 “成本分配和标记”。
- 控制对资源的访问。您可以在 AWS Identity and Access Management （ IAM ）策略中使用标签来控制对 Amazon Kendra 资源的访问。您可以将这些策略附加到 IAM 角色或用户，以激活基于标签的访问控制。有关更多信息，请参阅基于标签的身份验证。

您可以使用 AWS Management Console、 AWS Command Line Interface (AWS CLI) 或 Amazon Kendra API 创建和管理标签。

标记资源

如果您使用的是 Amazon Kendra 控制台，则可以在创建资源时标记资源或稍后添加资源。您还可以使用控制台来更新或删除标签。

如果您使用的是 AWS Command Line Interface (AWS CLI) 或 Amazon Kendra API，请使用以下操作来管理资源的标签：

- CreateDataSource— 在创建数据源时应用标签。
- CreateFaq—创建常见问题解答时应用标签。
- CreateIndex—创建索引时应用标签。
- ListTagsForResource— 查看与资源关联的标签。
- TagResource— 为资源添加和修改标签。
- UntagResource— 从资源中移除标签。

标签限制

以下限制适用于 Amazon Kendra 资源上的标签：

- 最大标签数量 - 50
- 最大键长度 – 128 个字符
- 最大值长度 – 256 个字符
- 键和值的有效字符 - a-z、A-Z、空格和以下字符：_ . : / = + - 和 @
- 键和值区分大小写
- 请不要使用 aws: 作为键的前缀；它保留为供 AWS 使用

设置 Amazon Kendra

在使用 Amazon Kendra 之前，您必须拥有 Amazon Web Services (AWS) 账户。拥有 AWS 账户后，您可以通过亚马逊 Kendra 控制台、AWS CLI () 或软件开发工具包访问 Amazon Kendra。 AWS Command Line Interface AWS

本指南包括 Java 和 Python 的示例。 AWS CLI

主题

报名参加 AWS

当您注册亚马逊 Web Services (AWS) 时，您的账户会自动注册所有服务 AWS，包括亚马逊 Kendra。您只需为使用的服务付费。

如果您已经有一个 AWS 帐户，请跳到下一个任务。如果您还没有 AWS 账户，请使用以下步骤创建。

要注册 AWS

1. 打开 https://aws.amazon.com，然后选择 “创建 AWS 账户”。
2. 按照屏幕上的说明完成账户创建。请记下您的 12 位 AWS 账号。作为注册流程的一部分，您会收到一个电话，需要您使用电话键盘输入一个 PIN 码。
3. 创建 AWS Identity and Access Management (IAM) 管理员用户。有关说明，请参阅《AWS Identity and Access Management 用户指南》中的创建您的第一个 IAM 用户和组。

区域和端点

终端节点是作为 Web 服务入口点的 URL。每个终端节点都与特定 AWS 区域相关联。如果您组合使用 Amazon Kendra 控制台 AWS CLI、和 Amazon Kendra 软件开发工具包，请注意它们的默认区域，因为给定活动的所有亚马逊 Kendra 组件（索引、查询等）都必须在同一区域创建。有关 Amazon Kendra 支持的所有区域和端点的列表，请参阅区域和端点。

设置 AWS CLI

AWS 命令行界面 (AWS CLI) 是一款用于管理 AWS 服务（包括 Amazon Kendra）的统一开发者工具。我们建议您安装它。

1. 要安装 AWS CLI，请按照《AWS 命令行界面用户指南》中的 “安装AWS命令行界面” 中的说明进行操作。
2. 要配置 AWS CLI 和设置配置文件以调用 AWS CLI，请按照《AWS 命令行界面用户指南》中配置AWS CLI中的说明进行操作。
3. 要确认配置 AWS CLI 文件配置是否正确，请运行以下命令：

   ```
   aws configure --profile default
   ```

   如果您的配置文件已正确配置，您将看到类似于以下内容的输出：

   ```
   AWS Access Key ID [****************52FQ]:
   AWS Secret Access Key [****************xgyZ]:
   Default region name [us-west-2]:
   Default output format [json]:
   ```

4. 要验证是否已配置 AWS CLI 为与 Amazon Kendra 配合使用，请运行以下命令：

   ```
   aws kendra help
   ```

   如果配置 AWS CLI 正确，您将看到 Amazon Kendra、Amazon Kendra 运行时和 Amazon Kendra 事件支持的 AWS CLI 命令列表。

设置 AWS 软件开发工具包

下载并安装您要使用的 AWS 软件开发工具包。本指南提供了适用于 Python 的示例。有关其他 AWS 软件开发工具包的信息，请参阅适用于 Amazon Web Services 的工具。

Python 开发工具包的软件包名为 Boto3。

在运行以下 Python 命令之前，必须先为您的操作系统下载并安装 Python 3.6 或更高版本。对 Python 3.5 及更早版本的支持已过时。如果您的 Python 脚本目录中没有包含 pip，则可以下载 get-pip.py 并将其存储在 Scripts 目录中。您也可以使用终端程序将 Python 目录设置为 Path 或环境变量。

```
# Install the latest Boto3 release via pip
pip install boto3

# You can install a specific version of Boto3 for compatibility reasons
# Install Boto3 version 1.0 specifically
pip install boto3==1.0.0

# Make sure Boto3 is no older than version 1.15.0
pip install boto3>=1.15.0

# Avoid versions of Boto3 newer than version 1.15.3
pip install boto3<=1.15.3
```

要使用 Boto3，您必须使用 IAM 控制台为您的 AWS 账户设置身份验证证书。

IAM 的访问角色 Amazon Kendra

创建索引、数据源或常见问题解答时， Amazon Kendra 需要访问创建 AWS Amazon Kendra 资源所需的资源。在创建 Amazon Kendra 资源之前，必须先创建 AWS Identity and Access Management (IAM) 策略。当您调用该操作时，请提供附有策略的角色的 Amazon 资源名称 (ARN)。例如，如果您调用 BatchPutDocumentAPI 来添加 Amazon S3 存储桶中的文档，则需要为角色提供 Amazon Kendra 具有访问存储桶权限的策略。

您可以在 Amazon Kendra 控制台中创建新 IAM 角色或选择要使用的 IAM 现有角色。控制台显示的角色的角色名称中包含字符串“kendra”或“Kendra”。

以下主题提供了所需策略的详细信息。如果您使用 Amazon Kendra 控制台创建 IAM 角色，则会为您创建这些策略。

主题

IAM 索引的角色

创建索引时，必须为 IAM 角色提供写入的权限 Amazon CloudWatch。您还必须提供允许代入该角色 Amazon Kendra 的信任策略。以下是必须提供的策略。

IAM 索引的角色

允许访问 CloudWatch 日志 Amazon Kendra 的角色策略。

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "cloudwatch:PutMetricData",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "cloudwatch:namespace": "AWS/Kendra"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "logs:DescribeLogGroups",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "logs:CreateLogGroup",
            "Resource": "arn:aws:logs:your-region:your-account-id:log-group:/aws/
kendra/*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:DescribeLogStreams",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": "arn:aws:logs:your-region:your-account-id:log-group:/aws/
kendra/*:log-stream:*"
        }
    ]
}
```

允许访问 Amazon Kendra 的角色策略 AWS Secrets Manager。如果您使用用户上下文 Secrets Manager 作为关键位置，则可以使用以下策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
```

```
            "Effect":"Allow",
            "Action":"cloudwatch:PutMetricData",
            "Resource":"*",
            "Condition":{
               "StringEquals":{
                  "cloudwatch:namespace":"AWS/Kendra"
               }
            }
         },
         {
            "Effect":"Allow",
            "Action":"logs:DescribeLogGroups",
            "Resource":"*"
         },
         {
            "Effect":"Allow",
            "Action":"logs:CreateLogGroup",
            "Resource":"arn:aws:logs:your-region:your-account-id:log-group:/aws/kendra/*"
         },
         {
            "Effect":"Allow",
            "Action":[
               "logs:DescribeLogStreams",
               "logs:CreateLogStream",
               "logs:PutLogEvents"
            ],
            "Resource":"arn:aws:logs:your-region:your-account-id:log-group:/aws/kendra/
*:log-stream:*"
         },
         {
            "Effect":"Allow",
            "Action":[
               "secretsmanager:GetSecretValue"
            ],
            "Resource":[
               "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
            ]
         },
         {
            "Effect":"Allow",
            "Action":[
               "kms:Decrypt"
            ],
            "Resource":[
```

```
            "arn:aws:kms:your-region:your-account-id:key/key-id"
         ],
         "Condition":{
            "StringLike":{
               "kms:ViaService":[
                  "secretsmanager.your-region.amazonaws.com"
               ]
            }
         }
      }
   ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM BatchPutDocumentAPI 的角色

> **Warning**
>
> Amazon Kendra 不使用向 Amazon Kendra 委托人授予与 S3 存储桶交互的权限的存储桶策略。它使用 IAM 角色。请确保该成员 Amazon Kendra 未作为可信成员包含在存储桶策略中，以避免在意外向任意委托人授予权限时出现任何数据安全问题。但是，您可以添加存储桶策略，以便在不同的账户中使用 Amazon S3 存储桶。有关更多信息，请参阅跨账户使用 Amazon S3 的策略。有关 S3 数据来源的 IAM 角色的信息，请参阅 IAM 角色。

使用 BatchPutDocumentAPI 为 Amazon S3 存储桶中的文档编制索引时，必须为 IAM 角色 Amazon Kendra 提供对存储桶的访问权限。您还必须提供允许代入该角色 Amazon Kendra 的信任策略。如果存储桶中的文档已加密，则必须提供使用 AWS KMS 客户主密钥 (CMK) 解密文档的权限。

IAM BatchPutDocumentAPI 的角色

允许 Amazon Kendra 访问 Amazon S3 存储桶的必需角色策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucket-name/*"
            ]
        }
    ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

建议您在信任策略中包含 aws:sourceAccount 和 aws:sourceArn。这会限制权限并安全地检查aws:sourceAccount和aws:sourceArn是否与sts:AssumeRole操作的 IAM 角色策略中提供的

相同。这样可以防止未经授权的实体访问您的 IAM 角色及其权限。有关更多信息，请参阅有关困惑的副手问题的 AWS Identity and Access Management指南。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "kendra.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "your-account-id"
                },
                "StringLike": {
                    "aws:SourceArn": "arn:aws:kendra:your-region:your-account-id:index/
*"
                }
            }
        }
    ]
}
```

一种可选的角色策略 Amazon Kendra ，允许使用 AWS KMS 客户主密钥 (CMK) 解密存储桶中的文档。 Amazon S3

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:kms:your-region:your-account-id:key/key-id"
            ]
        }
    ]
```

```
}
```

IAM 数据源的角色

使用 CreateDataSourceAPI 时，必须 Amazon Kendra 授予有权访问资源的 IAM 角色。所需的特定权限取决于数据来源。

IAM Adobe 体验管理器数据源的角色

当您使用 Adobe Experience Manager 时，需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Adobe 体验管理器进行身份验证的权限。
- 调用 Adobe Experience Manager 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过连接 Adobe Experience Manager 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
```

```
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/[[key-id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.{{your-region}}.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
```

```
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Alfresco 数据源的角色

当您使用 Alfresco 时，需要为角色提供以下策略。

- 允许访问您的 AWS Secrets Manager 密钥以对您的 Alfresco 进行身份验证。
- 调用 Alfresco 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以将 Alfresco 数据源连接到。 Amazon Kendra Amazon VPC如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
```

```
    "Resource": [
      "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/[[key-id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.{{your-region}}.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
```

```
            "Action":"sts:AssumeRole"
        }
    ]
}
```

IAMAurora (MySQL) 数据源的角色

使用 Aurora (MySQL) 时，您需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Aurora (MySQL) 进行身份验证的权限。
- 允许调用 Aurora (MySQL) 连接器所需的公共 API。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过连接 Aurora (MySQL) 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
```

```
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
```

```
}
```

IAMAurora (PostgreSQL) 数据源的角色

当你使用 Aurora (PostgreSQL) 时，你为角色提供以下策略。

- 允许访问您的 AWS Secrets Manager 密钥以对您的 Aurora (PostgreSQL) 进行身份验证。
- 调用 Aurora (PostgreSQL) 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以将 Aurora (PostgreSQL) 数据源连接到。 Amazon Kendra Amazon VPC如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
```

```
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAMAmazon FSx 数据源的角色

使用时 Amazon FSx，您可以为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Amazon FSx 文件系统进行身份验证的权限。
- 访问您的 Amazon FSx 文件系统所在位置 Amazon Virtual Private Cloud (VPC) 的权限。
- 获取 Amazon FSx 文件系统活动目录域名的权限。
- 调用 Amazon FSx 连接器所需的公共 API 的权限。
- 允许调用 `BatchPutDocument` 和 `BatchDeleteDocument` API 来更新索引的权限。

```
{
    "Version": "2012-10-17",
        "Statement": [
        {
          "Effect": "Allow",
          "Action": [
            "secretsmanager:GetSecretValue"
          ],
          "Resource": [
            "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:
{{secret-id}}"
          ]
        },
        {
          "Effect": "Allow",
          "Action": [
            "kms:Decrypt"
          ],
          "Resource": [
            "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/{{key-id}}"
          ],
          "Condition": {
            "StringLike": {
              "kms:ViaService": [
                "secretsmanager.{{your-region}}.amazonaws.com"
              ]
            }
          }
        },
        {
          "Effect": "Allow",
```

```
            "Action":[
              "ec2:CreateNetworkInterface",
              "ec2:DeleteNetworkInterface"
            ],
            "Resource": [
                  "arn:aws:ec2:{{your-region}}:{{your-account-id}}:network-interface/*",
                  "arn:aws:ec2:{{your-region}}:{{your-account-id}}:subnet/[[subnet-ids]]"
            ]
          },
          {
            "Effect": "Allow",
            "Action": [
              "ec2:DescribeSubnets",
              "ec2:DescribeNetworkInterfaces"
            ],
            "Resource": "*"
          },
          {
            "Effect": "Allow",
            "Action": [
              "ec2:CreateNetworkInterfacePermission"
            ],
            "Resource": "arn:aws:ec2:{{your-region}}:{{your-account-id}}:network-
interface/*",
            "Condition": {
              "StringEquals": {
                "ec2:AuthorizedService": "kendra.*.amazonaws.com"
              },
              "ArnEquals": {
                "ec2:Subnet": [
                  "arn:aws:ec2:{{your-region}}:{{your-account-id}}:subnet/[[subnet-ids]]"
                ]
              }
            }
          },
          {
            "Sid": "AllowsKendraToGetDomainNameOfActiveDirectory",
            "Effect": "Allow",
            "Action": "ds:DescribeDirectories",
            "Resource": "*"
          },
          {
            "Sid": "AllowsKendraToCallRequiredFsxAPIs",
            "Effect": "Allow",
```

```
          "Action": [
              "fsx:DescribeFileSystems"
          ],
          "Resource": "*"
        },
        {
          "Sid": "iamPassRole",
          "Effect": "Allow",
          "Action": "iam:PassRole",
          "Resource": "*",
          "Condition": {
            "StringEquals": {
              "iam:PassedToService": [
                "kendra.*.amazonaws.com"
              ]
            }
          }
        },
        {
          "Effect": "Allow",
          "Action": [
            "kendra:BatchPutDocument",
            "kendra:BatchDeleteDocument"
          ],
          "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/
{{index-id}}"
        }
        ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
```

```
}
```

IAM 数据库数据源的角色

当您使用数据库作为数据源时，您需要提供一个 Amazon Kendra 具有连接所需的权限的角色。其中包括：

- 访问包含网站用户名和密码的 AWS Secrets Manager 密钥的权限。有关密钥内容的更多信息，请参阅数据来源。
- 允许使用 AWS KMS 客户主密钥 (CMK) 解密存储的用户名和密码密钥。 Secrets Manager
- 使用 `BatchPutDocument` 和 `BatchDeleteDocument` 操作更新索引的权限。
- 访问包含用于与网站通信的 SSL 证书的 Amazon S3 存储桶的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过将数据库数据源连接到 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue"
            ],
            "Resource": [
                "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:kms:your-region:your-account-id:key/key-id"
            ]
```

```
        },
        {
            "Effect": "Allow",
            "Action": [
                "kendra:BatchPutDocument",
                "kendra:BatchDeleteDocument"
            ],
            "Resource": [
                "arn:aws:kendra:your-region:your-account-id:index/index-id"
            "Condition": {
                "StringLike": {
                    "kms:ViaService": [
                        "kendra.your-region.amazonaws.com"
                    ]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucket-name/*"
            ]
        }
    ]
}
```

您可以对数据来源使用两种可选策略。

如果您对包含用于与通信的 SSL 证书的 Amazon S3 存储桶进行了加密，请提供 Amazon Kendra 允许访问该密钥的策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
```

```
                "arn:aws:kms:your-region:your-account-id:key/key-id"
            ]
        }
    ]
}
```

如果您使用的是 VPC，请提供 Amazon Kendra 允许访问所需资源的策略。对于所需策略，请参阅 数据来源、VPC 的 IAM 角色。

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAMAmazon RDS （微软 SQL Server）数据源的角色

当你使用 Amazon RDS （Microsoft SQL Server）数据源连接器时，需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Amazon RDS （Microsoft SQL Server）数据源实例进行身份验证的权限。
- 允许调用 Amazon RDS （Microsoft SQL Server）数据源连接器所需的公共 API。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 你可以 Amazon Kendra 通过连接 Amazon RDS （微软 SQL Server）数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
```

```
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAMAmazon RDS (MySQL) 数据源的角色

使用 Amazon RDS (MySQL) 数据源连接器时，您需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Amazon RDS (MySQL) 数据源实例进行身份验证的权限。
- 有权调用 Amazon RDS (MySQL) 数据源连接器所需的公共 API。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过连接 Amazon RDS (MySQL) 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
```

```
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
"arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
```

```
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAMAmazon RDS (Oracle) 数据源的角色

使用 Amazon RDS Oracle 数据源连接器时，您需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Amazon RDS (Oracle) 数据源实例进行身份验证的权限。
- 有权调用 Amazon RDS (Oracle) 数据源连接器所需的公共 API。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过连接 O Amazon RDS racle 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
```

```
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
"arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
```

```
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAMAmazon RDS (PostgreSQL) 数据源的角色

当您使用 Amazon RDS (PostgreSQL) 数据源连接器时，您需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Amazon RDS (PostgreSQL) 数据源实例进行身份验证的权限。
- 调用 Amazon RDS (PostgreSQL) 数据来源连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以将 Amazon RDS (PostgreSQL) 数据源连接到。 Amazon Kendra Amazon VPC如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
```

```
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAMAmazon S3 数据源的角色

> **Warning**
>
> Amazon Kendra 不使用向 Amazon Kendra 委托人授予与 S3 存储桶交互的权限的存储桶策略。相反，它使用 IAM 角色。请确保该成员 Amazon Kendra 未作为可信成员包含在存储桶策略中，以避免在意外向任意委托人授予权限时出现任何数据安全问题。但是，您可以添加存储桶策略，以便在不同的账户中使用 Amazon S3 存储桶。有关更多信息，请参阅 跨账户使用 Amazon S3 的策略（向下滚动）。

当您使用 Amazon S3 存储桶作为数据源时，您需要提供一个有权访问存储桶以及使用`BatchPutDocument`和`BatchDeleteDocument`操作的角色。如果 Amazon S3 存储桶中的文档已加密，则必须提供使用 AWS KMS 客户主密钥 (CMK) 解密文档的权限。

以下角色策略必须 Amazon Kendra 允许代入角色。继续向下滚动以查看代入角色的信任策略。

允许 Amazon Kendra 将 Amazon S3 存储桶用作数据源的必需角色策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "s3:GetObject"
            ],
```

```
            "Resource": [
                "arn:aws:s3:::bucket-name/*"
            ],
            "Effect": "Allow"
        },
        {
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::bucket-name"
            ],
            "Effect": "Allow"
        },
        {
            "Effect": "Allow",
            "Action": [
                "kendra:BatchPutDocument",
                "kendra:BatchDeleteDocument"
            ],
            "Resource": [
                "arn:aws:kendra:your-region:your-account-id:index/index-id"
            ]
        }
    ]
}
```

一种可选的角色策略 Amazon Kendra ，允许使用 AWS KMS 客户主密钥 (CMK) 解密存储桶中的文档。 Amazon S3

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:kms:your-region:your-account-id:key/key-id"
            ]
        }
    ]
```

```
}
```

一种可选的角色策略，Amazon Kendra 允许在使用 Amazon S3 存储桶时访问存储桶 Amazon VPC，且无需激活 AWS KMS 或共享 AWS KMS 权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Action": [
        "s3:GetObject"
      ],
      "Resource": [
        "arn:aws:s3:::{{bucket-name}}/*"
      ],
      "Effect": "Allow"
    },
    {
      "Action": [
        "s3:ListBucket"
      ],
      "Resource": [
        "arn:aws:s3:::{{bucket-name}}"
      ],
      "Effect": "Allow"
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:{{your-region}}:{{your-account-id}}:subnet/[[subnet-ids]]",
        "arn:aws:ec2:{{your-region}}:{{your-account-id}}:security-group/[[security-
group]]"
      ]
    },
    {
        "Effect": "Allow",
        "Action": [
            "ec2:CreateNetworkInterface"
        ],
```

```
        "Resource": "arn:aws:ec2:{{your-region}}:{{your-account-id}}:network-interface/
*",
        "Condition": {
            "StringLike": {
                "aws:RequestTag/AWS_KENDRA": "kendra_{{your-account-id}}_{{index-
id}}_{data-source-id}}_*"
            }
        }
    },
    {
        "Effect": "Allow",
        "Action": [
            "ec2:CreateTags"
        ],
        "Resource": "arn:aws:ec2:{{your-region}}:{{your-account-id}}:network-interface/
*",
        "Condition": {
            "StringEquals": {
                "ec2:CreateAction": "CreateNetworkInterface"
            }
        }
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeSubnets"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeNetworkInterfaces"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterfacePermission"
      ],
      "Resource": "arn:aws:ec2:{{your-region}}:{{your-accoount-id}}:network-interface/
*",
      "Condition": {
```

```
        "StringEquals": {
          "ec2:AuthorizedService": "kendra.amazonaws.com"
        },
        "ArnEquals": {
          "ec2:Subnet": [
            "arn:aws:ec2:{{your-region}}:{{your-account-id}}:subnet/[[subnet-ids]]"
          ]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
      ],
      "Resource": [
        "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}",
        "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "kendra:BatchPutDocument",
        "kendra:BatchDeleteDocument"
      ],
      "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}"
    }
  ]
}
```

一项可选的角色策略 Amazon Kendra ，允许在使用时访问 Amazon S3 存储桶 Amazon VPC，并激活 AWS KMS 权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
```

```
      "Action": [
        "s3:GetObject"
      ],
      "Resource": [
        "arn:aws:s3:::{{bucket-name}}/*"
      ],
      "Effect": "Allow"
    },
    {
      "Action": [
        "s3:ListBucket"
      ],
      "Resource": [
        "arn:aws:s3:::{{bucket-name}}"
      ],
      "Effect": "Allow"
    },
    {
      "Effect": "Allow",
      "Action": [
        "kms:Decrypt"
      ],
      "Resource": [
        "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/{{key-id}}"
      ],
      "Condition": {
        "StringLike": {
          "kms:ViaService": [
            "s3.{{your-region}}.amazonaws.com"
          ]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:{{your-region}}:{{your-account-id}}:subnet/[[subnet-ids]]",
        "arn:aws:ec2:{{your-region}}:{{your-account-id}}:security-group/[[security-
group]]"
      ]
    },
```

```
    {
        "Effect": "Allow",
        "Action": [
            "ec2:CreateNetworkInterface"
        ],
        "Resource": "arn:aws:ec2:{{your-region}}:{{your-account-id}}:network-interface/
*",
        "Condition": {
            "StringLike": {
                "aws:RequestTag/AWS_KENDRA": "kendra_{{your-account-id}}_{{index-
id}}_{data-source-id}}_*"
            }
        }
    },
    {
        "Effect": "Allow",
        "Action": [
            "ec2:CreateTags"
        ],
        "Resource": "arn:aws:ec2:{{your-region}}:{{your-account-id}}:network-interface/
*",
        "Condition": {
            "StringEquals": {
                "ec2:CreateAction": "CreateNetworkInterface"
            }
        }
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeSubnets"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeNetworkInterfaces"
      ],
      "Resource": "*"
    },
    {
      "Effect": "Allow",
      "Action": [
```

```
        "ec2:CreateNetworkInterfacePermission"
      ],
      "Resource": "arn:aws:ec2:{{your-region}}:{{your-account-id}}:network-interface/
*",
      "Condition": {
        "StringEquals": {
          "ec2:AuthorizedService": "kendra.amazonaws.com"
        },
        "ArnEquals": {
          "ec2:Subnet": [
            "arn:aws:ec2:{{your-region}}:{{your-account-id}}:subnet/[[subnet-ids]]"
          ]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
      ],
      "Resource": [
        "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}",
        "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "kendra:BatchPutDocument",
        "kendra:BatchDeleteDocument"
      ],
      "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}"
    }
  ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

跨账户使用 Amazon S3 的策略

如果您的 Amazon S3 存储桶与您用于 Amazon Kendra 索引的账户位于不同的账户中，则可以创建跨账户使用该存储桶的策略。

当 Amazon S3 存储桶与您的 Amazon Kendra 索引位于不同的账户中时，使用您的存储桶作为数据源的角色策略。请注意，`s3:PutObject` 和 `s3:PutObjectAcl` 是可选的，如果要为访问控制列表包含配置文件，则可以使用此选项。

```
{
    "Version": "2012-10-17",
    "Statement": [
         {
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::$bucket-in-other-account/*"
            ],
            "Effect": "Allow"
        },
        {
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::$bucket-in-other-account/*"
            ],
            "Effect": "Allow"
```

```
        },
        {
            "Effect": "Allow",
            "Action": [
                "kendra:BatchPutDocument",
                "kendra:BatchDeleteDocument"
            ],
            "Resource": [
                "arn:aws:kendra:$your-region:$your-account-id:index/$index-id"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject",
                "s3:PutObject",
                "s3:PutObjectAcl"
            ],
            "Resource": "arn:aws:s3:::$bucket-in-other-account/*"
        }
    ]
}
```

允许 Amazon S3 数据源角色跨账户访问 Amazon S3 存储桶的存储桶策略。请注意，s3:PutObject 和 s3:PutObjectAcl 是可选的，如果要为访问控制列表包含配置文件，则可以使用此选项。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "AWS": "$kendra-s3-connector-role-arn"
            },
            "Action": [
                "s3:GetObject",
                "s3:PutObject",
                "s3:PutObjectAcl"
            ],
            "Resource": [
                "arn:aws:s3:::$bucket-in-other-account/*"
            ]
        },
```

```
        {
            "Effect": "Allow",
            "Principal": {
                "AWS": "$kendra-s3-connector-role-arn"
            },
            "Action": "s3:ListBucket",
            "Resource": "arn:aws:s3:::$bucket-in-other-account"
        }
    ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAMAmazon Kendra Web Crawler 数据源的角色

使用 Amazon Kendra Web Crawler 时，您需要为角色提供以下策略：

- 访问包含连接网站或由基本身份验证支持的 Web 代理服务器的凭据的 AWS Secrets Manager 密钥的权限。有关机密报告内容的更多信息，请参阅使用 Web 爬网程序数据来源。
- 允许使用 AWS KMS 客户主密钥 (CMK) 解密存储的用户名和密码密钥。 Secrets Manager
- 使用 `BatchPutDocument` 和 `BatchDeleteDocument` 操作更新索引的权限。
- 如果您使用 Amazon S3 存储桶来存储种子网址或站点地图列表，请添加访问该 Amazon S3 存储桶的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过 Amazon VPC连接 Amazon Kendra Web Crawler 数据源。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account:key/key-id"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
```

```
}
```

如果您将种子网址或站点地图存储在 Amazon S3 存储桶中，则必须向该角色添加此权限。

```
,
{"Effect": "Allow",
     "Action": [
         "s3:GetObject"
      ],
      "Resource": [
         "arn:aws:s3:::bucket-name/*"
      ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAMAmazon WorkDocs 数据源的角色

使用时 Amazon WorkDocs，您可以为角色提供以下策略

- 验证与您的 Amazon WorkDocs 站点存储库对应的目录 ID（组织 ID）的权限。
- 获取包含您的 Amazon WorkDocs 网站目录的 Active Directory 域名的权限。
- 调用 Amazon WorkDocs 连接器所需的公共 API 的权限。
- 允许调用 `BatchPutDocument` 和 `BatchDeleteDocument` API 来更新索引的权限。

```
{
```

```
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowsKendraToGetDomainNameOfActiveDirectory",
      "Effect": "Allow",
      "Action": "ds:DescribeDirectories",
      "Resource": "*"
    },
    {
      "Sid": "AllowsKendraToCallRequiredWorkDocsAPIs",
      "Effect": "Allow",
      "Action": [
        "workdocs:GetDocumentPath",
        "workdocs:GetGroup",
        "workdocs:GetDocument",
        "workdocs:DownloadDocumentVersions",
        "workdocs:DescribeUsers",
        "workdocs:DescribeFolderContents",
        "workdocs:DescribeActivities",
        "workdocs:DescribeComments",
        "workdocs:GetFolder",
        "workdocs:DescribeResourcePermissions",
        "workdocs:GetFolderPath",
        "workdocs:DescribeInstances"
      ],
      "Resource": "*"
    },
    {
      "Sid": "iamPassRole",
      "Effect": "Allow",
      "Action": "iam:PassRole",
      "Resource": "*",
      "Condition": {
        "StringEquals": {
          "iam:PassedToService": [
            "kendra.amazonaws.com"
          ]
        }
      }
    },
    {
      "Sid": "AllowsKendraToCallBatchPutDeleteAPIs",
      "Effect": "Allow",
      "Action": [
```

```
        "kendra:BatchPutDocument",
        "kendra:BatchDeleteDocument"
      ],
      "Resource": [
        "arn:aws:kendra:your-region:account-id:index/$index-id"
      ]
    }
  ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Box 数据源的角色

当您使用 Box 时，为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Slack 进行身份验证的权限。
- 调用 Box 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过连接 Box 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/[[key-id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.{{your-region}}.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-d}}:index/{{index-
id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
```

```
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Confluence 数据源的角色

IAM Confluence 连接器 v1.0 的角色

当您使用 Confluence Server 作为数据来源时，您需要为角色提供以下策略：

- 访问包含连接 Confluence 所需凭据的 AWS Secrets Manager 密钥的权限。有关密钥内容的更多信息，请参阅 Confluence 数据来源。
- 允许使用 AWS KMS 客户主密钥 (CMK) 解密存储的用户名和密码密钥。 Secrets Manager
- 使用 `BatchPutDocument` 和 `BatchDeleteDocument` 操作更新索引的权限。

> **Note**
>
> 您可以通过连接 Confluence 数据源。 Amazon Kendra Amazon VPC如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/key-id"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
}
```

如果您使用的是 VPC，请提供 Amazon Kendra 允许访问所需资源的策略。对于所需策略，请参阅 数据来源、VPC 的 IAM 角色。

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Confluence 连接器 v2.0 的角色

对于 Confluence 连接器 v2.0 数据来源，您需要为角色提供以下策略。

- 访问包含 Confluence 身份验证凭据的 AWS Secrets Manager 密钥的权限。有关密钥内容的更多信息，请参阅 Confluence 数据来源。
- 允许使用 AWS KMS 客户主密钥 (CMK) 解密存储的用户名和密码密钥。 AWS Secrets Manager
- 使用 `BatchPutDocument` 和 `BatchDeleteDocument` 操作更新索引的权限。

您还必须附上允许代入该角色 Amazon Kendra 的信任策略。

> **Note**
>
> 您可以通过连接 Confluence 数据源。 Amazon Kendra Amazon VPC如果您使用的是 Amazon VPC，则需要添加其他权限。

允许连接 Amazon Kendra 到 Confluence 的角色策略。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "secretsmanager:GetSecretValue"
      ],
```

```
      "Resource": [
        "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "kms:Decrypt"
      ],
      "Resource": [
        "arn:aws:kms:your-region:your-account-id:key/key-id"
      ],
      "Condition": {
        "StringLike": {
          "kms:ViaService": [
            "secretsmanager.your-region.amazonaws.com"
          ]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
      ],
      "Resource": [
        "arn:aws:kendra:your-region:your-account-id:index/index-id",
        "arn:aws:kendra:your-region:your-account-id:index/index-id/data-source/*"
      ]
    }
    {
      "Effect": "Allow",
      "Action": [
        "kendra:BatchPutDocument",
        "kendra:BatchDeleteDocument"
      ],
      "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
    }
  ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Dropbox 数据源的角色

当您使用 Dropbox 时，为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Dropbox 进行身份验证的权限。
- 调用 Dropbox 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过连接 Dropbox 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
"Version": "2012-10-17",
  "Statement": [
  {"Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
```

```
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {"Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/[[key-id]]"
    ],
    "Condition": {"StringLike": {"kms:ViaService": [
          "secretsmanager.{{your-region}}.amazonaws.com"
        ]
      }
    }
  },
  {"Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"]
  },
  {"Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
```

```
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Drupal 数据源的角色

当您使用 Drupal 时，为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Drupal 进行身份验证的权限。
- 调用 Drupal 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以将 Drupal 数据源连接 Amazon Kendra 到 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
```

```
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
```

```
            "Action":"sts:AssumeRole"
        }
    ]
}
```

IAM GitHub 数据源的角色

使用时 GitHub，您可以为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您进行身份验证的权限 GitHub。
- 允许为 GitHub 连接器调用所需的公共 API。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过将 GitHub 数据源连接到 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/[[key-id]]"
```

```
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.{{your-region}}.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
```

```
    ]
}
```

IAM Gmail 数据源的角色

当您使用 Gmail 时，为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Gmail 进行身份验证的权限。
- 调用 Gmailconnector 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过将 Gmail 数据源连接到 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
"Version": "2012-10-17",
  "Statement": [
  {"Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {"Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/[[key-id]]"
    ],
    "Condition": {"StringLike": {"kms:ViaService": [
          "secretsmanager.{{your-region}}.amazonaws.com"
        ]
```

```
      }
    }
  },
  {"Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"]
  },
  {"Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Google 云端硬盘数据源的角色

当您使用 Google Workspace 云端硬盘数据源时，您提供的 Amazon Kendra 角色具有连接到该网站所需的权限。其中包括：

- 获取和解 AWS Secrets Manager 密包含客户帐号电子邮件、管理员帐号电子邮件地址和连接 Google 云端硬盘网站所需的私钥的密钥的权限。有关密钥内容的更多信息，请参阅 Google Drive 数据来源。
- 使用BatchPutDocument和 BatchDeleteDocumentAPI 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过连接 Google 云端硬盘数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

以下 IAM 策略提供了必要的权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/key-id"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com"
        ]
      }
    }
  },
```

```
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM IBM DB2 数据源的角色

当您使用 IBM DB2 数据来源连接器时，需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对 IBM DB2 数据源实例进行身份验证的权限。
- 调用 IBM DB2 数据来源连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以将 IBM DB2 数据源连接 Amazon Kendra 到 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
```

```
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Jira 数据源的角色

当您使用 Jira 时，为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对 Jira 进行身份验证的权限。
- 调用 Jira 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过 Amazon VPC连接 Jira 数据源。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
```

```
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/[[key-id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.{{your-region}}.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
```

```
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 微软 Exchange 数据源的角色

当你使用 Microsoft Exchange 数据源时，你需要提供一个 Amazon Kendra 具有连接到该站点所需的权限的角色。其中包括：

- 允许获取和解 AWS Secrets Manager 密包含连接到 Microsoft Exchange 站点所需的应用程序 ID 和密钥的密钥。有关密钥内容的更多信息，请参阅 Microsoft Exchange 数据来源。
- 使用BatchPutDocument和 BatchDeleteDocumentAPI 的权限。

> **Note**
>
> 你可以 Amazon Kendra 通过连接微软 Exchange 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

以下 IAM 策略提供了必要的权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
```

```
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/key-id"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
}
```

如果您要将要索引的用户列表存储在 Amazon S3 存储桶中，则还必须提供使用 S3 `GetObject` 操作的权限。以下 IAM 策略提供了必要权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
```

```
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Action": [
      "s3:GetObject"
    ],
    "Resource": [
      "arn:aws:s3:::bucket-name/*"
    ],
    "Effect": "Allow"
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/[[key-ids]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com",
          "s3.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 微软 OneDrive 数据源的角色

当你使用 Microsoft OneDrive 数据源时，你提供的 Amazon Kendra 角色具有连接到该站点所需的权限。其中包括：

- 获取和解 AWS Secrets Manager 密包含连接到站点所需的应用程序 ID 和密钥的密钥的 OneDrive 权限。有关密钥内容的更多信息，请参阅 Microsoft OneDrive 数据源。
- 使用BatchPutDocument和 BatchDeleteDocumentAPI 的权限。

> **Note**
>
> 你可以 Amazon Kendra 通过连接 Microsoft OneDrive 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

以下 IAM 策略提供了必要的权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
```

```
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/key-id"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
}
```

如果您要将要索引的用户列表存储在 Amazon S3 存储桶中，则还必须提供使用 S3 `GetObject` 操作的权限。以下 IAM 策略提供了必要权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
```

```
    ]
  },
  {
    "Action": [
      "s3:GetObject"
    ],
    "Resource": [
      "arn:aws:s3:::bucket-name/*"
    ],
    "Effect": "Allow"
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/[[key-ids]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com",
          "s3.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
```

```
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 微软 SharePoint 数据源的角色

IAM SharePoint连接器 v1.0 的角色

对于 Microsoft SharePoint 连接器 v1.0 数据源，您需要为角色提供以下策略。

- 访问包含 SharePoint 网站用户名和密码的 AWS Secrets Manager 密钥的权限。有关密钥内容的更多信息，请参阅 Microsoft SharePoint 数据源。
- 允许使用 AWS KMS 客户主密钥 (CMK) 解密存储的用户名和密码密钥。 AWS Secrets Manager
- 使用 `BatchPutDocument` 和 `BatchDeleteDocument` 操作更新索引的权限。
- 访问包含用于与 SharePoint 网站通信的 SSL 证书的 Amazon S3 存储桶的权限。

您还必须附上允许代入该角色 Amazon Kendra 的信任策略。

> **Note**
>
> 你可以 Amazon Kendra 通过连接 Microsoft SharePoint 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue"
            ],
            "Resource": [
```

```
                "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:kms:your-region:your-account-id:key/key-id"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "kendra:BatchPutDocument",
                "kendra:BatchDeleteDocument"
            ],
            "Resource": [
                "arn:aws:kendra:your-region:your-account-id:index/index-id"
            ],
            "Condition": {
                "StringLike": {
                    "kms:ViaService": [
                        "kendra.your-region.amazonaws.com"
                    ]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucket-name/*"
            ]
        }
    ]
}
```

如果您对包含用于与 SharePoint 网站通信的 SSL 证书的 Amazon S3 存储桶进行了加密，请提供 Amazon Kendra 允许访问该密钥的策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:kms:your-region:your-account-id:key/key-id"
            ]
        }
    ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM SharePoint连接器 v2.0 的角色

对于 Microsoft conn SharePoint ector v2.0 数据源，您需要为角色提供以下策略。

- 访问包含 SharePoint 站点身份验证凭据的 AWS Secrets Manager 密钥的权限。有关密钥内容的更多信息，请参阅 Microsoft SharePoint 数据源。
- 允许使用 AWS KMS 客户主密钥 (CMK) 解密存储的用户名和密码密钥。 AWS Secrets Manager
- 使用 `BatchPutDocument` 和 `BatchDeleteDocument` 操作更新索引的权限。
- 访问包含用于与 SharePoint 网站通信的 SSL 证书的 Amazon S3 存储桶的权限。

您还必须附上允许代入该角色 Amazon Kendra 的信任策略。

> **Note**
>
> 你可以 Amazon Kendra 通过连接 Microsoft SharePoint 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "secretsmanager:GetSecretValue"
      ],
      "Resource": [
        "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "kms:Decrypt"
      ],
      "Resource": [
        "arn:aws:kms:your-region:your-account-id:key/key-id"
      ],
      "Condition": {
        "StringLike": {
          "kms:ViaService": [
            "secretsmanager.your-region.amazonaws.com"
          ]
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
```

```
        "kendra:DescribePrincipalMapping"
      ],
      "Resource": [
        "arn:aws:kendra:your-region:your-account-id:index/index-id",
        "arn:aws:kendra:your-region:your-account-id:index/index-id/data-source/*"
      ]
    },
    {
      "Action": [
        "s3:GetObject"
      ],
      "Resource": [
        "arn:aws:s3:::bucket-name/key-name"
      ],
      "Effect": "Allow"
    },
    {
      "Effect": "Allow",
      "Action": [
        "kendra:BatchPutDocument",
        "kendra:BatchDeleteDocument"
      ],
      "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:your-region:your-account-id:subnet/subnet-ids",
        "arn:aws:ec2:your-region:your-account-id:security-group/security-group"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": "arn:aws:ec2:region:account_id:network-interface/*",
      "Condition": {
        "StringLike": {
          "aws:RequestTag/AWS_KENDRA": "kendra_your-account-id_index-id_*"
        }
```

```
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:your-region:your-account-id:network-interface/*",
      "Condition": {
        "StringEquals": {
          "ec2:CreateAction": "CreateNetworkInterface"
        }
      }
    },

{
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterfacePermission"
      ],
      "Resource": "arn:aws:ec2:your-region:your-account-id:network-interface/*",
      "Condition": {
        "StringLike": {
          "aws:ResourceTag/AWS_KENDRA": "kendra_your-account-id_index-id_*"
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeNetworkInterfaces",
        "ec2:DescribeAvailabilityZones",
        "ec2:DescribeNetworkInterfaceAttribute",
        "ec2:DescribeVpcs",
        "ec2:DescribeRegions",
        "ec2:DescribeNetworkInterfacePermissions",
        "ec2:DescribeSubnets"
      ],
      "Resource": "*"
    }
  ]
}
```

如果您对包含用于与 SharePoint 网站通信的 SSL 证书的 Amazon S3 存储桶进行了加密，请提供 Amazon Kendra 允许访问该密钥的策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:kms:your-region:youraccount-id:key/key-id"
            ]
        }
    ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 微软 SQL Server 数据源的角色

当您使用 Microsoft SQL Server 时，需要为角色提供以下策略。

- 允许访问你的 AWS Secrets Manager 密钥来验证你的 Microsoft SQL Server 实例。
- 调用 Microsoft SQL Server 数据来源连接器所需的公共 API 的权限。

- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMapp` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 你可以 Amazon Kendra 通过连接微软 SQL Server 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
```

```
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 微软 Teams 数据源的角色

当你使用 Microsoft Team Amazon Kendra s 数据源时，你提供的角色具有连接到该站点所需的权限。其中包括：

- 获取和解 AWS Secrets Manager 密包含连接到 Microsoft Teams 所需的客户端 ID 和客户机密的密钥的权限。有关密钥内容的更多信息，请参阅 Microsoft Teams 数据来源。

> **Note**
>
> 你可以 Amazon Kendra 通过连接 Microsoft Teams 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

以下 IAM 策略提供了必要的权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:client-id:secret:secret-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/key-id"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
```

```
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 微软 Yammer 数据源的角色

当你使用 Microsoft Yammer 数据源时，你提供的 Amazon Kendra 角色具有连接到该站点所需的权限。其中包括：

- 允许获取和解 AWS Secrets Manager 密包含连接到 Microsoft Yammer 网站所需的应用程序 ID 和密钥的密钥。有关密钥内容的更多信息，请参阅 Microsoft Yammer 数据来源。
- 使用BatchPutDocument和 BatchDeleteDocumentAPI 的权限。

> **Note**
>
> 你可以 Amazon Kendra 通过 Amazon VPC连接 Microsoft Yammer 数据源。如果您使用的是 Amazon VPC，则需要添加其他权限。

以下 IAM 策略提供了必要的权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
```

```
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/key-id"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
}
```

如果您要将要索引的用户列表存储在 Amazon S3 存储桶中，则还必须提供使用 S3 `GetObject` 操作的权限。以下 IAM 策略提供了必要权限：

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
```

```
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Action": [
      "s3:GetObject"
    ],
    "Resource": [
      "arn:aws:s3:::bucket-name/*"
    ],
    "Effect": "Allow"
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/[[key-ids]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com",
          "s3.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM MySQL 数据源的角色

当您使用 MySQL 数据来源连接器时，需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以验证您的 My SQL 数据源实例的权限。
- 调用 MySQL 数据来源连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过连接 MySQL 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
```

```
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
```

```
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Oracle 数据源的角色

当您使用 Oracle 数据来源连接器时，需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对 Oracle 数据源实例进行身份验证的权限。
- 调用 Oracle 数据来源连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过连接 Oracle 数据源 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
```

```
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
```

```
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM PostgreSQL 数据源的角色

当您使用 PostgreSQL 数据来源连接器时，需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 PostgreSQL 数据源实例进行身份验证的权限。
- 调用 PostgreSQL 数据来源连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以将 PostgreSQL 数据源连接到。 Amazon Kendra Amazon VPC如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{region}}:{{account_id}}:secret:[[secret_id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
```

```
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{region}}:{{account_id}}:key/[[key_id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.*.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}",
 "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{region}}:{{account_id}}:index/{{index_id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
```

```
        },
        "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Quip 数据源的角色

当您使用 Quip 时，需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对 Quip 进行身份验证的权限。
- 调用 Quip 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过 Amazon VPC连接 Quip 数据源。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
```

```
      "arn:aws:kms:{{your-region}}:{{yoour-account-id}}:key/[[key-id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.{{your-region}}.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{your-
index-id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{your-index-
id}}/data-source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
```

```
      }
    ]
}
```

IAM Salesforce 数据源的角色

当您使用 Salesforce 作为数据来源时，需要为角色提供以下策略：

- 访问包含 Salesforce 网站用户名和密码的 AWS Secrets Manager 密钥的权限。有关清单报告内容的更多信息，请参阅 Salesforce 数据来源。
- 允许使用 AWS KMS 客户主密钥 (CMK) 解密存储的用户名和密码密钥。 Secrets Manager
- 使用 `BatchPutDocument` 和 `BatchDeleteDocument` 操作更新索引的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过 Amazon VPC连接 Salesforce 数据源。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/key-id"
    ],
    "Condition": {
```

```
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:account-id:index/index-id"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM ServiceNow 数据源的角色

当您使用 ServiceNow 作为数据源时，您需要为角色提供以下策略：

- 访问包含 ServiceNow 网站用户名和密码的 Secrets Manager 密钥的权限。有关密钥内容的更多信息，请参阅ServiceNow 数据来源。
- 允许使用 AWS KMS 客户主密钥 (CMK) 解密存储的用户名和密码密钥。 Secrets Manager
- 使用 `BatchPutDocument` 和 `BatchDeleteDocument` 操作更新索引的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过将 ServiceNow 数据源连接到 Amazon VPC。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:your-region:your-account-id:secret:secret-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/key-id"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.your-region.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:your-region:your-account-id:index/index-id"
  }]
```

```
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Slack 数据源的角色

当您使用 Slack 时，需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Slack 进行身份验证的权限。
- 调用 Slack 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过 Amazon VPC连接 Slack 数据源。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
```

```
      "secretsmanager:GetSecretValue"
    ],
    "Resource": [
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{your-region}}:{{account-id}}:key/[[key-id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.{{region}}.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM Zendesk 数据源的角色

当您使用 Zendesk 时，需要为角色提供以下策略。

- 访问您的 AWS Secrets Manager 密钥以对您的 Zendesk 套件进行身份验证的权限。
- 调用 Zendesk 连接器所需的公共 API 的权限。
- 调用 `BatchPutDocument`、`BatchDeleteDocument`、`PutPrincipalMapping`、`DeletePrincipalMappi` 和 `ListGroupsOlderThanOrderingId` API 的权限。

> **Note**
>
> 您可以 Amazon Kendra 通过 Amazon VPC连接 Zendesk 数据源。如果您使用的是 Amazon VPC，则需要添加其他权限。

```
{
  "Version": "2012-10-17",
  "Statement": [
  {
    "Effect": "Allow",
    "Action": [
      "secretsmanager:GetSecretValue"
    ],
```

```
    "Resource": [
      "arn:aws:secretsmanager:{{your-region}}:{{your-account-id}}:secret:[[secret-id]]"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt"
    ],
    "Resource": [
      "arn:aws:kms:{{your-region}}:{{your-account-id}}:key/[[key-id]]"
    ],
    "Condition": {
      "StringLike": {
        "kms:ViaService": [
          "secretsmanager.{{your-region}}.amazonaws.com"
        ]
      }
    }
  },
  {
    "Effect": "Allow",
    "Action": [
        "kendra:PutPrincipalMapping",
        "kendra:DeletePrincipalMapping",
        "kendra:ListGroupsOlderThanOrderingId",
        "kendra:DescribePrincipalMapping"
    ],
    "Resource": ["arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-
id}}", "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}/data-
source/*"]
  },
  {
    "Effect": "Allow",
    "Action": [
      "kendra:BatchPutDocument",
      "kendra:BatchDeleteDocument"
    ],
    "Resource": "arn:aws:kendra:{{your-region}}:{{your-account-id}}:index/{{index-id}}"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

虚拟私有云 (VPC) IAM 角色

如果您使用虚拟私有云 (VPC) 连接到数据源，则必须提供以下额外权限。

VPC IAM 角色

```
{
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": [
        "arn:aws:ec2:{{region}}:{{account_id}}:subnet/[[subnet_ids]]",
        "arn:aws:ec2:{{region}}:{{account_id}}:security-group/[[security_group]]"
      ]
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterface"
      ],
      "Resource": "arn:aws:ec2:{{region}}:{{account_id}}:network-interface/*",
      "Condition": {
        "StringLike": {
          "aws:RequestTag/AWS_KENDRA": "kendra_{{account_id}}_{{index_id}}_*"
        }
      }
    },
    {
```

```
      "Effect": "Allow",
      "Action": [
        "ec2:CreateTags"
      ],
      "Resource": "arn:aws:ec2:{{region}}:{{account_id}}:network-interface/*",
      "Condition": {
        "StringEquals": {
          "ec2:CreateAction": "CreateNetworkInterface"
        }
      }
    },

{
      "Effect": "Allow",
      "Action": [
        "ec2:CreateNetworkInterfacePermission"
      ],
      "Resource": "arn:aws:ec2:{{region}}:{{account_id}}:network-interface/*",
      "Condition": {
        "StringLike": {
          "aws:ResourceTag/AWS_KENDRA": "kendra_{{account_id}}_{{index_id}}_*"
        }
      }
    },
    {
      "Effect": "Allow",
      "Action": [
        "ec2:DescribeNetworkInterfaces",
        "ec2:DescribeAvailabilityZones",
        "ec2:DescribeNetworkInterfaceAttribute",
        "ec2:DescribeVpcs",
        "ec2:DescribeRegions",
        "ec2:DescribeNetworkInterfacePermissions",
        "ec2:DescribeSubnets"
      ],
      "Resource": "*"
    }
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
```

```
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 常见问题解答 (FAQ) 的角色

使用 CreateFaqAPI 将问题和答案加载到索引中时，必须为 IAM 角色 Amazon Kendra 提供对包含源文件的 Amazon S3 存储桶的访问权限。如果源文件已加密，则必须提供使用 AWS KMS 客户主密钥 (CMK) 解密文件的权限。

IAM 常见问题解答中的角色

允许 Amazon Kendra 访问 Amazon S3 存储桶的必需角色策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucket-name/*"
            ]
        }
    ]
}
```

一项可选的角色策略 Amazon Kendra ，允许使用 AWS KMS 客户主密钥 (CMK) 解密存储桶中的文件。 Amazon S3

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:kms:your-region:your-account-id:key/key-id"
            ],
            "Condition": {
                "StringLike": {
                    "kms:ViaService": [
                        "kendra.your-region.amazonaws.com"
                    ]
                }
            }
        }
    ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 查询建议的角色

当您使用 Amazon S3 文件作为查询建议阻止列表时，您需要提供一个有权访问该 Amazon S3 文件和 Amazon S3 存储桶的角色。如果 Amazon S3 存储桶中的阻止列表文本 Amazon S3 文件（该文件）已加密，则必须提供使用 AWS KMS 客户主密钥 (CMK) 解密文档的权限。

IAM 查询建议的角色

允许 Amazon Kendra 使用该 Amazon S3 文件作为查询建议屏蔽列表的必需角色策略。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {"Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucket-name/*"
            ]
        }
    ]
}
```

一种可选的角色策略 Amazon Kendra ，允许使用 AWS KMS 客户主密钥 (CMK) 解密存储桶中的文档。 Amazon S3

```
{
    "Version": "2012-10-17",
    "Statement": [
        {"Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:kms:your-region:your-account-id:key/key-id"
            ]
        }
    ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
```

```
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 用于用户和组的主体映射的角色

当您使用 PutPrincipalMappingAPI 将用户映射到他们的群组以按用户上下文筛选搜索结果时，您需要提供属于某个群组的用户或子群组的列表。如果您的列表中某个群组的用户或子群组超过 1000 个，则需要提供一个有权访问您的列表 Amazon S3 文件和 Amazon S3 存储桶的角色。如果 Amazon S3 存储桶中列表的文本 Amazon S3 文件（文件）已加密，则必须提供使用 AWS KMS 客户主密钥 (CMK) 解密文档的权限。

IAM 主体映射的角色

一项必需的角色策略 Amazon Kendra ，允许将该 Amazon S3 文件用作属于某个群组的用户和子群组的列表。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {"Effect": "Allow",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucket-name/*"
            ]
        }
    ]
}
```

一种可选的角色策略 Amazon Kendra ，允许使用 AWS KMS 客户主密钥 (CMK) 解密存储桶中的文档。 Amazon S3

```
{
    "Version": "2012-10-17",
    "Statement": [
```

```
        {"Effect": "Allow",
            "Action": [
                "kms:Decrypt"
            ],
            "Resource": [
                "arn:aws:kms:your-region:your-account-id:key/key-id"
            ]
        }
    ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

建议您在信任策略中包含 aws:sourceAccount 和 aws:sourceArn。这会限制权限并安全地检查aws:sourceAccount和aws:sourceArn是否与sts:AssumeRole操作的 IAM 角色策略中提供的相同。这样可以防止未经授权的实体访问您的 IAM 角色及其权限。有关更多信息，请参阅有关困惑的副手问题的 AWS Identity and Access Management指南。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "kendra.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole",
```

```
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "your-account-id"
                },
                "StringLike": {
                    "aws:SourceArn": "arn:aws:kendra:your-region:your-account-id:index-
id/*"
                }
            }
        }
    ]
}
```

IAM 的角色 AWS IAM Identity Center

当您使用该UserGroupResolutionConfiguration对象从 AWS IAM Identity Center 身份源获取群组和用户的访问权限级别时，您需要提供一个具有访问权限的角色 IAM Identity Center。

IAM 的角色 AWS IAM Identity Center

允许 Amazon Kendra 访问的必需角色策略 IAM Identity Center。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "sso-directory:SearchUsers",
                "sso-directory:ListGroupsForUser",
                "sso-directory:DescribeGroups",
                "sso:ListDirectoryAssociations"
            ],
            "Resource": [
                 "*"
            ]
        },
        {
          "Sid": "iamPassRole",
          "Effect": "Allow",
          "Action": "iam:PassRole",
          "Resource": "*",
          "Condition": {
```

```
            "StringEquals": {
              "iam:PassedToService": [
                "kendra.amazonaws.com"
              ]
            }
          }
        }
    ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

IAM 角色换 Amazon Kendra 体验

当您使用CreateExperience或 UpdateExperienceAPI 创建或更新搜索应用程序时，必须提供一个有权访问必要操作和 IAM Identity Center 的角色。

IAMAmazon Kendra 搜索体验角色

允许 Amazon Kendra 访问Query操作、QuerySuggestionsSubmitFeedback操作、操作和存储您的用户和群组信息的 IAM Identity Center 所需的角色策略。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowsKendraSearchAppToCallKendraApi",
      "Effect": "Allow",
```

```
      "Action": [
        "kendra:GetQuerySuggestions",
        "kendra:Query",
        "kendra:DescribeIndex",
        "kendra:ListFaqs",
        "kendra:DescribeDataSource",
        "kendra:ListDataSources",
        "kendra:DescribeFaq",
        "kendra:SubmitFeedback"
      ],
      "Resource": [
        "arn:aws:kendra:your-region:your-account-id:index/index-id"
      ]
    },
    {
      "Sid": "AllowKendraSearchAppToDescribeDataSourcesAndFaq",
      "Effect": "Allow",
      "Action": [
        "kendra:DescribeDataSource",
        "kendra:DescribeFaq"
      ],
      "Resource": [
        "arn:aws:kendra:your-region:your-account-id:index/index-id/data-source/data-
source-id",
        "arn:aws:kendra:your-region:your-account-id:index/index-id/faq/faq-id"
      ]
    },
    {
      "Sid": "AllowKendraSearchAppToCallSSODescribeUsersAndGroups",
      "Effect": "Allow",
      "Action": [
        "sso-directory:ListGroupsForUser",
        "sso-directory:SearchGroups",
        "sso-directory:SearchUsers",
        "sso-directory:DescribeUser",
        "sso-directory:DescribeGroup",
        "sso-directory:DescribeGroups",
        "sso-directory:DescribeUsers",
        "sso:ListDirectoryAssociations"
      ],
      "Resource": [
        "*"
      ],
      "Condition": {
```

```
        "StringLike": {
          "kms:ViaService": [
            "kendra.your-region.amazonaws.com"
          ]
        }
      }
    }
  ]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

建议您在信任策略中包含 aws:sourceAccount 和 aws:sourceArn。这会限制权限并安全地检查aws:sourceAccount和aws:sourceArn是否与sts:AssumeRole操作的 IAM 角色策略中提供的相同。这样可以防止未经授权的实体访问您的 IAM 角色及其权限。有关更多信息，请参阅有关困惑的副手问题的 AWS Identity and Access Management指南。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "kendra.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole",
            "Condition": {
```

```
                "StringEquals": {
                    "aws:SourceAccount": "your-account-id"
                },
                "StringLike": {
                    "aws:SourceArn": "arn:aws:kendra:your-region:your-account-id:index-
id/*"
                }
            }
        }
    ]
}
```

IAM 自定义文档扩充的角色

当您使用该CustomDocumentEnrichmentConfiguration对象对文档元数据和内容进行高级更改时，必须提供一个具有运行和 PreExtractionHookConfiguration / 或PostExtractionHookConfiguration所需权限的角色。您可以配置 Lambda 函数，以便 PreExtractionHookConfiguration 和/或 PostExtractionHookConfiguration 在提取过程中对文档元数据和内容进行高级更改。如果您选择为 Amazon S3 存储桶激活服务器端加密，则必须提供使用 AWS KMS 客户主密钥 (CMK) 加密和解密存储在存储桶中的对象的权限。 Amazon S3

IAM 自定义文档扩充的角色

允许运行的必需角色策略 Amazon Kendra ，PreExtractionHookConfiguration并对存储 Amazon S3 桶PostExtractionHookConfiguration进行加密。

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Action": [
      "s3:GetObject",
      "s3:PutObject"
    ],
    "Resource": [
      "arn:aws:s3:::bucket-name/*"
    ],
    "Effect": "Allow"
  },
  {
    "Action": [
      "s3:ListBucket"
```

```
    ],
    "Resource": [
      "arn:aws:s3:::bucket-name"
    ],
    "Effect": "Allow"
  },
  {
    "Effect": "Allow",
    "Action": [
      "kms:Decrypt",
      "kms:GenerateDataKey"
    ],
    "Resource": [
      "arn:aws:kms:your-region:your-account-id:key/key-id"
    ]
  },
  {
    "Effect": "Allow",
    "Action": [
      "lambda:InvokeFunction"
    ],
    "Resource": "arn:aws:lambda:your-region:your-account-id:function:lambda-function"
  }]
}
```

一个可选的角色策略 Amazon Kendra ，允许在PostExtractionHookConfiguration不加密存储 Amazon S3 桶的情况下运行PreExtractionHookConfiguration。

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Action": [
      "s3:GetObject",
      "s3:PutObject"
    ],
    "Resource": [
      "arn:aws:s3:::bucket-name/*"
    ],
    "Effect": "Allow"
  },
  {
    "Action": [
      "s3:ListBucket"
```

```
    ],
    "Resource": [
      "arn:aws:s3:::bucket-name"
    ],
    "Effect": "Allow"
  },
  {
    "Effect": "Allow",
    "Action": [
      "lambda:InvokeFunction"
    ],
    "Resource": "arn:aws:lambda:your-region:your-account-id:function:lambda-function"
  }]
}
```

允许担任角色 Amazon Kendra 的信任策略。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Principal":{
            "Service":"kendra.amazonaws.com"
         },
         "Action":"sts:AssumeRole"
      }
   ]
}
```

建议您在信任策略中包含 aws:sourceAccount 和 aws:sourceArn。这会限制权限并安全地检查aws:sourceAccount和aws:sourceArn是否与sts:AssumeRole操作的 IAM 角色策略中提供的相同。这样可以防止未经授权的实体访问您的 IAM 角色及其权限。有关更多信息，请参阅有关困惑的副手问题的 AWS Identity and Access Management指南。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
```

```
                    "kendra.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole",
            "Condition": {
                "StringEquals": {
                    "aws:SourceAccount": "your-account-id"
                },
                "StringLike": {
                    "aws:SourceArn": "arn:aws:kendra:your-region:your-account-id:index-
id/*"
                }
            }
        }
    ]
}
```

部署 Amazon Kendra

当您需要将 Amazon Kendra 搜索部署到网站时，我们会提供源代码，您可以在 React 中使用这些源代码，让您的应用程序抢占先机。根据 MIT 许可修订版，源代码免费提供。您可以按原样使用，也可以根据自己的需求进行更改。提供的 React 应用程序是帮助您入门的示例。该应用程序不能用于生产。

要部署无代码的搜索应用程序并生成具有访问控制功能的搜索页面的端点 URL，请参阅 Amazon Kendra Experience Builder。

以下示例代码为现有的 React Web 应用程序添加了 Amazon Kendra 搜索功能：

- https://kendrasamples.s3.amazonaws.com/kendrasamples-react-app.zip - 开发人员可用于在现有的 React Web 应用程序中构建实用搜索体验的示例文件。

这些示例以 Amazon Kendra 控制台的搜索页面为模型进行创建。它们具有相同的搜索和显示搜索结果的功能。您可以使用整个示例，也可以只选择其中一个功能来供自己使用。

要在 Amazon Kendra 控制台中查看搜索页面的三个组成部分，请从右侧菜单中选择代码图标 (</>)。将鼠标指针悬停在各个部分上可查看组件的简短描述并获取组件源代码的 URL。

主题

概述

您可以将示例代码添加到现有 React Web 应用程序中来激活搜索。示例代码包括一个自述文件，其中包含设置新 React 开发环境的步骤。示例代码中的示例数据可用于演示搜索。示例代码中的搜索文件和组件结构如下：

- 主搜索页面 (`Search.tsx`) - 这是包含所有组件的主页。您可以在这里将应用程序与 Amazon Kendra API 集成。
- 搜索栏 - 这是用户输入搜索词并调用搜索功能的组件。
- 结果 - 这是显示 Amazon Kendra 的结果的组件。它由三个部分组成：建议答案、常见问题解答结果和推荐文档。
- 分面 - 该组件在搜索结果中显示分面，并允许您选择一个分面来缩小搜索范围。
- 分页 - 这是对来自 Amazon Kendra 的响应进行分页的组件。

先决条件

在开始之前，您需要：

- 已安装 Node.js 和 npm。需要使用 Node.js 版本 19 或更早版本。
- 已下载和安装 Python 3 或 Python 2。
- SDK for Java 或 AWS SDK for JavaScript，以便对 Amazon Kendra 调用 API。
- 现有 React Web 应用程序。示例代码包括一个自述文件，其中包含有关如何设置新 React 开发环境的步骤（包括使用所需的框架/库）。您也可以按照 React 文档中的快速入门说明创建 React Web 应用程序。
- 在开发环境中配置的必需库和依赖项。示例代码包括一个自述文件，其中列出了所需的库和软件包依赖关系。请注意，`sass` 是必需的，因为已弃用 `node-sass`。如果您之前安装过 `node-sass`，请将其卸载并安装 `sass`。

设置示例

向 React 应用程序添加 Amazon Kendra 搜索功能的完整过程见代码示例中包含的自述文件。

开始使用 kendrasamples-react-app.zip

1. 确保您已经完成先决条件，包括下载和安装 Node.js 和 npm。

2. 下载 kendrasamples-react-app.zip 并解压缩。
3. 打开您的终端并转到 aws-kendra-example-react-app/src/services/。打开 local-dev-credentials.json 并提供您的凭证。不要将此文件添加到任何公共存储库中。
4. 转到 aws-kendra-example-react-app 并在 package.json 中安装依赖项。运行 npm install。
5. 在本地服务器上启动应用程序的演示版。运行 npm start。在键盘上输入 Cmd/Ctrl + C 可停止本地服务器。
6. 您可以通过访问 package.json 并更新主机和端口来更改端口或主机（例如，IP 地址）："start": "HOST=[host] PORT=[port] react-scripts start"。如果您使用的是 Windows："start": "set HOST=[host] && set PORT=[port] && react-scripts start"。
7. 如果您有注册的网站域名，则可以在应用程序名称后的 package.json 中指定域名。例如，"homepage": "https://mywebsite.com"。必须再次运行 npm install 才能更新新的依赖项，然后运行 npm start。
8. 要构建应用程序，请运行 npm build。将构建目录中的内容上传到您的托管服务提供商。

> **Warning**
>
> React 应用程序未准备好投入生产。这是部署用于 Amazon Kendra 搜索的应用程序的示例。

主搜索页面

主搜索页面 (Search.tsx) 包含所有示例搜索组件。它包括用于输出的搜索栏组件、用于显示来自 Query API 的响应的结果组件，以及用于分页浏览响应的分页组件。

搜索组件

搜索组件提供了一个用于输入查询文本的文本框。onSearch 函数是一个钩子，它可以在 Search.tsx 中调用主函数来进行 Amazon Kendra Query API 调用。

结果组件

结果组件显示来自 Query API 的响应。结果显示在三个不同的区域中。

- 建议答案 - 这些是 Query API 返回的热门结果。它最多包含三个建议的答案。在响应中，它们的结果类型为 ANSWER。
- 常见问题答案 - 这些是响应返回的常见问题答案结果。常见问题答案会单独添加到索引中。在响应中，它们的类型为 QUESTION_ANSWER。有关更多信息，请参阅问题和答案。
- 推荐文档 - 这些是 Amazon Kendra 在回复中返回的其他文档。在 Query API 的响应中，它们的类型为 DOCUMENT。

结果组件共享一组用于突出显示、标题、链接等功能的组件。必须存在共享组件才能让结果组件正常工作。

分面组件

分面组件列出了搜索结果中可用的分面。每个分面都按特定的维度（例如，作者）来对响应进行分类。通过从列表中选择一个分面，您可以将搜索范围缩小到特定的分面。

选择一个分面后，该组件会使用属性筛选条件调用 Query，该筛选条件将搜索范围限制为与该分面匹配的文档。

分页组件

分页组件允许您在多个页面中显示来自 Query API 的搜索结果。它使用 PageSize 和 PageNumber 参数调用 Query API 以获取特定页面的结果。

无需代码即可打造搜索体验

无需任何前端代码即可构建和部署 Amazon Kendra 搜索应用程序。Amazon Kendra只需点击几下，Experience Builder 即可帮助您构建和部署功能齐全的搜索应用程序，这样您就可以立即开始搜索。您可以自定义设计搜索页面并调整搜索，以根据用户的需求量身定制体验。Amazon Kendra 生成搜索页面的唯一且完全托管的端点 URL，以便开始搜索您的文档和常见问题解答。您可以快速建立搜索体验的概念证明，并与他人分享。

您可以使用构建器中提供的搜索体验模板来自定义搜索。您可以邀请其他人合作打造您的搜索体验，也可以评估搜索结果以进行调整。一旦您的搜索体验准备就绪，可供用户开始搜索，就只需共享安全端点 URL。

搜索 Experience Builder 的工作原理

构建搜索体验的总体过程如下所述：

1. 您可以通过为搜索体验命名、描述并选择要用于搜索体验的数据来源来创建搜索体验。
2. 您可以在 AWS IAM Identity Center 中配置您的用户和组列表，然后为其分配搜索体验的访问权限。您将自己列为体验的所有者。有关更多信息，请参阅the section called “提供对搜索页面的访问权限”。
3. 您可以打开 Amazon Kendra Experience Builder 来设计和调整搜索页面。您可以与您分配了自己的编辑访问权限或查看搜索访问权限的其他人共享您的搜索体验的端点 URL。

您可以调用 CreateExperience API 来创建和配置您的搜索体验。如果您使用控制台，则可以选择您的索引，然后在导航菜单中选择体验 来配置您的体验。

设计和调整您的搜索体验

创建和配置搜索体验后，您可以使用端点 URL 来打开搜索体验，以拥有编辑者访问权限的所有者的身份开始自定义搜索。在搜索框中键入查询，然后使用侧面板上的编辑选项对搜索进行自定义，以查看它们如何应用于您的页面。准备好发布时，选择发布。您还可以在切换到实时视图（查看搜索页面的最新发布版本）和切换到构建模式（编辑或自定义搜索页面）之间切换。

以下是您可以自定义搜索体验的方法。

筛选条件

添加分面搜索或按文档属性筛选。这包括自定义属性。您可以使用自己配置的元数据字段添加筛选条件。例如，要按每个城市类别进行分面搜索，请使用包含所有城市类别的 `_category` 自定义文档属性。

建议的答案

将机器学习生成的答案添加到用户的查询中。例如，对于“这门课程有多难？”。Amazon Kendra 可以检索所有涉及课程难度的文档中最相关的文本，并推荐最相关的答案。

常见问题

添加常见问题解答文档以提供常见问题的答案。例如，“完成这门课程需要多少小时？”。Amazon Kendra 可以使用包含此问题答案的常见问题解答文档并给出正确的答案。

排序

添加对搜索结果的排序，以便您的用户可以按相关性、创建时间、上次更新时间和其他排序标准来组织结果。

文档

配置文档或搜索结果在搜索页面上的显示方式。您可以配置页面上显示多少结果，包括页码等分页，激活用户反馈按钮，以及安排文档元数据字段在搜索结果中的显示方式。

语言

选择一种语言来筛选所选语言的搜索结果或文档。

搜索框

配置搜索框的大小和占位符文本，并允许提出查询建议。

相关性优化

为文档元数据字段添加增强功能，以便在用户搜索文档时提高这些字段的权重。您可以添加从 1 开始并逐渐增加到 10 的权重。您可以增强文本、日期和数值字段的类型。例如，要赋予 `_last_updated_at` 和 `_created_at` 比其他字段更高的权重或重要性，请根据这些字段的重要性将这些字段的权重设为 1 到 10。您可以为每个搜索应用程序或体验应用不同的相关性调整配置。

提供对搜索页面的访问权限

通过 IAM 身份中心访问您的搜索体验。配置搜索体验时，您授予身份中心目录中列出的其他人访问您的 Amazon Kendra 搜索页面的权限。他们会收到一封电子邮件，指示他们使用自己在 IAM Identity Center 中的证书登录以访问搜索页面。您必须在 AWS Organizations 中的组织级别或账户持有人级别设置 IAM 身份中心。有关设置 IAM IAM Identy 的更多信息，请参阅 IAM Identity Center 入门。

您可以根据自己的搜索体验在 IAM Identity Center 中激活用户身份，并使用 API 或控制台分配查看者或所有者访问权限。

- 查看者：允许发布查询、接收与搜索相关的建议答案，并向 Amazon Kendra 提供反馈，以便不断改进搜索。
- 所有者：允许自定义搜索页面的设计、调整搜索以及将搜索应用程序用作查看器。目前不支持在控制台中禁用查看者的访问权限。

要向其他人分配访问您的搜索体验的访问权限，请先使用 ExperienceConfiguration 对象在 IAM Identity Center 中使用您的 Amazon Kendra 体验激活用户身份。您可以指定包含用户标识符（例如用户名或电子邮件地址）的字段名称。然后，您可以使用 AssociateEntitiesToExperience API 向您的用户列表授予访问您的搜索体验的权限，并使用 AssociatePersonasToEntities API 将他们

的权限定义为查看者或所有者。您可以使用 EntityConfiguration 对象指定每个用户或组，并使用 EntityPersonaConfiguraton 对象指定该用户或组是查看者还是所有者。

要使用控制台向其他人分配访问您的搜索体验的权限，您首先需要创建体验并确认您的身份以及您是所有者。然后，您可以将其他用户或组指定为查看者或所有者。在控制台中，选择您的索引，然后在导航菜单中选择体验。创建体验后，您可以从列表中选择您的体验。转到访问权限管理，将用户或组分配为查看者或所有者。

配置搜索体验

以下是配置或创建搜索体验的示例。

Console

创建 Amazon Kendra 搜索体验

1. 在左侧导航窗格的索引下，选择体验，然后选择创建体验。
2. 在配置体验页面上，输入体验的名称和描述，选择您的内容来源，然后为您的体验选择 IAM 角色。有关 IAM 角色的更多信息，请参阅 适用于 Amazon Kendra 体验的 IAM 角色。
3. 在通过 Identity Center 目录确认您的身份页面上，选择您的用户 ID，例如，您的电子邮件地址。如果您没有 Identity Center 目录，只需输入您的全名和电子邮件即可创建身份中心目录。这包括作为体验用户的您，并会自动为您分配所有者访问权限。
4. 在查看以打开 Experience Builder 页面上，查看您的配置详细信息，然后选择创建体验并打开 Experience Builder 以开始编辑您的搜索页面。

CLI

创建 Amazon Kendra 体验

```
aws kendra create-experience \
 --name experience-name \
 --description "experience description" \
 --index-id index-id \
 --role-arn arn:aws:iam::account-id:role/role-name \
 --configuration '{"ExperienceConfiguration":[{"ContentSourceConfiguration":
{"DataSourceIds":["data-source-1","data-source-2"]},
 "UserIdentityConfiguration":"identity attribute name"}]}'

aws kendra describe-experience \
 --endpoints experience-endpoint-URL(s)
```

Python

创建 Amazon Kendra 体验

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Create an experience.")

# Provide a name for the experience
name = "experience-name"
# Provide an optional description for the experience
description = "experience description"
# Provide the index ID for the experience
index_id = "index-id"
# Provide the IAM role ARN required for Amazon Kendra experiences
role_arn = "arn:aws:iam::${account-id}:role/${role-name}"
# Configure the experience
configuration = {"ExperienceConfiguration":
        [{
            "ContentSourceConfiguration":{"DataSourceIds":["data-source-1","data-
source-2"]},
            "UserIdentityConfiguration":"identity attribute name"
        }]
    }

try:
    experience_response = kendra.create_experience(
        Name = name,
        Description = description,
        IndexId = index_id,
        RoleArn = role_arn,
        Configuration = configuration
    )

    pprint.pprint(experience_response)

    experience_endpoints = experience_response["Endpoints"]

    print("Wait for Amazon Kendra to create the experience.")
```

```
    while True:
        # Get the details of the experience, such as the status
        experience_description = kendra.describe_experience(
            Endpoints = experience_endpoints
        )
        status = experience_description["Status"]
        print(" Creating experience. Status: "+status)
        time.sleep(60)
        if status != "CREATING":
            break

except  ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

创建 Amazon Kendra

```
package com.amazonaws.kendra;

import java.util.concurrent.TimeUnit;
import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.CreateExperienceRequest;
import software.amazon.awssdk.services.kendra.model.CreateExperienceResponse;
import software.amazon.awssdk.services.kendra.model.DescribeExperienceRequest;
import software.amazon.awssdk.services.kendra.model.DescribeExperienceResponse;
import software.amazon.awssdk.services.kendra.model.ExperienceStatus;


public class CreateExperienceExample {

    public static void main(String[] args) throws InterruptedException {
        System.out.println("Create an experience");

        String experienceName = "experience-name";
        String experienceDescription = "experience description";
        String indexId = "index-id";
        String experienceRoleArn = "arn:aws:iam::account-id:role/role-name";

        KendraClient kendra = KendraClient.builder().build();
```

```
        CreateExperienceRequest createExperienceRequest = CreateExperienceRequest
            .builder()
            .name(experienceName)
            .description(experienceDescription)
            .roleArn(experienceRoleArn)
            .configuration(
                ExperienceConfiguration
                    .builder()
                    .contentSourceConfiguration(
                        ContentSourceConfiguration(
                            .builder()
                            .dataSourceIds("data-source-1","data-source-2")
                            .build()
                        )
                    )
                    .userIdentityConfiguration(
                        UserIdentityConfiguration(
                            .builder()
                            .identityAttributeName("identity-attribute-name")
                            .build()
                        )
                    ).build()
            ).build();

        CreateExperienceResponse createExperienceResponse =
kendra.createExperience(createExperienceRequest);
        System.out.println(String.format("Experience response %s",
createExperienceResponse));

        String experienceEndpoints = createExperienceResponse.endpoints();

        System.out.println(String.format("Wait for Kendra to create the
experience.", experienceEndpoints));
        while (true) {
            DescribeExperienceRequest describeExperienceRequest =
DescribeExperienceRequest.builder().endpoints(experienceEndpoints).build();
            DescribeExperienceResponse describeEpxerienceResponse =
kendra.describeExperience(describeExperienceRequest);
            ExperienceStatus status = describeExperienceResponse.status();
            TimeUnit.SECONDS.sleep(60);
            if (status != ExperienceStatus.CREATING) {
                break;
            }
        }
```

```
        System.out.println("Experience creation is complete.");
    }
}
```

调整容量

Amazon Kendra 以容量单位为索引提供资源。每个容量单位都为索引提供了额外的资源。文档和查询的存储有单独的容量单位。您只能向 Amazon Kendra 企业版索引添加容量单位。无法向开发者版本索引添加容量。

文档存储容量单位为您的索引提供了以下额外存储空间。

- 100,000 份文档或 30 GB 存储空间。

查询容量单位为您的索引提供以下额外查询。

- 每秒 0.1 次查询，或者每天大约 8000 次查询。

每个索引的基本容量等于 1 个容量单位（30 GB 存储空间和每秒 0.1 个查询）。每增加一个容量单位都需要支付额外费用。有关详细信息，请参阅 Amazon Kendra 定价。

您最多可以添加 100 个额外容量单位到存储中，并查询索引的资源。如果您需要更多设备，只需联系支持团队。

为了适应您的使用要求，您每天最多可以将容量单位提高 5 倍。您不能将文档存储容量减少到索引中存储的文档数量以下。例如，如果您要存储 150,000 个文档，则不能将存储容量减少到 1 个额外单位以下。

您可以在控制台中查看索引正在使用的资源，方法是选择索引名称以打开索引设置和其他信息，也可以使用 DescribeIndexAPI。

Amazon Kendra 当超过索引容量时，还会返回异常。当所有文档的提取总大小超过索引的限制时，您会得到 `ServiceQuotaExceededException`。当文档数量超过索引限制时，每个文档都会得到一个 `InvalidRequest`。如果每秒的查询数量超出限制，则会得到 `ThrottlingException`。有关限制的更多信息，请参阅 Amazon Kendra的限额。

累积的查询将持续长达 24 小时。

查看容量

通过选择索引名称来访问详细信息，即可在 Amazon Kendra 控制台中查看您的索引正在使用的资源。控制台还提供使用情况图表，因此您可以确定索引使用的存储空间和查询容量。您可以使用此信息来帮助您计划何时添加更多容量。

要查看文档存储和查询，请使用（控制台）

1. 登录 AWS Management Console 并打开 Amazon Kendra 控制台，网址为 https://console.aws.amazon.com/kendra/home。
2. 从索引列表中选择要访问的索引。
3. 滚动至设置部分，查看当前的文档存储总量和查询容量。

要使用 Amazon Kendra API 查看容量，请使用 DescribeIndexAPI 中的`CapacityUnits`参数。

添加和删除容量

如果您需要为索引增加容量，则可以使用控制台或 Amazon Kendra API 进行添加。

添加或删除存储或查询容量（控制台）

1. 登录 AWS Management Console 并打开 Amazon Kendra 控制台，网址为 https://console.aws.amazon.com/kendra/home。
2. 从索引列表中选择要访问的索引。
3. 选择编辑，或者从操作下拉列表中选择编辑。
4. 选择下一步以进入配置详细信息页面。
5. 添加或移除文档存储和/或查询容量单位。
6. 继续选择下一步以进入评论页面，然后选择更新以保存您的更改。

在更新索引的容量后，您所做的更改可能需要几分钟才能生效。

要使用 Amazon Kendra API 添加或移除容量，请使用 UpdateIndexAPI 中的`CapacityUnits`参数。

Amazon Kendra 智能排名容量

容量单位每秒为重新评分执行计划提供以下额外的重新评分请求。重新评分执行计划是用于配置重新评分 API 的资源。

- 每秒 0.01 个请求

每个重新评分执行计划都附带一个等于 1 个容量单位（每秒 0.01 个请求）的基本容量。每增加一个容量单位都需要支付额外费用。有关详细信息，请参阅 Amazon Kendra 定价。

您最多可以添加 1000 个额外容量单位来执行重新评分执行计划。如果您需要更多设备，只需联系支持团队。

查询建议容量

使用查询建议时，基本查询容量为每秒 2.5 次GetQuerySuggestions调用。GetQuerySuggestions 容量是索引预置查询容量的五倍，或每秒 2.5 个调用的基本容量，以较高者为准。例如，索引的基本容量为每秒 0.1 个查询，GetQuerySuggestions 容量的基本容量为每秒 2.5 个调用。如果您在索引每秒共 0.2 个查询的基础上再添加 0.1 个查询，GetQuerySuggestions 容量为每秒 2.5 个调用（比每秒 0.2 个查询的五倍还要高）。

Amazon Kendra 经验容量

搜索体验容量

Amazon Kendra 开始限制你的 Amazon Kendra 体验 QueryQuerySuggestions，SubmitFeedback每秒 15 个请求，每秒 40 个请求用于查询爆发。对于查询容量单位超过 150 的索引，这些限制仍然适用。

例如，您的索引的查询容量单位为 150，因此您的搜索体验应用程序每秒可以处理 15 个请求。但是，如果您扩展到 200 个查询容量单位，那么您的搜索体验应用程序每秒仍只能处理 15 个请求。如果您将索引限制为 100 个查询容量单位，则您的搜索体验应用程序每秒只能处理 10 个请求。

自适应查询暴增

Amazon Kendra 预配置的基本容量为 1 个查询容量单位。您每天使用 8000 个查询，最低吞吐量为每秒 0.1 个查询（每个查询容量单位）。累积的查询将持续长达 24 小时，并且可以容纳大量流量。允许的突发量会有所不同，因为它取决于集群在任何给定时间的负载。预配置足够的查询容量单位来处理您的峰值负载水平。

处理超出预配置吞吐量的意外突发流量的一种自适应方法是内置 Amazon Kendra的自适应查询爆发。Amazon Kendra的企业版提供了自适应查询突发功能。

自适应查询突发是一项内置功能，允许您应用未使用的查询容量来处理意外流量。 Amazon Kendra 以每秒预配置的查询速率累积未使用的查询，最多不超过您为索引预配置的最大查询数。 Amazon Kendra 这些累积的查询用于超出分配容量的意外流量。自适应查询突发的最佳性能可能会有所不同，具体取决于多个因素，例如索引总大小、查询复杂性、累积的未使用查询以及索引的总体负载。建议您自行进行负载测试，以准确测量容量暴增。

开始使用

本节介绍如何创建数据源以及如何将文档添加到 Amazon Kendra 索引。提供了 AWS 控制台、使用的 Python 程序和使用的 AWS SDK for Python (Boto3) Java 程序的说明 AWS SDK for Java。 AWS CLI

主题

先决条件

以下步骤是入门练习的先决条件。这些步骤向您展示了如何设置账户、创建 Amazon Kendra 允许代表您拨打电话的 IAM 角色以及如何为 Amazon S3 存储桶中的文档编制索引。以 S3 存储桶为例，但您可以使用其他 Amazon Kendra 支持的数据源。选择数据来源。

注册获取 AWS 账户

如果您没有 AWS 账户，请完成以下步骤来创建一个。

要注册 AWS 账户

1. 打开 https://portal.aws.amazon.com/billing/signup。
2. 按照屏幕上的说明进行操作。

   在注册时，将接到一通电话，要求使用电话键盘输入一个验证码。

   当您注册时 AWS 账户，就会创建AWS 账户根用户一个。根用户有权访问该账户中的所有 AWS 服务 和资源。作为安全最佳实践，请为用户分配管理访问权限，并且只使用根用户来执行需要根用户访问权限的任务。

AWS 注册过程完成后会向您发送一封确认电子邮件。在任何时候，您都可以通过转至 https://aws.amazon.com/ 并选择我的账户来查看当前的账户活动并管理您的账户。

创建具有管理访问权限的用户

注册后，请保护您的安全 AWS 账户 AWS 账户根用户 AWS IAM Identity Center，启用并创建管理用户，这样您就不会使用 root 用户执行日常任务。

保护你的 AWS 账户根用户

1. 选择 Root 用户并输入您的 AWS 账户 电子邮件地址，以账户所有者的身份登录。AWS Management Console在下一页上，输入您的密码。

   要获取使用根用户登录方面的帮助，请参阅《AWS 登录 用户指南》中的以根用户身份登录。
2. 为您的根用户启用多重身份验证 (MFA)。

   有关说明，请参阅 I A M 用户指南中的为 AWS 账户 根用户启用虚拟 MFA 设备（控制台）。

创建具有管理访问权限的用户

1. 启用 IAM Identity Center

   有关说明，请参阅《AWS IAM Identity Center 用户指南》中的启用 AWS IAM Identity Center。
2. 在 IAM Identity Center 中，为用户授予管理访问权限。

   有关使用 IAM Identity Center 目录 作为身份源的教程，请参阅《用户指南》 IAM Identity Center 目录中的使用默认设置配置AWS IAM Identity Center 用户访问权限。

以具有管理访问权限的用户身份登录

- 要使用您的 IAM Identity Center 用户身份登录，请使用您在创建 IAM Identity Center 用户时发送到您的电子邮件地址的登录网址。

  有关使用 IAM Identity Center 用户登录的帮助，请参阅AWS 登录 用户指南中的登录 AWS 访问门户。

将访问权限分配给其他用户

1. 在 IAM Identity Center 中，创建一个权限集，该权限集遵循应用最低权限的最佳做法。

有关说明，请参阅《AWS IAM Identity Center 用户指南》中的创建权限集。

2. 将用户分配到一个组，然后为该组分配单点登录访问权限。

   有关说明，请参阅《AWS IAM Identity Center 用户指南》中的添加组。

- 如果您使用的是包含要测试的文档的 S3 存储桶 Amazon Kendra，请在您使用的同一区域创建一个 S3 存储桶 Amazon Kendra。有关说明，请参阅 《Amazon Simple Storage Service 用户指南》中的创建和配置 S3 存储桶。

  将您的文档上载到 S3 存储桶。有关说明，请参阅《Amazon Simple Storage Service 用户指南》中的上传、下载和管理对象。

  如果您使用的是其他数据来源，则必须具有活动站点和凭证才能连接到该数据来源。

如果您使用控制台来开始使用，请从 Amazon Kendra 控制台入门 开始。

Amazon Kendra 资源： AWS CLI、SDK、控制台

如果您使用 CLI、开发工具包或控制台，则需要某些权限。

要用 Amazon Kendra 于 CLI、SDK 或控制台，您必须拥有 Amazon Kendra 允许代表您创建和管理资源的权限。根据您的用例，这些权限包括访问 Amazon Kendra API 本身（ AWS KMS keys 如果您想通过自定义 CMK 加密数据），如果您想与搜索体验集成 AWS IAM Identity Center 或创建搜索体验，则包括 Identity Center 目录。有关不同使用案例的完整权限列表，请参阅 IAM 角色。

首先，您必须将以下权限授予您的 IAM 用户。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "Stmt1644430853544",
      "Action": [
        "kms:CreateGrant",
        "kms:DescribeKey"
      ],
      "Effect": "Allow",
      "Resource": "*"
    },
    {
```

```
      "Sid": "Stmt1644430878150",
      "Action": "kendra:*",
      "Effect": "Allow",
      "Resource": "*"
    },
    {
      "Sid": "Stmt1644430973706",
      "Action": [
        "sso:AssociateProfile",
        "sso:CreateManagedApplicationInstance",
        "sso:DeleteManagedApplicationInstance",
        "sso:DisassociateProfile",
        "sso:GetManagedApplicationInstance",
        "sso:GetProfile",
        "sso:ListDirectoryAssociations",
        "sso:ListProfileAssociations",
        "sso:ListProfiles"
      ],
      "Effect": "Allow",
      "Resource": "*"
    },
    {
      "Sid": "Stmt1644430999558",
      "Action": [
        "sso-directory:DescribeGroup",
        "sso-directory:DescribeGroups",
        "sso-directory:DescribeUser",
        "sso-directory:DescribeUsers"
      ],
      "Effect": "Allow",
      "Resource": "*"
    },
    {
      "Sid": "Stmt1644431025960",
      "Action": [
        "identitystore:DescribeGroup",
        "identitystore:DescribeUser",
        "identitystore:ListGroups",
        "identitystore:ListUsers"
      ],
      "Effect": "Allow",
      "Resource": "*"
    }
  ]
```

```
}
```

其次，如果您使用 CLI 或 SDK，则还必须创建要访问的 IAM 角色和策略 Amazon CloudWatch Logs。如果您使用控制台，则无需为此创建 IAM 角色和策略。您可以将其作为控制台过程的一部分创建。

为和 SDK 创建允许 Amazon Kendra 访问您的 IAM 角色 AWS CLI 和策略 Amazon CloudWatch Logs。

1. 登录 AWS Management Console 并打开 IAM 控制台，网址为 https://console.aws.amazon.com/iam/。
2. 在左侧菜单中，选择策略，然后选择创建策略。
3. 选择 JSON，并将默认策略替换为以下策略：

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "cloudwatch:PutMetricData"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "cloudwatch:namespace": "AWS/Kendra"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:DescribeLogGroups"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogGroup"
            ],
            "Resource": [
                "arn:aws:logs:region:account ID:log-group:/aws/kendra/*"
```

```
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:DescribeLogStreams",
                "logs:CreateLogStream",
                "logs:PutLogEvents"
            ],
            "Resource": [
                "arn:aws:logs:region:account ID:log-group:/aws/kendra/*:log-
stream:*"
            ]
        }
    ]
}
```

4. 选择查看策略。
5. 将策略命名为 "KendraPolicyForGettingStartedIndex"，然后单击创建策略。
6. 在左侧菜单中，选择角色，然后选择创建角色。
7. 选择 “其他 AWS 账户”，然后在 “账户 ID” 中键入您的账户 ID。选择下一步：权限。
8. 选择您创建的策略，然后选择下一步：标签。
9. 请勿添加任何标签。选择 下一步: 审核。
10. 将角色命名为 "KendraRoleForGettingStartedIndex"，然后单击创建角色。
11. 找到您刚才创建的角色。选择角色名称以打开摘要。选择信任关系，然后选择编辑信任关系。
12. 将现有信任关系替换为内容：

```
{
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Principal": {
          "Service": "kendra.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
      }
    ]
}
```

13. 选择更新信任策略。

第三，如果您使用存储文档或使用 S3 进行测试 Amazon Kendra，则还必须创建 IAM 角色和策略才能访问您的存储桶。 Amazon S3 如果您正在使用其他数据来源，请参阅 数据来源的IAM 角色。

创建允许 Amazon Kendra 访问 Amazon S3 存储桶并为其编制索引的 IAM 角色和策略。

1. 登录 AWS Management Console 并打开 IAM 控制台，网址为 https://console.aws.amazon.com/iam/。
2. 在左侧菜单中，选择策略，然后选择创建策略。
3. 选择 JSON，并将默认策略替换为以下策略：

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::bucket name/*"
            ],
            "Effect": "Allow"
        },
        {
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::bucket name"
            ],
            "Effect": "Allow"
        },
        {
            "Effect": "Allow",
            "Action": [
                "kendra:BatchPutDocument",
                "kendra:BatchDeleteDocument"
            ],
            "Resource": "arn:aws:kendra:region:account ID:index/*"
        }
```

```
    ]
}
```

4. 选择查看策略。
5. 将策略命名为 KendraPolicyForGettingStartedDataSource “”，然后选择创建策略。
6. 在左侧菜单中，选择角色，然后选择创建角色。
7. 选择 “其他 AWS 账户”，然后在 “账户 ID” 中键入您的账户 ID。选择下一步：权限。
8. 选择您创建的策略，然后选择下一步：标签。
9. 请勿添加任何标签。选择 下一步: 审核。
10. 将角色命名为 “KendraRoleForGettingStartedDataSource”，然后选择 “创建角色”。
11. 找到您刚才创建的角色。选择角色名称以打开摘要。选择信任关系，然后选择编辑信任关系。
12. 将现有信任关系替换为内容：

```
{
    "Version": "2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Principal": {
          "Service": "kendra.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
      }
    ]
}
```

13. 选择更新信任策略。

根据您想要如何使用 Amazon Kendra API，请执行以下任一操作。

- 入门 (AWS CLI)
- 入门 (AWS SDK for Java)
- 入门 (AWS SDK for Python (Boto3))

Amazon Kendra 控制台入门

以下过程说明如何使用 AWS 控制台创建和测试 Amazon Kendra 索引。在这些过程中，您将为索引创建索引和数据来源。最后，您可以通过提出搜索请求来测试您的索引。

步骤 1：创建索引（控制台）

1. 登录 AWS 管理控制台并打开控制 Amazon Kendra 台，网址为 https://console.aws.amazon.com/kendra/。
2. 在索引部分选择创建索引。
3. 在指定索引详细信息页面上，为您的索引指定名称和描述。
4. 在 IAM 角色中，选择创建新角色，然后为角色提供一个名称。该 IAM 角色将带有前缀“AmazonKendra-”。
5. 将所有其他字段保留为默认值。选择下一步。
6. 在配置用户访问权限控制页面上选择下一步。
7. 在配置详细信息页面中，选择开发者版本。
8. 选择创建以创建索引。
9. 等待您的索引创建完成。 Amazon Kendra 为您的索引配置硬件。此操作可能需要一些时间。

步骤 2：为索引添加数据来源（控制台）

1. 查看可用的数据源以 Amazon Kendra 连接到您的文档并为其编制索引。
2. 在导航窗格中，选择数据来源，然后为所选数据来源选择添加数据来源。
3. 按照步骤配置数据来源。

步骤 3：搜索索引（控制台）

1. 在导航窗格中，选择用于搜索索引的选项。
2. 输入适合您的索引的搜索词。将显示排名靠前的结果和排名靠前的文档结果。

入门 (AWS CLI)

以下过程说明如何使用创建 Amazon Kendra 索引 AWS CLI。该过程会创建数据来源、索引，并对索引运行查询。

创建 Amazon Kendra 索引 (CLI)

1. 完成先决条件。
2. 输入以下命令来创建索引。

```
aws kendra create-index \
 --name cli-getting-started-index \
 --description "Index for CLI getting started guide." \
 --role-arn arn:aws:iam::account id:role/KendraRoleForGettingStartedIndex
```

3. 等待 Amazon Kendra 创建索引。使用以下命令检查进度。当状态字段为 ACTIVE 时，继续执行下一步。

```
aws kendra describe-index \
 --id index id
```

4. 在命令提示符下，输入以下命令以创建数据来源。

```
aws kendra create-data-source \
 --index-id index id \
 --name data source name \
 --role-arn arn:aws:iam::account id:role/KendraRoleForGettingStartedDataSource \
 --type S3 \
 --configuration '{"S3Configuration":{"BucketName":"S3 bucket name"}}'
```

如果您使用模板架构连接到数据来源，请配置模板架构。

```
aws kendra create-data-source \
 --index-id index id \
 --name data source name \
 --role-arn arn:aws:iam::account id:role/KendraRoleForGettingStartedDataSource \
 --type TEMPLATE \
 --configuration '{"TemplateConfiguration":{"Template":{JSON schema}}}'
```

5. 创建数据源需要 Amazon Kendra 一段时间。输入以下命令以检查进度。当状态为 ACTIVE 时，继续执行下一步。

```
aws kendra describe-data-source \
 --id data source ID \
 --index-id index ID
```

6. 输入以下命令以同步数据来源。

```
aws kendra start-data-source-sync-job \
 --id data source ID \
 --index-id index ID
```

7. Amazon Kendra 将为您的数据源编制索引。所需时间取决于文档的数量。您可以使用以下命令检查同步作业的状态。当状态为 ACTIVE 时，继续执行下一步。

```
aws kendra describe-data-source \
 --id data source ID \
 --index-id index ID
```

8. 输入以下命令以执行查询。

```
aws kendra query \
 --index-id index ID \
 --query-text "search term"
```

搜索的结果将以 JSON 格式显示。

入门 (AWS SDK for Python (Boto3))

以下程序是在 Python 程序 Amazon Kendra 中使用的示例。此程序执行以下操作：

1. 使用CreateIndex操作创建新索引。
2. 等待索引创建完成。它使用该DescribeIndex操作来监视索引的状态。
3. 一旦索引处于活动状态，它就会使用CreateDataSource操作创建数据源。
4. 等待数据来源创建完成。它使用该DescribeDataSource操作来监视数据源的状态。
5. 当数据源处于活动状态时，它会使用StartDataSourceSyncJob操作将索引与数据源的内容同步。

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")
```

```
print("Create an index.")

# Provide a name for the index
index_name = "python-getting-started-index"
# Provide an optional decription for the index
description = "Getting started index"
# Provide the IAM role ARN required for indexes
index_role_arn = "arn:aws:iam::${accountId}:role/KendraRoleForGettingStartedIndex"

try:
    index_response = kendra.create_index(
        Description = description,
        Name = index_name,
        RoleArn = index_role_arn
    )

    pprint.pprint(index_response)

    index_id = index_response["Id"]

    print("Wait for Amazon Kendra to create the index.")

    while True:
        # Get the details of the index, such as the status
        index_description = kendra.describe_index(
            Id = index_id
        )
        # When status is not CREATING quit.
        status = index_description["Status"]
        print(" Creating index. Status: "+status)
        time.sleep(60)
        if status != "CREATING":
            break

    print("Create an S3 data source.")

    # Provide a name for the data source
    data_source_name = "python-getting-started-data-source"
    # Provide an optional description for the data source
    data_source_description = "Getting started data source."
    # Provide the IAM role ARN required for data sources
    data_source_role_arn = "arn:aws:iam::${accountId}:role/
KendraRoleForGettingStartedDataSource"
    # Provide the data source connection information
```

```
    S3_bucket_name = "S3-bucket-name"
    data_source_type = "S3"
    # Configure the data source
    configuration = {"S3Configuration":
        {
            "BucketName": S3_bucket_name
        }
    }

    """
    If you connect to your data source using a template schema,
    configure the template schema
    configuration = {"TemplateConfiguration":
        {
            "Template": {JSON schema}
        }
    }
    """

    data_source_response = kendra.create_data_source(
        Name = data_source_name,
        Description = data_source_name,
        RoleArn = data_source_role_arn,
        Type = data_source_type,
        Configuration = configuration,
        IndexId = index_id
    )

    pprint.pprint(data_source_response)

    data_source_id = data_source_response["Id"]

    print("Wait for Amazon Kendra to create the data source.")

    while True:
        # Get the details of the data source, such as the status
        data_source_description = kendra.describe_data_source(
            Id = data_source_id,
            IndexId = index_id
        )
        # If status is not CREATING, then quit
        status = data_source_description["Status"]
        print(" Creating data source. Status: "+status)
        time.sleep(60)
```

```
        if status != "CREATING":
            break

    print("Synchronize the data source.")

    sync_response = kendra.start_data_source_sync_job(
        Id = data_source_id,
        IndexId = index_id
    )

    pprint.pprint(sync_response)

    print("Wait for the data source to sync with the index.")

    while True:

        jobs = kendra.list_data_source_sync_jobs(
            Id = data_source_id,
            IndexId = index_id
        )

        # For this example, there should be one job
        status = jobs["History"][0]["Status"]

        print(" Syncing data source. Status: "+status)
        if status != "SYNCING":
            break
        time.sleep(60)

except  ClientError as e:
        print("%s" % e)

print("Program ends.")
```

入门 (AWS SDK for Java)

以下程序是在 Java 程序 Amazon Kendra 中使用的示例。此程序执行以下操作：

1. 使用CreateIndex操作创建新索引。
2. 等待索引创建完成。它使用该DescribeIndex操作来监视索引的状态。
3. 一旦索引处于活动状态，它就会使用CreateDataSource操作创建数据源。

4. 等待数据来源创建完成。它使用该DescribeDataSource操作来监视数据源的状态。

5. 当数据源处于活动状态时，它会使用StartDataSourceSyncJob操作将索引与数据源的内容同步。

```
package com.amazonaws.kendra;

import java.util.concurrent.TimeUnit;
import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.CreateDataSourceRequest;
import software.amazon.awssdk.services.kendra.model.CreateDataSourceResponse;
import software.amazon.awssdk.services.kendra.model.CreateIndexRequest;
import software.amazon.awssdk.services.kendra.model.CreateIndexResponse;
import software.amazon.awssdk.services.kendra.model.DataSourceConfiguration;
import software.amazon.awssdk.services.kendra.model.DataSourceStatus;
import software.amazon.awssdk.services.kendra.model.DataSourceSyncJob;
import software.amazon.awssdk.services.kendra.model.DataSourceSyncJobStatus;
import software.amazon.awssdk.services.kendra.model.DataSourceType;
import software.amazon.awssdk.services.kendra.model.DescribeDataSourceRequest;
import software.amazon.awssdk.services.kendra.model.DescribeDataSourceResponse;
import software.amazon.awssdk.services.kendra.model.DescribeIndexRequest;
import software.amazon.awssdk.services.kendra.model.DescribeIndexResponse;
import software.amazon.awssdk.services.kendra.model.IndexStatus;
import software.amazon.awssdk.services.kendra.model.ListDataSourceSyncJobsRequest;
import software.amazon.awssdk.services.kendra.model.ListDataSourceSyncJobsResponse;
import software.amazon.awssdk.services.kendra.model.S3DataSourceConfiguration;
import software.amazon.awssdk.services.kendra.model.StartDataSourceSyncJobRequest;
import software.amazon.awssdk.services.kendra.model.StartDataSourceSyncJobResponse;


public class CreateIndexAndDataSourceExample {

    public static void main(String[] args) throws InterruptedException {
        System.out.println("Create an index");

        String indexDescription = "Getting started index for Kendra";
        String indexName = "java-getting-started-index";
        String indexRoleArn = "arn:aws:iam::<your AWS account ID>:role/<name of an IAM
 role>";

        System.out.println(String.format("Creating an index named %s", indexName));
        KendraClient kendra = KendraClient.builder().build();

        CreateIndexRequest createIndexRequest = CreateIndexRequest
```

```
            .builder()
            .description(indexDescription)
            .name(indexName)
            .roleArn(indexRoleArn)
            .build();
        CreateIndexResponse createIndexResponse =
kendra.createIndex(createIndexRequest);
        System.out.println(String.format("Index response %s", createIndexResponse));

        String indexId = createIndexResponse.id();

        System.out.println(String.format("Waiting until the index with index ID %s is
created", indexId));
        while (true) {
            DescribeIndexRequest describeIndexRequest =
DescribeIndexRequest.builder().id(indexId).build();
            DescribeIndexResponse describeIndexResponse =
kendra.describeIndex(describeIndexRequest);
            IndexStatus status = describeIndexResponse.status();
            if (status != IndexStatus.CREATING) {
                break;
            }

            TimeUnit.SECONDS.sleep(60);
        }

        System.out.println("Creating an S3 data source");
        String dataSourceName = "java-getting-started-data-source";
        String dataSourceDescription = "Getting started data source";
        String s3BucketName = "an-aws-kendra-test-bucket";
        String dataSourceRoleArn = "arn:aws:iam::<your AWS account ID>:role/<name of an
IAM role>";

        CreateDataSourceRequest createDataSourceRequest = CreateDataSourceRequest
            .builder()
            .indexId(indexId)
            .name(dataSourceName)
            .description(dataSourceDescription)
            .roleArn(dataSourceRoleArn)
            .type(DataSourceType.S3)
            .configuration(
                DataSourceConfiguration
                    .builder()
                    .s3Configuration(
```

```
                        S3DataSourceConfiguration
                            .builder()
                            .bucketName(s3BucketName)
                            .build()
                    ).build()
            ).build();

        CreateDataSourceResponse createDataSourceResponse =
kendra.createDataSource(createDataSourceRequest);
        System.out.println(String.format("Response of creating data source: %s",
createDataSourceResponse));

        String dataSourceId = createDataSourceResponse.id();
        System.out.println(String.format("Waiting for Kendra to create the data source
%s", dataSourceId));
        DescribeDataSourceRequest describeDataSourceRequest = DescribeDataSourceRequest
            .builder()
            .indexId(indexId)
            .id(dataSourceId)
            .build();

        while (true) {
            DescribeDataSourceResponse describeDataSourceResponse =
kendra.describeDataSource(describeDataSourceRequest);

            DataSourceStatus status = describeDataSourceResponse.status();
            System.out.println(String.format("Creating data source. Status: %s",
status));
            if (status != DataSourceStatus.CREATING) {
                break;
            }

            TimeUnit.SECONDS.sleep(60);
        }

        System.out.println(String.format("Synchronize the data source %s",
dataSourceId));
        StartDataSourceSyncJobRequest startDataSourceSyncJobRequest =
StartDataSourceSyncJobRequest
            .builder()
            .indexId(indexId)
            .id(dataSourceId)
            .build();
```

```
        StartDataSourceSyncJobResponse startDataSourceSyncJobResponse =
 kendra.startDataSourceSyncJob(startDataSourceSyncJobRequest);
        System.out.println(String.format("Waiting for the data
 source to sync with the index %s for execution ID %s", indexId,
 startDataSourceSyncJobResponse.executionId()));

        // For this particular list, there should be just one job
        ListDataSourceSyncJobsRequest listDataSourceSyncJobsRequest =
 ListDataSourceSyncJobsRequest
            .builder()
            .indexId(indexId)
            .id(dataSourceId)
            .build();

        while (true) {
            ListDataSourceSyncJobsResponse listDataSourceSyncJobsResponse =
 kendra.listDataSourceSyncJobs(listDataSourceSyncJobsRequest);
            DataSourceSyncJob job = listDataSourceSyncJobsResponse.history().get(0);
            System.out.println(String.format("Syncing data source. Status: %s",
 job.status()));

            if (job.status() != DataSourceSyncJobStatus.SYNCING) {
                break;
            }

            TimeUnit.SECONDS.sleep(60);

        }

        System.out.println("Index setup is complete");
    }
}
```

Amazon S3 数据来源入门（控制台）

您可以使用 Amazon Kendra 控制台来开始使用 Amazon S3 存储桶作为数据存储。使用该控制台时，您可以指定为存储桶的内容编制索引所需的连接信息。有关更多信息，请参阅Amazon S3。

使用默认配置，按照以下过程创建基本 S3 存储桶数据来源。此过程假定您已按照 Amazon Kendra 控制台入门 中的步骤 1 创建了索引。

使用 Amazon Kendra 控制台创建 S3 存储桶

1. 登录 AWS Management Console 并打开 Amazon Kendra 控制台 (https://console.aws.amazon.com/kendra/home)。
2. 从索引列表中，选择要向其添加数据来源的索引。
3. 选择添加数据来源。
4. 从数据来源连接器列表中选择 Amazon S3。
5. 在定义属性页面上，为您的数据来源指定一个名称和一个可选的描述。将标签字段留空。选择 Next（下一步）以继续。
6. 在输入数据来源位置字段中，输入包含您的文档的 S3 存储桶的名称。您可以直接输入名称，也可以通过选择浏览来指定该名称。该存储桶必须与索引位于同一区域。
7. 在 IAM 角色中，选择创建新角色，然后输入一个角色名称。有关更多信息，请参阅 Amazon S3 S3 数据来源的 IAM 角色。
8. 在设置同步运行计划部分，选择按需运行。
9. 选择 Next（下一步）以继续。
10. 在审核和创建页面上，查看您的 S3 数据来源的详细信息。如果要进行更改，请选择要更改的项目旁的编辑按钮。如果您对自己的选择感到满意，请选择创建以创建您的 S3 数据来源。

选择创建后，Amazon Kendra 会开始创建数据来源。创建数据来源可能需要几分钟时间。完成后，数据来源的状态将从正在创建变为活动。

创建数据来源后，您需要将 Amazon Kendra 索引与数据来源同步。选择立即同步以开始同步过程。同步数据来源可能需要几分钟到几小时，具体取决于文档的数量和大小。

MySQL 数据库数据来源入门（控制台）

通过 Amazon Kendra 控制台，您可以开始使用 MySQL 数据库作为数据来源。使用该控制台时，您可以指定为 MySQL 数据库的内容建立索引所需的连接信息。有关更多信息，请参阅使用数据库数据来源。

首先，您需要创建 MySQL 数据库，然后才能为该数据库创建数据来源。

按照以下过程创建 MySQL 的基本数据库。此过程假定您已按照Amazon Kendra 控制台入门中的步骤 1 创建了索引。

创建 MySQL 数据库

1. 登录 AWS Management Console 并通过以下网址打开 Amazon RDS 控制台：https://console.aws.amazon.com/rds/。
2. 从导航窗格中，选择子网组，然后选择创建数据库子网组。
3. 为子网组命名，然后选择 虚拟私有云 (VPC)。有关配置 VPC 的更多信息，请参阅配置 Amazon Kendra 以使用 VPC。
4. 添加 VPC 的私有子网。您的私有子网是未连接到 NAT 的子网。选择 Create（创建）。
5. 在导航窗格中，选择数据库，然后选择创建数据库。
6. 使用以下参数创建数据库。将所有其他参数保留为默认值。
   - 引擎选项 - MySQL
   - 模板 - 免费套餐
   - 凭证设置 - 输入并确认密码
   - 在连接下，选择其他连接配置。进行以下选择：
     - 子网组 - 选择您在步骤 4 中创建的子网组。
     - VPC 安全组 - 选择包含您在 VPC 中创建的入站和出站规则的组。例如，**`DataSourceSecurityGroup`**。有关配置 VPC 的更多信息，请参阅配置 Amazon Kendra 以使用 VPC。
   - 在其他配置下，将初始数据库名称设置为 **`content`**。
7. 选择 Create database（创建数据库）。
8. 从数据集列表中选择您的新数据库。记录数据库端点。
9. 创建数据库后，您必须创建一个用于保存文档的表。创建表不在这些说明的讨论范围之内。创建表时，请注意以下几点：
   - 数据库名称 - **`content`**
   - 表名称 - **`documents`**
   - 列 - **`ID`**、**`Title`**、**`Body`** 和 **`LastUpdate`**。如果需要，可以添加其他列。

现在，您已创建 MySQL 数据库，可以为该数据库创建一个数据来源。

创建 MySQL 数据来源

1. 登录 AWS Management Console并打开 Amazon Kendra 控制台 (https://console.aws.amazon.com/kendra/home)。

2. 从导航窗格中选择索引，然后选择您的索引。
3. 选择添加数据来源，然后选择 Amazon RDS。
4. 键入数据来源的名称和描述，然后选择下一步。
5. 选择 MySQL。
6. 在连接访问权限下，输入以下信息：
   - 端点 - 您之前创建的数据库的端点。
   - 端口 - 数据库的端口号。MySQL 的默认端口号是 3306。
   - 身份验证类型 - 选择新建。
   - 新的密钥容器名称 - 数据库凭证的 Secrets Manager 容器的名称。
   - 用户名 - 对数据库具有管理权限的用户的名称。
   - 密码 - 用户的密码，然后选择保存身份验证。
   - 数据库名称 - **`content`**。
   - 表名称 - **`documents`**。
   - IAM 角色 - 选择创建新角色，然后输入该角色的名称。
7. 在列配置中，输入：
   - 文档 ID 列名称 - **`ID`**
   - 文档标题列名称 - **`Title`**
   - 文档数据列名称 - **`Body`**
8. 在列更改检测中，输入：
   - 更改检测列 - **`LastUpdate`**
9. 在配置 VPC 和安全组中，提供：
   - 在虚拟私有云 (VPC) 中，选择您的 VPC。
   - 在子网中，选择您在 VPC 中创建的私有子网。
   - 在 VPC 安全组中 - 选择包含您在 VPC 中创建的入站和出站规则的安全组。例如，**`DataSourceSecurityGroup`**。
10. 在设置同步运行计划中，选择按需运行，然后选择下一步。
11. 在数据来源字段映射中，选择下一步。
12. 检查数据来源的配置，确保正确无误。如果您认为所有设置都正确，请选择创建。

AWS IAM Identity Center 身份源入门（控制台）

AWS IAM Identity Center 身份源包含有关您的用户和群组的信息。这对于设置用户上下文筛选非常有用，在这种 Amazon Kendra 筛选中，根据用户或其群组对文档的访问权限筛选不同用户的搜索结果。

要创建 IAM Identity Center 身份源，您必须激活 IAM Identity Center 并在 AWS Organizations 中创建一个组织。当您首次激活 IAM Identity Center 并创建组织时，它会自动默认将 Identity Center 目录作为身份源。您可以将其更改为 Active Directory（Amazon 托管或自管理）或外部身份提供商，以作为您的身份来源。您必须遵循正确的指导，请参阅更改 IAM Identity Center 身份来源。每个组织只能有一个身份源。

为了向用户和组分配不同级别的文档访问权限，在将文档提取到索引中时，您需要将您的用户和组包含在访问控制列表中。这允许您的用户和群组根据其访问权限级别搜索文档。 Amazon Kendra 当您发出查询时，用户 ID 必须与 IAM Identity Center 中的用户名完全匹配。

您还必须授予使用 IAM 身份中心所需的权限 Amazon Kendra。有关更多信息，请参阅IAM Identity Center 的 IAM 角色。

设置 IAM Identity Center 身份源

1. 打开 IAM Identity Center 控制台。
2. 选择启用 IAM 身份中心，然后选择创建 AWS 组织。

   默认情况下会创建 Identity Center 目录，并向您发送一封电子邮件以验证与组织关联的电子邮件地址。
3. 要将群组添加到您的 AWS 组织，请在导航窗格中选择 “群组”。
4. 在组页面上，选择创建组并在对话框中输入组的名称和描述。选择 创建。
5. 要向组织添加用户，请在导航窗格中选择用户。
6. 在 Users (用户) 页面上，选择 Add user (添加用户)。在用户详细信息下，指定所有必填字段。对于密码，选择向用户发送电子邮件。选择下一步。
7. 要将用户添加到组，请选择组并选择一个组。
8. 在详细信息页面上的组成员下，选择添加用户。
9. 在向组添加用户页面上，选择要添加为组成员的用户。您可以选择多个用户以添加到组中。
10. 要将您的用户和组列表与 IAM Identity Center 同步，请将您的身份源更改为 Active Directory 或外部身份提供商。

Identity Center 目录是默认的身份源，如果您没有让提供商管理自己的列表，则需要您使用此来源手动添加用户和组。要更改身份源，您必须按照正确的指导操作，请参阅更改 IAM Identity Center 身份源。

> **Note**
>
> 如果使用 Active Directory 或外部身份提供商作为身份源，则在指定跨域身份管理系统 (SCIM) 协议时，必须将用户的电子邮件地址映射到 IAM Identity Center 用户名。有关更多信息，请参阅 SCIM 上有关启用 IAM Identity Center 的《IAM Identity Center 指南》。

设置 IAM Identity Center 身份源后，您可以在创建或编辑索引时在控制台中将其激活。转到索引设置中的用户访问控制并编辑您的设置，以允许从 IAM Identity Center 获取用户组信息。

您也可以使用UserGroupResolutionConfiguration对象激活 IAM 身份中心。您提供 `UserGroupResolutionMode` as `AWS_SSO` 并创建一个 IAM 角色来授予调用`sso:ListDirectoryAssociations`、、`sso-directory:SearchUsers`sso-directory:ListGroupsForUser、的权限`sso-directory:DescribeGroups`。

> **Warning**
>
> Amazon Kendra 目前不支持使用`UserGroupResolutionConfiguration` AWS 组织成员账户作为您的 IAM Identity Center 身份源。您必须在组织的管理账户中创建索引才能使用 `UserGroupResolutionConfiguration`。

下面概述了如何设置数据来源 `UserGroupResolutionConfiguration` 和用户访问控制来根据用户上下文筛选搜索结果。这假设您已经为索引创建了索引和 IAM 角色。您可以使用 CreateIndexAPI 创建索引并提供 IAM 角色。

使用 **`UserGroupResolutionConfiguration`** 和用户上下文筛选设置数据来源

1. 创建授予访问 IAM Identity Center 身份源的权限的 IAM 角色。
2. `UserGroupResolutionConfiguration`通过将模式设置为进行配置，`AWS_SSO`然后调用更新您的索引UpdateIndex以使用 IAM Identity Center。
3. 如果要使用基于令牌的用户访问控制根据用户上下文筛选搜索结果，请UserContextPolicy将其设置为呼叫`USER_TOKEN`时。`UpdateIndex`否则， Amazon Kendra 会搜索大多数数据源连接器的

每个文档的访问控制列表。您还可以通过在 UserContext 中提供用户和组信息，在查询 API 中根据用户上下文筛选搜索结果。您也可以使用将用户映射到他们的群组，PutPrincipalMapping这样您只需在发出查询时提供用户 ID。

4. 创建一个 IAM 角色，以授予访问数据来源的权限。
5. 配置数据来源。您必须提供所需的连接信息以连接到您的数据来源。
6. 使用 CreateDataSourceAPI 创建数据源。提供 DataSourceConfiguration 对象，包括索引的 ID TemplateConfiguration、数据来源的 IAM 角色、数据来源类型，并为您的数据来源命名。您也可以更新您的数据来源。

更改 IAM Identer Identer 身份源

> **Warning**
>
> 在 IAM Identity Center 设置中更改您的身份源可能会影响用户和组信息的保留。为了安全地执行此操作，建议查看更改身份源的注意事项。当您更改身份源时，会生成一个新的身份源 ID。在将模式设置为 in 之前，请检查您使用的 ID 是否正确UserGroupResolutionConfiguration。AWS_SSO

更改 IAM Identer 身份源

1. 打开 IAM Identity Center > 控制台。
2. 选择设置。
3. 在设置页面的身份源下，选择更改。
4. 在更改身份源页面上，选择您的首选身份源，然后选择下一步。

创建索引

您可以使用控制台或通过调用 CreateIndexAPI 来创建索引。您可以将 AWS Command Line Interface (AWS CLI) 或 SDK 与 API 配合使用。创建索引后，您可以直接向索引中添加文档，也可以从数据来源添加文档。

要创建索引，您必须提供 () 角色的 Amazon 资源名称 AWS Identity and Access Management (ARNIAM) 以供索引访问。 CloudWatch有关更多信息，请参阅索引的 IAM 角色。

以下选项卡提供了使用创建索引的过程，以及使用 AWS Management Console、Python 和 Java 开发工具包的代码示例。 AWS CLI

Console

创建索引

1. 登录 AWS 管理控制台并打开控制 Amazon Kendra 台，网址为 https://console.aws.amazon.com/kendra/。
2. 在索引部分选择创建索引。
3. 在指定索引详细信息中，指定索引都名称和描述。
4. 在IAM 角色中提供一个 IAM 角色。要查找角色，请在您的账户中选择包含“kendra”一词的角色，或者输入其他角色的名称。有关该角色所需权限的更多信息，请参阅索引的 IAM 角色。
5. 选择下一步。
6. 在配置用户访问权限控制页面上选择下一步。创建索引后，您可以更新索引以使用令牌进行访问权限控制。有关更多信息，请参阅控制对文档的访问权限。
7. 在预配置详细信息页面上选择创建。
8. 索引可能需要一些时间才能创建完成。查看索引列表以了解索引创建进度。当索引的状态为 `ACTIVE` 时，您的索引就已经准备就绪。

AWS CLI

创建索引

1. 使用以下命令创建索引。`role-arn`必须是可以运行 Amazon Kendra 操作的 IAM 角色的 Amazon 资源名称 (ARN)。有关更多信息，请参阅 IAM 角色。

该命令针对 Linux 和 macOS 编排了格式。如果您使用 Windows，请将 Unix 行继续符 (\) 替换为脱字号 (^)。

```
aws kendra create-index \
 --name index name \
 --description "index description" \
 --role-arn arn:aws:iam::account ID:role/role name
```

2. 索引可能需要一些时间才能创建完成。要检查索引的状态，请在以下命令中使用 `create-index` 返回的索引 ID。当索引的状态为 `ACTIVE` 时，您的索引就已经准备就绪。

```
aws kendra describe-index \
 --index-id index ID
```

Python

创建索引

- 在下面的代码示例中为以下变量提供值：
  - `description` - 正在创建的索引的描述。该项为可选项。
  - `index_name` - 正在创建的索引的名称。
  - `role_arn`— 可以运行 Amazon Kendra API 的角色的亚马逊资源名称 (ARN)。有关更多信息，请参阅 IAM 角色。

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Create an index.")

# Provide a name for the index
index_name = "index-name"
# Provide an optional description for the index
description = "index description"
```

```
# Provide the IAM role ARN required for indexes
role_arn = "arn:aws:iam::${account id}:role/${role name}"

try:
    index_response = kendra.create_index(
        Name = index_name,
        Description = description,
        RoleArn = role_arn
    )

    pprint.pprint(index_response)

    index_id = index_response["Id"]

    print("Wait for Amazon Kendra to create the index.")

    while True:
        # Get the details of the index, such as the status
        index_description = kendra.describe_index(
            Id = index_id
        )
        # If status is not CREATING, then quit
        status = index_description["Status"]
        print(" Creating index. Status: "+status)
        if status != "CREATING":
            break
        time.sleep(60)

except  ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

创建索引

- 在下面的代码示例中为以下变量提供值：
  - `description` - 正在创建的索引的描述。该项为可选项。
  - `index_name` - 正在创建的索引的名称。

- `role_arn`— 可以运行 Amazon Kendra API 的角色的亚马逊资源名称 (ARN)。有关更多信息，请参阅 IAM 角色。

```
package com.amazonaws.kendra;

import java.util.concurrent.TimeUnit;
import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.CreateIndexRequest;
import software.amazon.awssdk.services.kendra.model.CreateIndexResponse;
import software.amazon.awssdk.services.kendra.model.DescribeIndexRequest;
import software.amazon.awssdk.services.kendra.model.DescribeIndexResponse;
import software.amazon.awssdk.services.kendra.model.IndexStatus;


public class CreateIndexExample {

    public static void main(String[] args) throws InterruptedException {

        String indexDescription = "Getting started index for Kendra";
        String indexName = "java-getting-started-index";
        String indexRoleArn = "arn:aws:iam::<your AWS account ID>:role/
KendraRoleForGettingStartedIndex";

        System.out.println(String.format("Creating an index named %s",
 indexName));
        CreateIndexRequest createIndexRequest = CreateIndexRequest
            .builder()
            .description(indexDescription)
            .name(indexName)
            .roleArn(indexRoleArn)
            .build();
        KendraClient kendra = KendraClient.builder().build();
        CreateIndexResponse createIndexResponse =
 kendra.createIndex(createIndexRequest);
        System.out.println(String.format("Index response %s",
 createIndexResponse));

        String indexId = createIndexResponse.id();

        System.out.println(String.format("Waiting until the index with ID %s is
 created.", indexId));
        while (true) {
```

```
            DescribeIndexRequest describeIndexRequest =
 DescribeIndexRequest.builder().id(indexId).build();
            DescribeIndexResponse describeIndexResponse =
 kendra.describeIndex(describeIndexRequest);
            IndexStatus status = describeIndexResponse.status();
            if (status != IndexStatus.CREATING) {
                break;
            }

            TimeUnit.SECONDS.sleep(60);
        }

        System.out.println("Index creation is complete.");
    }
}
```

创建索引后，您可以向其中添加文档。您可以直接添加，也可以创建定期更新索引的数据来源。

主题

- 通过批量上传将文档直接添加到索引中
- 将常见问题解答 (FAQ) 添加到索引中
- 创建自定义文档字段
- 使用令牌控制用户访问文档

通过批量上传将文档直接添加到索引中

您可以使用 BatchPutDocumentAPI 将文档直接添加到索引。您无法使用控制台直接添加文档。如果使用控制台，则可以连接到数据来源，以便向索引中添加文档。您可以从 S3 存储桶添加文档，也可以将文档作为二进制数据提供。有关支持的文档类型的列表， Amazon Kendra 请参阅文档类型。

使用 BatchPutDocument 将文档添加到索引中是一种异步操作。调用 BatchPutDocument API 后，您可以使用 BatchGetDocumentStatusAPI 来监控为文档编制索引的进度。当您使用文档 ID 列表调用 BatchGetDocumentStatus API 时，它会返回文档的状态。当文档状态为 INDEXED 或 FAILED 时，表明文档处理已完成。当状态为 FAILED 时，BatchGetDocumentStatus API 会返回无法为文档编制索引的原因。

如果您想在文档提取过程中更改内容和文档元数据字段或属性，请参阅 Amazon Kendra 自定义文档富集。如果要使用自定义数据来源，则使用 BatchPutDocument API 提交的每个文档都需要将数据来源 ID 和执行 ID 作为属性或字段。有关更多信息，请参阅自定义数据来源的必需属性。

> **Note**
>
> 每个索引的每个文档 ID 必须是唯一的。在创建数据来源时，您不能使用文档的唯一 ID 来编制索引，然后使用 BatchPutDocument API 为相同的文档编制索引，反之亦然。您可以删除数据来源，然后使用 BatchPutDocument API 为相同的文档编制索引，反之亦然。对于同一组文档，将 BatchPutDocument 和 BatchDeleteDocument API 与 Amazon Kendra 数据来源连接器结合使用可能会导致数据不一致。我们建议使用 Amazon Kendra 自定义数据来源连接器。

以下开发人员指南文档展示了如何将文档直接添加到索引中。

主题

使用 BatchPutDocument API 添加文档

以下示例通过调用BatchPutDocument将文本块添加到索引中。您可以使用 BatchPutDocument API 将文档直接添加到索引中。有关支持的文档类型的列表， Amazon Kendra 请参阅文档类型。

有关使用 AWS CLI 和软件开发工具包创建索引的示例，请参阅创建索引。要设置 CLI 和开发工具包，请参阅设置 Amazon Kendra。

> **Note**
>
> 添加到索引中的文件必须采用 UTF-8 编码的字节流。

在以下示例中，已将 UTF-8 编码的文本添加到索引中。

CLI

在中 AWS Command Line Interface，使用以下命令。该命令针对 Linux 和 macOS 编排了格式。如果您使用 Windows，请将 Unix 行继续符 (\) 替换为脱字号 (^)。

```
aws kendra batch-put-document \
   --index-id index-id \
   --documents '{"Id":"doc-id-1", "Blob":"Amazon.com is an online retailer.",
 "ContentType":"PLAIN_TEXT", "Title":"Information about Amazon.com"}'
```

Python

```
import boto3

kendra = boto3.client("kendra")

# Provide the index ID
index_id = "index-id"

# Provide the title and text
title = "Information about Amazon.com"
text = "Amazon.com is an online retailer."

document = {
    "Id": "1",
    "Blob": text,
    "ContentType": "PLAIN_TEXT",
    "Title": title
}

documents = [
    document
]

result = kendra.batch_put_document(
    IndexId = index_id,
    Documents = documents
)

print(result)
```

Java

```
package com.amazonaws.kendra;


import software.amazon.awssdk.core.SdkBytes;
import software.amazon.awssdk.services.kendra.KendraClient;
```

```
import software.amazon.awssdk.services.kendra.model.BatchPutDocumentRequest;
import software.amazon.awssdk.services.kendra.model.BatchPutDocumentResponse;
import software.amazon.awssdk.services.kendra.model.ContentType;
import software.amazon.awssdk.services.kendra.model.Document;

public class AddDocumentsViaAPIExample {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();

        String indexId = "yourIndexId";

        Document testDoc = Document
            .builder()
            .title("The title of your document")
            .id("a_doc_id")
            .blob(SdkBytes.fromUtf8String("your text content"))
            .contentType(ContentType.PLAIN_TEXT)
            .build();

        BatchPutDocumentRequest batchPutDocumentRequest = BatchPutDocumentRequest
            .builder()
            .indexId(indexId)
            .documents(testDoc)
            .build();

        BatchPutDocumentResponse result =
 kendra.batchPutDocument(batchPutDocumentRequest);

        System.out.println(String.format("BatchPutDocument Result: %s", result));
    }
}
```

从 S3 存储桶添加文档

您可以使用 BatchPutDocumentAPI 将 Amazon S3 存储桶中的文档直接添加到索引中。在同一次调用中最多可以添加 10 个文档。使用 S3 存储桶时，必须为 IAM 角色提供访问包含您的文档的存储桶的权限。该角色在 `RoleArn` 参数中指定。

使用 BatchPutDocumentAPI 从 Amazon S3 存储桶添加文档是一次性操作。要使索引与存储桶的内容保持同步，请创建 Amazon S3 数据源。有关更多信息，请参阅 Amazon S3 数据来源。

有关使用 AWS CLI 和软件开发工具包创建索引的示例，请参阅创建索引。要设置 CLI 和开发工具包，请参阅设置 Amazon Kendra。有关创建 S3 存储桶的信息，请参阅 Amazon Simple Storage Service 文档。

在以下示例中，使用 `BatchPutDocument` API 将两个 Microsoft Word 文档添加到了索引中。

Python

```
import boto3

kendra = boto3.client("kendra")

# Provide the index ID
index_id = "index-id"
# Provide the IAM role ARN required to index documents in an S3 bucket
role_arn = "arn:aws:iam::${acccountID}:policy/${roleName}"

doc1_s3_file_data = {
    "Bucket": "bucket-name",
    "Key": "document1.docx"
}

doc1_document = {
    "S3Path": doc1_s3_file_data,
    "Title": "Document 1 title",
    "Id": "doc_1"
}

doc2_s3_file_data = {
    "Bucket": "bucket-name",
    "Key": "document2.docx"
}

doc2_document = {
    "S3Path": doc2_s3_file_data,
    "Title": "Document 2 title",
    "Id": "doc_2"
}

documents = [
    doc1_document,
    doc2_document
]
```

```
result = kendra.batch_put_document(
    Documents = documents,
    IndexId = index_id,
    RoleArn = role_arn
)

print(result)
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.BatchPutDocumentRequest;
import software.amazon.awssdk.services.kendra.model.BatchPutDocumentResponse;
import software.amazon.awssdk.services.kendra.model.Document;
import software.amazon.awssdk.services.kendra.model.S3Path;

public class AddFilesFromS3Example {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();

        String indexId = "yourIndexId";
        String roleArn = "yourIndexRoleArn";

        Document pollyDoc = Document
            .builder()
            .s3Path(
                S3Path.builder()
                .bucket("an-aws-kendra-test-bucket")
                .key("What is Amazon Polly.docx")
                .build())
            .title("What is Amazon Polly")
            .id("polly_doc_1")
            .build();

        Document rekognitionDoc = Document
            .builder()
            .s3Path(
                S3Path.builder()
                .bucket("an-aws-kendra-test-bucket")
                .key("What is Amazon Rekognition.docx")
```

```
                .build())
            .title("What is Amazon rekognition")
            .id("rekognition_doc_1")
            .build();

        BatchPutDocumentRequest batchPutDocumentRequest = BatchPutDocumentRequest
            .builder()
            .indexId(indexId)
            .roleArn(roleArn)
            .documents(pollyDoc, rekognitionDoc)
            .build();

        BatchPutDocumentResponse result =
 kendra.batchPutDocument(batchPutDocumentRequest);

        System.out.println(String.format("BatchPutDocument result: %s", result));
    }
}
```

将常见问题解答 (FAQ) 添加到索引中

您可以使用控制台或 CreateFaqAPI 将常见问题 (FAQ) 直接添加到索引中。将常见问题解答添加到索引中是一种异步操作。您将常见问题解答的数据放在 Amazon Simple Storage Service 存储桶中的文件中。您可以使用 CSV 或 JSON 文件作为常见问题解答的输入：

- 基本 CSV - CSV 文件，其中每行包含一个问题、答案和一个可选的源 URI。
- 自定义 CSV - 包含问题、答案和自定义字段/属性的标头的 CSV 文件，您可以使用这些字段/属性对常见问题答案进行分面、显示或排序。您还可以定义访问权限控制字段，将常见问题解答响应限制为仅允许特定用户和组查看。
- JSON - 包含问题、答案和自定义字段/属性的 JSON 文件，您可以使用这些字段/属性对常见问题解答进行分面、显示或排序。您还可以定义访问权限控制字段，将常见问题解答响应限制为仅允许特定用户和组查看。

例如，下面是一个基本 CSV 文件，它提供了有关美国华盛顿州斯波坎和美国密苏里州山景城免费诊所的问题的答案。

```
How many free clinics are in Spokane WA?, 13
How many free clinics are there in Mountain View Missouri?, 7
```

> **Note**
>
> FAQ 文件必须是 UTF-8 编码的文件。

主题

为常见问题解答文件创建索引字段

使用自定义 CSV 或 JSON 文件进行输入时，可以为常见问题声明自定义字段。例如，您可以创建一个自定义字段，将每个常见问题解答的问题分配一个业务部门。例如，在响应中返回常见问题解答时，您可以使用部门作为一个分面，将搜索范围缩小到“人力资源”或“财务”部门。

自定义字段必须映射到索引字段。在控制台中，您可以使用分面定义页面来创建索引字段。使用 API 时，必须先使用 UpdateIndexAPI 创建索引字段。

常见问题解答文件中的字段/属性类型必须与关联的索引字段的类型相匹配。例如，“部门”字段是一个 `STRING_LIST` 类型字段。因此，您必须在常见问题解答文件中以字符串列表的形式提供部门字段的值。您可以使用控制台中的 Facet 定义页面或使用 DescribeIndexAPI 来检查索引字段的类型。

创建映射到自定义属性的索引字段时，您可以将其标记为可显示、可分面或可排序。无法将自定义属性设置为可搜索。

除了自定义属性外，您还可以在自定义 CSV 或 JSON 文件中使用 Amazon Kendra 保留或常用的字段。有关更多信息，请参阅文档属性或字段。

基本 CSV 文件

如果要为常见问题解答使用简单的结构，请使用基本 CSV 文件。在基本 CSV 文件中，每行都有两个或三个字段：问题、答案和指向包含更多信息的文档的可选来源 URI。

文件内容必须符合逗号分隔值 (CSV) 文件的 RFC 4180 通用格式和 MIME 类型。

下面是基本 CSV 格式的常见问题解答文件。

```
How many free clinics are in Spokane WA?, 13, https://s3.region.company.com/bucket-
name/directory/faq.csv
How many free clinics are there in Mountain View Missouri?, 7, https://
s3.region.company.com/bucket-name/directory/faq.csv
```

自定义 CSV 文件

如果您想在常见问题解答的问题中添加自定义字段/属性，请使用自定义 CSV 文件。对于自定义 CSV 文件，您可以使用 CSV 文件中的标头行来定义其他属性。

CSV 文件必须包含以下两个必需字段：

- `_question` - 常见问题
- `_answer` - 常见问题的答案

您的文件可以包含 Amazon Kendra 保留字段和自定义字段。以下是自定义 CSV 文件的示例。

```
_question,_answer,_last_updated_at,custom_string
How many free clinics are in Spokane WA?, 13, 2012-03-25T12:30:10+01:00, Note: Some
 free clinics require you to meet certain criteria in order to use their services
How many free clinics are there in Mountain View Missouri?, 7,
 2012-03-25T12:30:10+01:00, Note: Some free clinics require you to meet certain
 criteria in order to use their services
```

自定义文件内容必须符合逗号分隔值 (CSV) 文件的 RFC 4180 通用格式和 MIME 类型。

下面列出了自定义字段的类型：

- Date - ISO 8601 编码的日期和时间值。

  例如，2012-03-25T12:30:10+01:00 是中部欧洲时间 2012 年 3 月 25 日中午 12:30（10 秒）的 ISO 8601 日期-时间格式。
- Long - 数字，例如，1234。
- String - 字符串值。如果字符串包含逗号，请用双引号 (") 将整个值括起来（例如，`"custom attribute, and more"`）。
- String list - 字符串值的列表。列出使用引号 (") 括起来的逗号分隔值列表（例如，`"item1, item2, item3"`）。如果列表仅包含一个条目，则可以省略引号（例如，`item1`）。

自定义 CSV 文件可以包含用户访问权限控制字段。您可以使用这些字段将常见问题解答限制为仅某些用户和组可访问。要根据用户上下文进行筛选，用户必须在查询中提供用户和组信息。否则，将返回所有相关的常见问题解答。有关更多信息，请参阅用户上下文筛选。

下面列出了常见问题解答的用户上下文筛选条件：

- _acl_user_allow - 允许列表中的用户可以在查询响应中查看常见问题解答。常见问题解答不会返回给其他用户。
- _acl_user_deny - 拒绝列表中的用户无法在查询响应中查看常见问题解答。当常见问题解答与查询相关时，会将其返回给所有其他用户。
- _acl_group_allow - 属于允许群组的用户可以在查询响应中查看常见问题解答。常见问题解答不会返回给属于其他组的成员的用户。
- _acl_group_deny - 属于拒绝组成员的用户无法在查询响应中查看常见问题解答。当常见问题解答与查询相关时，会将其返回给其他组。

在用引号括起来的逗号分隔列表中提供允许和拒绝列表的值（例如，"user1,user2,user3"）。您可以将用户或组包含在允许列表或拒绝列表中，但不能将同一用户或组同时包括在允许列表或拒绝列表中。如果将用户或组包含在两者中，则会收到错误。

下面是包含用户上下文信息的自定义 CSV 文件的示例。

```
_question, _answer, _acl_user_allow, _acl_user_deny, _acl_group_allow, _acl_group_deny
How many free clinics are in Spokane WA?, 13, "userID6201,userID7552",
 "userID1001,userID2020", groupBasicPlusRate, groupPremiumRate
```

JSON 文件

您可以使用 JSON 文件为索引提供问题、答案和字段。您可以将任何 Amazon Kendra 保留字段或自定义字段添加到常见问题解答中。

下面是 JSON 文件的架构。

```
{
    "SchemaVersion": 1,
    "FaqDocuments": [
        {
            "Question": string,
            "Answer": string,
            "Attributes": {
```

```
                string: object
                additional attributes
            },
            "AccessControlList": [
               {
                   "Name": string,
                   "Type": enum( "GROUP" | "USER" ),
                   "Access": enum( "ALLOW" | "DENY" )
               },
               additional user context
            ]
        },
        additional FAQ documents
    ]
}
```

以下示例 JSON 文件显示了两个常见问题解答文档。其中一个文档只包含所需的问题和答案。另一个文档还包含其他字段和用户上下文或访问权限控制信息。

```
{
    "SchemaVersion": 1,
    "FaqDocuments": [
        {
            "Question": "How many free clinics are in Spokane WA?",
            "Answer": "13"
        },
        {
            "Question": "How many free clinics are there in Mountain View Missouri?",
            "Answer": "7",
            "Attributes": {
                "_source_uri": "https://s3.region.company.com/bucket-name/directory/
faq.csv",
                "_category": "Charitable Clinics"
            },
            "AccessControlList": [
               {
                   "Name": "user@amazon.com",
                   "Type": "USER",
                   "Access": "ALLOW"
               },
               {
                   "Name": "Admin",
                   "Type": "GROUP",
```

```
                    "Access": "ALLOW"
                }
            ]
        }
    ]
}
```

下面列出了自定义字段的类型：

- Date - 使用 ISO 8601 编码的日期和时间值的 JSON 字符串值。例如，2012-03-25T12:30:10+01:00 是中部欧洲时间 2012 年 3 月 25 日中午 12:30（10 秒）的 ISO 8601 日期-时间格式。
- Long - JSON 数值，例如，1234。
- String - JSON 字符串值（例如，"custom attribute"）。
- String list - 字符串值的 JSON 数组（例如，["item1,item2,item3"]）。

JSON 文件可以包含用户访问控制字段。您可以使用这些字段将常见问题解答限制为仅某些用户和组可访问。要根据用户上下文进行筛选，用户必须在查询中提供用户和组信息。否则，将返回所有相关的常见问题解答。有关更多信息，请参阅用户上下文筛选。

您可以将用户或组包含在允许列表或拒绝列表中，但不能将同一用户或组同时包括在允许列表或拒绝列表中。如果将用户或组包含在两者中，则会收到错误。

下面是 JSON 常见问题解答中包含用户访问权限控制的示例。

```
"AccessControlList": [
                {
                    "Name": "group or user name",
                    "Type": "GROUP | USER",
                    "Access": "ALLOW | DENY"
                },
                additional user context
            ]
```

使用常见问题解答文件

将常见问题解答输入文件存储在 S3 存储桶中后，您可以使用控制台或 CreateFaq API 将问题和答案编入索引中。如果要更新常见问题解答，请删除该常见问题解答并重新创建。您可以使用 DeleteFaq API 删除常见问题解答。

您必须提供一个有权访问包含您的源文件的 S3 存储桶的 IAM 角色。您可以在控制台或 RoleArn 参数中指定该角色。下面是将常见问题解答文件添加到索引中的示例。

Python

```
import boto3

kendra = boto3.client("kendra")

# Provide the index ID
index_id = "index-id"
# Provide the IAM role ARN required to index documents in an S3 bucket
role_arn = "arn:aws:iam::${accountId}:role/${roleName}"

# Provide the S3 bucket path information to the FAQ file
faq_path = {
    "Bucket": "bucket-name",
    "Key": "FreeClinicsUSA.csv"
}

response = kendra.create_faq(
    S3Path =  faq_path,
    Name = "FreeClinicsUSA",
    IndexId = index_id,
    RoleArn = role_arn
)

print(response)
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.CreateFaqRequest;
import software.amazon.awssdk.services.kendra.model.CreateFaqResponse;
import software.amazon.awssdk.services.kendra.model.S3Path;

public class AddFaqExample {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();

        String indexId = "yourIndexId";
```

```
        String roleArn = "your role for accessing S3 files";

        CreateFaqRequest createFaqRequest = CreateFaqRequest
            .builder()
            .indexId(indexId)
            .name("FreeClinicsUSA")
            .roleArn(roleArn)
            .s3Path(
                S3Path
                    .builder()
                    .bucket("an-aws-kendra-test-bucket")
                    .key("FreeClinicsUSA.csv")
                    .build())
            .build();

        CreateFaqResponse response = kendra.createFaq(createFaqRequest);

        System.out.println(String.format("The result of creating FAQ: %s",
 response));

    }
}
```

其他语言的常见问题解答文件（除英语外）

您可以使用支持的语言为常见问题解答编制索引。 Amazon Kendra 如果您未指定语言，则默认使用英文索引常见问题解答。您可以在调用CreateFaq操作时指定语言代码，也可以将常见问题解答的语言代码作为字段包含在常见问题元数据中。如果没有在元数据字段的元数据中指定常见问题解答的语言代码，则使用您在调用该 CreateFAQ 操作时指定的语言代码来为常见问题解答编制索引。要在控制台中以支持的语言为常见问题解答文档编制索引，请转到常见问题解答并选择添加常见问题解答。您可以从语言下拉列表中选择语言。

创建自定义文档字段

您可以在 Amazon Kendra 索引中为文档创建自定义属性或字段。例如，您可以创建一个名为“Department”的自定义字段或属性，其值为“HR”、“Sales”和“Manufacturing”。如果您将这些自定义字段或属性映射到您的 Amazon Kendra 索引，则可以使用它们筛选搜索结果，以便按照 “HR” 部门属性包括文档。

必须先在索引中创建字段，然后才能使用自定义字段或属性。使用控制台编辑数据源字段映射以添加自定义字段或使用 UpdateIndexAPI 创建索引字段。创建字段后无法更改字段的数据类型。

对于大多数数据来源，您可以将外部数据来源中的字段映射到 Amazon Kendra 中的相应字段。有关更多信息，请参阅映射数据来源字段。对于 S3 数据来源，您可以使用 JSON 元数据文件来创建自定义字段或属性。

最多可以创建 500 个自定义字段或属性。

您也可以使用 Amazon Kendra 保留字段或常用字段。有关更多信息，请参阅文档属性或字段。

主题

- 更新自定义文档字段

更新自定义文档字段

使用 UpdateIndex API，您可以使用 DocumentMetadataConfigurationUpdates 参数来添加自定义字段或属性。

以下 JSON 示例使用 DocumentMetadataConfigurationUpdates 来向索引添加名为“Department”的字段。

```
"DocumentmetadataConfigurationUpdates": [
   {
       "Name": "Department",
       "Type": "STRING_VALUE"
   }
]
```

以下各节包括使用BatchPutDocument和为 Amazon S3 数据源添加自定义属性或字段的示例。

主题

- 使用 BatchPutDocument API 添加自定义属性或字段
- 向 Amazon S3 数据来源添加自定义属性或字段

使用 BatchPutDocument API 添加自定义属性或字段

使用 BatchPutDocumentAPI 向索引添加文档时，可以将自定义字段或属性指定为其中的一部分`Attributes`。您可以在调用 API 时添加多个字段或属性。最多可以创建 500 个自定义字段或属性。以下示例是将“Department”添加到文档中的自定义字段或属性。

```
"Attributes":
    {
        "Department": "HR",
        "_category": "Vacation policy"
    }
```

向 Amazon S3 数据来源添加自定义属性或字段

使用 S3 存储桶作为索引的数据来源时，您可以将元数据与随附的元数据文件一起添加到文档中。您可以将元数据 JSON 文件放在与您的文档并排的目录结构中。有关更多信息，请参阅 S3 文档元数据。

您可以在 `Attributes` JSON 结构中指定自定义字段或属性。最多可以创建 500 个自定义字段或属性。例如，以下示例使用 `Attributes` 来定义三个自定义字段或属性以及一个保留字段。

```
"Attributes": {
        "brand": "Amazon Basics",
        "price": 1595,
        "_category": "sports",
        "subcategories": ["outdoors", "electronics"]
    }
```

以下步骤将引导您完成向 Amazon S3 数据源添加自定义属性的过程。

主题

- 第 1 步：创建亚马逊 Kendra 索引
- 步骤 2：更新索引以添加自定义文档字段
- 步骤 3：创建 Amazon S3 数据源并将数据源字段映射到自定义属性

第 1 步：创建亚马逊 Kendra 索引

按照中的创建索引步骤创建您的亚马逊 Kendra 索引。

步骤 2：更新索引以添加自定义文档字段

创建索引后，向其添加字段。以下过程说明如何使用控制台和 CLI 向索引添加字段。

Console

创建索引字段

1. 确保您已创建索引。
2. 然后，从左侧导航菜单的 “数据管理” 中选择 Facet 定义。
3. 在 “索引字段设置指南” 中，从 “索引字段” 中选择 “添加字段” 以添加自定义字段。
4. 在添加索引字段对话框中，执行以下操作：
   - 字段名-添加字段名称。
   - 数据类型-选择数据类型，无论是字符串、字符串列表还是日期。
   - 使用类型 -选择使用类型，包括 Facetable、可搜索、可显示和可排序。

   然后，选择 “添加”。

对要映射的任何其他字段重复最后一步。

CLI

```
aws kendra update-index  \
--region $region \
--endpoint-url $endpoint \
--application-id $applicationId \
--index-id $indexId  \
--document-metadata-configuration-updates \
"[
    {
        "Name": "string",
        "Type": "STRING_VALUE"|"STRING_LIST_VALUE"|"LONG_VALUE"|"DATE_VALUE",
        "Relevance": {
            "Freshness": true|false,
            "Importance": integer,
            "Duration": "string",
            "RankOrder": "ASCENDING"|"DESCENDING",
            "ValueImportanceMap": {"string": integer
            ...}
```

```
    },
    "Search": {
        "Facetable": true|false,
        "Searchable": true|false,
        "Displayable": true|false,
        "Sortable": true|false
        }
    }
...
]"
```

步骤 3：创建 Amazon S3 数据源并将数据源字段映射到自定义属性

要创建 Amazon S3 数据源并将字段映射到该数据源，请按照中的说明进行操作Amazon S3。

如果您使用的是 API，请在使用 AP CreateDataSourceI `configuration` 时使用下面的`fieldMappings`属性。

有关如何映射数据源字段的概述，请参阅映射数据来源字段。

使用令牌控制用户访问文档

您可以控制哪些用户或组可以访问索引中的某些文档或在搜索结果中查看某些文档。这称为用户上下文筛选。这是一种个性化搜索，其优点是可以控制对文档的访问权限。例如，并非所有在公司门户网站上搜索信息的团队都应该访问绝密的公司文档，这些文档也不应与所有用户相关。只有获得绝密文档访问权限的特定用户或团队组才能在搜索结果中看到这些文档。

Amazon Kendra 支持使用以下令牌类型进行基于令牌的用户访问控制：

- Open ID
- 带有共享密钥的 JWT
- 带有公有密钥的 JWT
- JSON

Amazon Kendra 为您的搜索应用程序提供高度安全的企业搜索。您的搜索结果反映了您组织的安全模型。客户有责任对用户进行身份验证和授权，使其能够访问其搜索应用程序。在搜索时，根据客户的搜索应用程序提供的用户 ID 以及 Amazon Kendra 连接器在爬取/编制期间收集的文档访问控制列表 (ACL)，Amazon Kendra 服务会筛选搜索结果。搜索结果返回指向原始文档存储库的 URL 以及简短的摘录。对完整文档的访问仍由原始存储库强制执行。

主题

使用 OpenID

要将 Amazon Kendra 索引配置为使用 OpenID 令牌进行访问控制，您需要来自 OpenID 提供程序的 JWKS（JSON Web 密钥集）网址。在大多数情况下，JWKS URL 采用以下格式：`https://domain-name/.well_known/jwks.json`（如果他们关注的是 OpenID 发现）。

以下示例说明在创建索引时如何使用 OpenID 进行用户访问控制。

Console

1. 选择创建索引以开始创建新索引。
2. 在指定索引详细信息页面上，为您的索引指定名称和描述。
3. 对于 IAM 角色，选择一个角色或选择创建新角色，并指定角色名称来创建新角色。IAM 角色将带有前缀 “AmazonKendra-”。
4. 将所有其他字段保留为默认值。选择下一步。
5. 在配置用户访问控制页面上，在访问控制设置下，选择是以使用令牌进行访问控制。
6. 在令牌配置下，选择 OpenID 作为令牌类型。
7. 指定签名密钥 URL。该 URL 应指向一组 JSON Web 密钥。
8. （可选）在高级配置下：
    a. 指定要在 ACL 检查中使用的用户名。
    b. 指定要在 ACL 检查中使用的一个或多个组。
    c. 指定将验证颁发机构的颁发机构。
    d. 指定客户端 ID。您必须指定与 JWT 中的受众相匹配的正则表达式。
9. 在配置详细信息页面中，选择开发者版本。
10. 选择创建以创建索引。
11. 等待您的索引创建完成。 Amazon Kendra 为您的索引配置硬件。此操作可能需要一些时间。

CLI

要使用 JSON 输入文件创建索引，请先 AWS CLI 使用所需参数创建一个 JSON 文件：

```
{
    "Name": "user-context",
    "Edition": "ENTERPRISE_EDITION",
    "RoleArn": "arn:aws:iam::account-id:role:/my-role",
    "UserTokenConfigurations": [
        {
            "JwtTokenTypeConfiguration": {
                "KeyLocation": "URL",
                "Issuer": "optional: specify the issuer url",
                "ClaimRegex": "optional: regex to validate claims in the token",
                "UserNameAttributeField": "optional: user",
                "GroupAttributeField": "optional: group",
                "URL": "https://example.com/.well-known/jwks.json"
            }
        }
    ],
    "UserContextPolicy": "USER_TOKEN"
}
```

您可以覆盖默认的用户和组字段名称。UserNameAttributeField 的默认值为“user”。GroupAttributeField 的默认值为“groups”。

接下来，使用输入文件调用 create-index。例如，如果您的 JSON 文件名为 create-index-openid.json，则可以使用以下名称：

```
aws kendra create-index --cli-input-json file://create-index-openid.json
```

Python

```
response = kendra.create_index(
    Name='user-context',
    Edition='ENTERPRISE_EDITION',
    RoleArn='arn:aws:iam::account-id:role:/my-role',
    UserTokenConfigurations=[
        {
            "JwtTokenTypeConfiguration": {
                "KeyLocation": "URL",
                "Issuer": "optional: specify the issuer url",
```

```
                "ClaimRegex": "optional: regex to validate claims in the token",
                "UserNameAttributeField": "optional: user",
                "GroupAttributeField": "optional: group",
                "URL": "https://example.com/.well-known/jwks.json"
            }
        }
    ],
    UserContextPolicy='USER_TOKEN'
)
```

使用带有共享密钥的 JSON Web 令牌（JWT）

以下示例说明如何在创建索引时将 JSON Web Token (JWT) 与共享密钥令牌一起用于用户访问控制。

Console

1. 选择创建索引以开始创建新索引。
2. 在指定索引详细信息页面上，为您的索引指定名称和描述。
3. 对于 IAM 角色，选择一个角色或选择创建新角色，并指定角色名称来创建新角色。该 IAM 角色将带有前缀 “AmazonKendra-”。
4. 将所有其他字段保留为默认值。选择下一步。
5. 在配置用户访问控制页面上，在访问控制设置下，选择是以使用令牌进行访问控制。
6. 在令牌配置下，选择带有共享密钥的 JWT 作为令牌类型。
7. 在用于签名共享密钥的参数下，选择密钥类型。您可以使用现有的 AWS Secrets Manager 共享密钥或创建一个新共享密钥。

   要创建新共享密钥，请选择新建，然后按照下列步骤操作：

   a. 在 “新 AWS Secrets Manager 密钥” 下，指定密钥名称。保存公有密钥时，将添加前缀 AmazonKendra-。
   b. 指定键 ID。键 ID 是一个提示，指示哪些键用于保护令牌的 JSON Web 签名。
   c. 为令牌选择签名算法。这是用于保护 ID 令牌的加密算法。有关 RSA 的更多信息，请参阅 RSA 密码术。
   d. 通过输入 base64 URL 编码的密钥来指定共享密钥。您也可以选择生成密钥来为您生成密钥。您必须确保该密钥是 base64 URL 编码的密钥。
   e. （可选）指定共享密钥何时有效。您可以指定密钥有效期的开始日期、截止日期，或同时指定两者。该密钥将在指定的时间间隔内有效。

f. 选择保存密钥以保存新密钥。

8. （可选）在高级配置下：

   a. 指定要在 ACL 检查中使用的用户名。

   b. 指定要在 ACL 检查中使用的一个或多个组。

   c. 指定将验证颁发机构的颁发机构。

   d. 指定声明 ID。您必须指定与 JWT 中的受众相匹配的正则表达式。

9. 在配置详细信息页面中，选择开发者版本。

10. 选择创建以创建索引。

11. 等待您的索引创建完成。 Amazon Kendra 为您的索引配置硬件。此操作可能需要一些时间。

CLI

您可以将 JWT 令牌与内部的共享密钥一起使用。 AWS Secrets Manager该密钥必须是 base64 URL 编码的密钥。你需要 Secrets Manager ARN，而且你的 Amazon Kendra 角色必须有权访问该`GetSecretValue`资源。 Secrets Manager 如果您使用对 Secrets Manager 资源进行加密 AWS KMS，则该角色还必须有权访问解密操作。

要使用 JSON 输入文件创建索引，请先 AWS CLI 使用所需参数创建一个 JSON 文件：

```
{
    "Name": "user-context",
    "Edition": "ENTERPRISE_EDITION",
    "RoleArn": "arn:aws:iam::account-id:role:/my-role",
    "UserTokenConfigurations": [
        {
            "JwtTokenTypeConfiguration": {
                "KeyLocation": "SECRET_MANAGER",
                "Issuer": "optional: specify the issuer url",
                "ClaimRegex": "optional: regex to validate claims in the token",
                "UserNameAttributeField": "optional: user",
                "GroupAttributeField": "optional: group",
                "SecretManagerArn": "arn:aws:secretsmanager:us-west-2:account
 id:secret:/my-user-context-secret
            }
        }
    ],
    "UserContextPolicy": "USER_TOKEN"
}
```

您可以覆盖默认的用户和组字段名称。UserNameAttributeField 的默认值为“user”。GroupAttributeField 的默认值为“groups”。

接下来，使用输入文件调用 create-index。例如，如果您的 JSON 文件名为 create-index-openid.json，则可以使用以下名称：

```
aws kendra create-index --cli-input-json file://create-index-openid.json
```

密钥必须采用以下格式 AWS Secrets Manager：

```
{
  "keys": [
    {
      "kid": "key_id",
      "alg": "HS256|HS384|HS512",
      "kty": "OCT",
      "use": "sig", //this value can be sig only for now
      "k": "secret",
      "nbf":"ISO1806 date format"
      "exp":"ISO1806 date format"
    }
  ]
}
```

有关 JWT 的更多信息，请参阅 jwt.io。

Python

您可以将 JWT 令牌与内部的共享密钥一起使用。 AWS Secrets Manager该密钥必须是 base64 URL 编码的密钥。你需要 Secrets Manager ARN，而且你的 Amazon Kendra 角色必须有权访问该GetSecretValue资源。 Secrets Manager 如果您使用对 Secrets Manager 资源进行加密 AWS KMS，则该角色还必须有权访问解密操作。

```
response = kendra.create_index(
    Name='user-context',
    Edition='ENTERPRISE_EDITION',
    RoleArn='arn:aws:iam::account-id:role:/my-role',
    UserTokenConfigurations=[
        {
            "JwtTokenTypeConfiguration": {
                "KeyLocation": "URL",
                "Issuer": "optional: specify the issuer url",
```

```
                "ClaimRegex": "optional: regex to validate claims in the token",
                "UserNameAttributeField": "optional: user",
                "GroupAttributeField": "optional: group",
                "SecretManagerArn": "arn:aws:secretsmanager:us-west-2:account
 id:secret:/my-user-context-secret"
            }
        }
    ],
    UserContextPolicy='USER_TOKEN'
)
```

使用带有公有密钥的 JSON Web 令牌（JWT）

以下示例说明如何在创建索引时使用带有公钥的 JSON Web Token (JWT) 进行用户访问控制。有关 JWT 的更多信息，请参阅 jwt.io。

Console

1. 选择创建索引以开始创建新索引。
2. 在指定索引详细信息页面上，为您的索引指定名称和描述。
3. 对于 IAM 角色，选择一个角色或选择创建新角色，并指定角色名称来创建新角色。该 IAM 角色将带有前缀 “AmazonKendra-”。
4. 将所有其他字段保留为默认值。选择下一步。
5. 在配置用户访问控制页面上，在访问控制设置下，选择是以使用令牌进行访问控制。
6. 在令牌配置下，选择带有公有密钥的 JWT 作为令牌类型。
7. 在用于签名公有密钥的参数下，选择密钥类型。您可以使用现有的 AWS Secrets Manager 密钥或创建一个新密钥。

   要创建新密钥，请选择新建，然后按照下列步骤操作：

   a. 在 “新 AWS Secrets Manager 密钥” 下，指定密钥名称。保存公有密钥时，将添加前缀 AmazonKendra-。
   b. 指定键 ID。键 ID 是一个提示，指示哪些键用于保护令牌的 JSON Web 签名。
   c. 为令牌选择签名算法。这是用于保护 ID 令牌的加密算法。有关 RSA 的更多信息，请参阅 RSA 密码术。
   d. 在证书属性下，指定可选的证书链。证书链由证书列表组成。它以服务器的证书开头，以根证书结尾。

e. 可选 指定指纹。它是证书的哈希值，可计算出所有证书数据及其签名。

f. 指定指数。这是 RSA 公有密钥的指数值。它表示为采用 Base64urlUInt 编码的值。

g. 指定模数。这是 RSA 公有密钥的指数值。它表示为采用 Base64urlUInt 编码的值。

h. 选择保存密钥以保存新密钥。

8. （可选）在高级配置下：

a. 指定要在 ACL 检查中使用的用户名。

b. 指定要在 ACL 检查中使用的一个或多个组。

c. 指定将验证颁发机构的颁发机构。

d. 指定客户端 ID。您必须指定与 JWT 中的受众相匹配的正则表达式。

9. 在配置详细信息页面中，选择开发者版本。

10. 选择创建以创建索引。

11. 等待您的索引创建完成。 Amazon Kendra 为您的索引配置硬件。此操作可能需要一些时间。

CLI

您可以将 JWT 与 AWS Secrets Manager内部的公有密钥一起使用。你需要 Secrets Manager ARN，而且你的 Amazon Kendra 角色必须有权访问该`GetSecretValue`资源。 Secrets Manager 如果您使用对 Secrets Manager 资源进行加密 AWS KMS，则该角色还必须有权访问解密操作。

要使用 JSON 输入文件创建索引，请先 AWS CLI 使用所需参数创建一个 JSON 文件：

```
{
    "Name": "user-context",
    "Edition": "ENTERPRISE_EDITION",
    "RoleArn": "arn:aws:iam::account id:role:/my-role",
    "UserTokenConfigurationList": [
        {
            "JwtTokenTypeConfiguration": {
                "KeyLocation": "SECRET_MANAGER",
                "Issuer": "optional: specify the issuer url",
                "ClaimRegex": "optional: regex to validate claims in the token",
                "UserNameAttributeField": "optional: user",
                "GroupAttributeField": "optional: group",
                "SecretManagerArn": "arn:aws:secretsmanager:us-west-2:account
 id:secret:/my-user-context-secret
            }
        }
```

```
    ],    "UserContextPolicy": "USER_TOKEN"
}
```

您可以覆盖默认的用户和组字段名称。UserNameAttributeField 的默认值为“user”。GroupAttributeField 的默认值为“groups”。

接下来，使用输入文件调用 create-index。例如，如果您的 JSON 文件名为 create-index-openid.json，则可以使用以下名称：

```
aws kendra create-index --cli-input-json file://create-index-openid.json
```

密钥必须采用以下格式 Secrets Manager：

```
{
  "keys": [
    {
      "alg": "RS256|RS384|RS512",
      "kty": "RSA", //this can be RSA only for now
      "use": "sig", //this value can be sig only for now
      "n": "modulus of standard pem",
      "e": "exponent of standard pem",
      "kid": "key_id",
      "x5t": "certificate thumprint for x.509 cert",
      "x5c": [
        "certificate chain"
      ]
    }
  ]
}
```

有关 JWT 的更多信息，请参阅 jwt.io。

Python

```
response = kendra.create_index(
    Name='user-context',
    Edition='ENTERPRISE_EDITION',
    RoleArn='arn:aws:iam::account id:role:/my-role',
    UserTokenConfigurationList=[
        {
            "JwtTokenTypeConfiguration": {
                "KeyLocation": "URL",
```

```
                "Issuer": "optional: specify the issuer url",
                "ClaimRegex": "optional: regex to validate claims in the token",
                "UserNameAttributeField": "optional: user",
                "GroupAttributeField": "optional: group",
                "SecretManagerArn": "arn:aws:secretsmanager:us-west-2:account
 id:secret:/my-user-context-secret"
            }
        }
    ],
    UserContextPolicy='USER_TOKEN'
)
```

使用 JSON

以下示例说明如何在创建索引时使用 JSON 进行用户访问控制。

> **Warning**
>
> JSON 令牌是未经验证的有效负载。只有当对 Amazon Kendra 的请求来自可信服务器而不是浏览器时，才应使用此选项。

Console

1. 选择创建索引以开始创建新索引。
2. 在指定索引详细信息页面上，为您的索引指定名称和描述。
3. 对于 IAM 角色，选择一个角色或选择创建新角色，并指定角色名称来创建新角色。该 IAM 角色将带有前缀 “AmazonKendra-”。
4. 将所有其他字段保留为默认值。选择下一步。
5. 在配置用户访问控制页面上，在访问控制设置下，选择是以使用令牌进行访问控制。
6. 在令牌配置下，选择 JSON 作为令牌类型。
7. 指定要在 ACL 检查中使用的用户名。
8. 指定要在 ACL 检查中使用的一个或多个组。
9. 选择下一步。
10. 在配置详细信息页面中，选择开发者版本。
11. 选择创建以创建索引。

12. 等待您的索引创建完成。 Amazon Kendra 为您的索引配置硬件。此操作可能需要一些时间。

CLI

要使用 JSON 输入文件创建索引，请先 AWS CLI 使用所需参数创建一个 JSON 文件：

```
{
    "Name": "user-context",
    "Edition": "ENTERPRISE_EDITION",
    "RoleArn": "arn:aws:iam::account-id:role:/my-role",
    "UserTokenConfigurations": [
        {
            "JsonTokenTypeConfiguration": {
                "UserNameAttributeField": "user",
                "GroupAttributeField": "group"
            }
        }
    ],
    "UserContextPolicy": "USER_TOKEN"
}
```

接下来，使用输入文件调用 `create-index`。例如，如果您的 JSON 文件名为 `create-index-openid.json`，则可以使用以下名称：

```
aws kendra create-index --cli-input-json file://create-index-openid.json
```

如果您未使用 Open ID AWS IAM Identity Center，则可以将 JSON 格式的令牌发送给我们。如果这样做，则必须指定 JSON 令牌中的哪个字段包含用户名，哪个字段包含组。组字段值必须是 JSON 字符串数组。例如，如果您使用 SAML，您的令牌可能类似于以下内容：

```
{
     "username" : "user1",
     "groups": [
        "group1",
        "group2"
     ]
}
```

`TokenConfiguration` 指定用户名和组字段名称：

```
{
```

```
    "UserNameAttributeField":"username",
    "GroupAttributeField":"groups"
}
```

Python

```
response = kendra.create_index(
    Name='user-context',
    Edition='ENTERPRISE_EDITION',
    RoleArn='arn:aws:iam::account-id:role:/my-role',
    UserTokenConfigurations=[
        {
            "JwtTokenTypeConfiguration": {
                "UserNameAttributeField": "user",
                "GroupAttributeField": "group",
            }
        }
    ],
    UserContextPolicy='USER_TOKEN'
)
```

创建数据来源连接器

您可以为创建数据源连接器 Amazon Kendra ，以便连接到您的文档并为其编制索引。 Amazon Kendra 可以连接到 Microsoft SharePoint、Google 云端硬盘和许多其他提供商。创建数据源连接器时，需要提供 Amazon Kendra 连接到源存储库所需的配置信息。与直接向索引添加文档不同，您可以定期扫描数据来源以更新索引。

例如，假设您有一个存储在存储 Amazon S3 桶中的税务文件存储库。不时更改现有文档并将新文档添加到存储库中。如果将存储库 Amazon Kendra 作为数据源添加到中，则可以通过在数据源和索引之间设置定期同步来使索引保持最新状态。

您可以选择使用控制台或 StartDataSourceSyncJobAPI 手动更新索引。否则，您可以设置一个计划来更新索引并使其与您的数据来源同步。

一个索引可以包含多个数据来源。每个数据来源可以有自己的更新计划。例如，您可以每天甚至每小时更新工作文档的索引，同时在存档发生变化时手动更新存档的文档。

如果您想在文档提取过程中更改元数据或属性和内容，请参阅 Amazon Kendra 自定义文档富集。

> **Note**
>
> 每个索引的每个文档 ID 必须是唯一的。在创建数据来源时，您不能使用文档的唯一 ID 来编制索引，然后使用 `BatchPutDocument` API 为相同的文档编制索引，反之亦然。您可以删除数据来源，然后使用 `BatchPutDocument` API 为相同的文档编制索引，反之亦然。将`BatchPutDocument`和 `BatchDeleteDocument` API 与 Amazon Kendra 数据源连接器结合使用同一组文档可能会导致数据不一致。我们建议使用 Amazon Kendra 自定义数据来源连接器。

> **Note**
>
> 添加到索引中的文件必须采用 UTF-8 编码的字节流。有关中文档的更多信息 Amazon Kendra，请参阅文档。

设置更新计划

将您的数据来源配置为使用控制台定期更新，或者在创建或更新数据来源时使用 `Schedule` 参数进行更新。参数的内容是一个字符串，它包含 `cron` 格式的计划字符串或空字符串，表示索引是按需更新的。有关 cron 表达式的格式，请参阅《Amazon CloudWatch Events 用户指南》中的规则计划表达式。 Amazon Kendra 仅支持 cron 表达式。它不支持 rate 表达式。

设置语言

您可以使用支持的语言为数据来源中的所有文档编制索引。在调用时，您可以为数据源中的所有文档指定语言代码CreateDataSource。如果文档没有在元数据字段中指定语言代码，则使用为数据来源级别的所有文档指定的语言代码为该文档编制索引。如果未指定语言，则在默认情况下， Amazon Kendra 会使用英语为数据来源中的文档编制索引。有关支持的语言（包括其代码）的更多信息，请参阅添加非英语语言文档。

在控制台中，您可以使用支持的语言为数据来源中的所有文档编制索引。转到数据来源并编辑您的数据来源，或者如果您要添加新的数据来源，请添加数据来源。在指定数据来源详细信息页面上，从语言下拉列表中选择一种语言。选择更新或继续输入配置信息以连接数据来源。

数据来源连接器

本节介绍如何使用 Amazon Kendra 和 Amazon Kendra API Amazon Kendra 连接到支持的数据库 AWS Management Console 和数据源存储库。

主题

- 数据来源模板架构
- Adobe Experience Manager
- Alfresco
- Aurora (MySQL)
- Aurora (PostgreSQL)
- Amazon FSx（视窗）
- Amazon FSx（NetApp ONTAP）
- Amazon RDS/Aurora
- Amazon RDS（微软 SQL Server）

- Amazon RDS (MySQL)
- Amazon RDS (Oracle)
- Amazon RDS (PostgreSQL)
- Amazon S3
- Amazon Kendra 网络爬虫
- Amazon WorkDocs
- Box
- Confluence
- 自定义数据来源连接器
- Dropbox
- Drupal
- GitHub
- Gmail
- Google Drive
- IBM DB2
- Jira
- Microsoft Exchange
- 微软 OneDrive
- 微软 SharePoint
- Microsoft SQL Server
- Microsoft Teams
- Micoft Yammer
- MySQL
- Oracle Database
- PostgreSQL
- Quip
- Salesforce
- ServiceNow
- Slack
- Zendesk

数据来源模板架构

以下是支持模板的数据来源的模板架构。

主题

- PostgreSQL 模板架构
- Salesforce 模板架构
- ServiceNow 模板架构
- Slack 模板架构
- Zendesk 模板架构

Adobe Experience Manager 模板架构

您可以将包含数据来源架构的 JSON 作为 TemplateConfiguration 对象的一部分。在连接配置或存储库端点详细信息中，您需要提供 Adobe Experience Manager 主机 URL、身份验证类型，以及您是使用 Adobe Experience Manager（AEM）即云服务还是 AEM On-Premise。此外，请将数据来源的类型指定为 AEM、身份验证凭证的密钥以及其他必要的配置。然后，当您 CreateDataSource 时，您可以将 TEMPLATE 指定为 Type。

您可以使用本开发者指南中提供的模板。有关更多信息，请参阅 Adobe Experience Manager JSON 架构。

下表描述了 AEM JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
aemUrl	Adobe Experience Manager 主机 URL。例如，如果您使用 AEM On-Premise，则需要包含主机名和端口：https://hostname:port。或者，如果您使用 AEM 即云服务，则可以使用作者 URL：https://author-xxxxxx-xxxxxxx.adobeaemcloud.com。
authType	您使用的身份验证类型，可以是 Basic 或 OAuth2。
deploymentType	您使用的 Adobe Experience Manager 的类型，可以是 CLOUD 或 ON_PREMISE 。

配置	描述
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• 页 • asset	将Adobe Experience Manager页面和资产的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。
timeZoneId	如果您使用 AEM 本地部署，并且服务器的时区与 AEM 连接器或索引的时区不同，则可以指定与 Amazon Kendra AEM 连接器或索引对齐的服务器时区。 AEM 本地部署的默认时区是 AE Amazon Kendra M 连接器或索引的时区。AEM 即云服务的默认时区是格林威治标准时间。
• pageRootPaths • assetRootPaths	页面和资产的根路径列表。例如，页面的根路径可以是 /content/sub，而资源的根路径可以是 /content/sub/asset1。
crawlAssets	为 `true` 则爬取资产。
crawlPages	为 `true` 则爬取页面。
• pagePathInclusion图案 • pageNameInclusion图案 • assetPathInclusion图案 • assetTypeInclusion图案 • assetNameInclusion图案	用于在 Adobe Experience Manager 数据来源中包含某些特定页面和资源的正则表达式模式的列表。与模式匹配的页面和资产将包含在索引中。与模式不匹配的页面和资产将从索引中排除。如果页面或资产同时匹配包含和排除模式，则排除模式优先，也就是说，内容不会包含在索引中。

配置	描述
• pagePathExclusion图案 • pageNameExclusion图案 • assetPathExclusion图案 • assetTypeInclusion图案 • assetNameInclusion图案	用于在 Adobe Experience Manager 数据来源中排除某些特定页面和资源的正则表达式模式的列表。与模式匹配的页面和资产将从索引中排除。与模式不匹配的页面和资产将包含在索引中。如果页面或资产同时匹配包含和排除模式，则排除模式优先，也就是说，内容不会包含在索引中。
pageComponents	您想要编入索引的特定页面组件的名称的列表。
contentFragmentVariations	您想要编入索引的 Adobe Experience Manager 内容片段的特定已保存变体的名称列表。
type	数据来源的类型。指定 `AEM` 作为数据来源类型。
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	包含连接到 Adobe Experience Man AWS Secrets Manager ager 所需的键值对的密钥的亚马逊资源名称 (ARN)。有关这些键值对的信息，请参阅 Adobe Experience Manager 的连接说明。
版本	当前支持的此模板的版本。

Adobe Experience Manager JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties":
  {
    "connectionConfiguration": {
      "type": "object",
      "properties":
      {
        "repositoryEndpointMetadata":
```

```
        {
          "type": "object",
          "properties":
          {
            "aemUrl":
            {
              "type": "string",
              "pattern": "https:.*"
            },
            "authType": {
              "type": "string",
              "enum": ["Basic", "OAuth2"]
            },
            "deploymentType": {
              "type": "string",
              "enum": ["CLOUD","ON_PREMISE"]
            }
          },
          "required":
          [
            "aemUrl",
            "authType",
            "deploymentType"
          ]
        }
      },
      "required":
      [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties":
      {
        "page":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
```

```
            [
              {
                "type": "object",
                "properties":
                {
                  "indexFieldName":
                  {
                    "type": "string"
                  },
                  "indexFieldType":
                  {
                    "type": "string",
                    "enum":
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName":
                  {
                    "type": "string"
                  },
                  "dateFieldFormat":
                  {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required":
                [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
```

```
        },
        "asset":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
                      [
                        "STRING",
                        "STRING_LIST",
                        "DATE",
                        "LONG"
                      ]
                    },
                    "dataSourceFieldName":
                    {
                      "type": "string"
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required":
                  [
                    "indexFieldName",
                    "indexFieldType",
```

```
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
      }
    }
  },
  "additionalProperties": {
    "type": "object",
    "properties":
    {
      "timeZoneId": {
        "type": "string",
        "enum": [
          "Africa/Abidjan",
          "Africa/Accra",
          "Africa/Addis_Ababa",
          "Africa/Algiers",
          "Africa/Asmara",
          "Africa/Asmera",
          "Africa/Bamako",
          "Africa/Bangui",
          "Africa/Banjul",
          "Africa/Bissau",
          "Africa/Blantyre",
          "Africa/Brazzaville",
          "Africa/Bujumbura",
          "Africa/Cairo",
          "Africa/Casablanca",
          "Africa/Ceuta",
          "Africa/Conakry",
          "Africa/Dakar",
          "Africa/Dar_es_Salaam",
          "Africa/Djibouti",
          "Africa/Douala",
          "Africa/El_Aaiun",
          "Africa/Freetown",
          "Africa/Gaborone",
```

```
            "Africa/Harare",
            "Africa/Johannesburg",
            "Africa/Juba",
            "Africa/Kampala",
            "Africa/Khartoum",
            "Africa/Kigali",
            "Africa/Kinshasa",
            "Africa/Lagos",
            "Africa/Libreville",
            "Africa/Lome",
            "Africa/Luanda",
            "Africa/Lubumbashi",
            "Africa/Lusaka",
            "Africa/Malabo",
            "Africa/Maputo",
            "Africa/Maseru",
            "Africa/Mbabane",
            "Africa/Mogadishu",
            "Africa/Monrovia",
            "Africa/Nairobi",
            "Africa/Ndjamena",
            "Africa/Niamey",
            "Africa/Nouakchott",
            "Africa/Ouagadougou",
            "Africa/Porto-Novo",
            "Africa/Sao_Tome",
            "Africa/Timbuktu",
            "Africa/Tripoli",
            "Africa/Tunis",
            "Africa/Windhoek",
            "America/Adak",
            "America/Anchorage",
            "America/Anguilla",
            "America/Antigua",
            "America/Araguaina",
            "America/Argentina/Buenos_Aires",
            "America/Argentina/Catamarca",
            "America/Argentina/ComodRivadavia",
            "America/Argentina/Cordoba",
            "America/Argentina/Jujuy",
            "America/Argentina/La_Rioja",
            "America/Argentina/Mendoza",
            "America/Argentina/Rio_Gallegos",
            "America/Argentina/Salta",
```

```
            "America/Argentina/San_Juan",
            "America/Argentina/San_Luis",
            "America/Argentina/Tucuman",
            "America/Argentina/Ushuaia",
            "America/Aruba",
            "America/Asuncion",
            "America/Atikokan",
            "America/Atka",
            "America/Bahia",
            "America/Bahia_Banderas",
            "America/Barbados",
            "America/Belem",
            "America/Belize",
            "America/Blanc-Sablon",
            "America/Boa_Vista",
            "America/Bogota",
            "America/Boise",
            "America/Buenos_Aires",
            "America/Cambridge_Bay",
            "America/Campo_Grande",
            "America/Cancun",
            "America/Caracas",
            "America/Catamarca",
            "America/Cayenne",
            "America/Cayman",
            "America/Chicago",
            "America/Chihuahua",
            "America/Ciudad_Juarez",
            "America/Coral_Harbour",
            "America/Cordoba",
            "America/Costa_Rica",
            "America/Creston",
            "America/Cuiaba",
            "America/Curacao",
            "America/Danmarkshavn",
            "America/Dawson",
            "America/Dawson_Creek",
            "America/Denver",
            "America/Detroit",
            "America/Dominica",
            "America/Edmonton",
            "America/Eirunepe",
            "America/El_Salvador",
            "America/Ensenada",
```

```
            "America/Fort_Nelson",
            "America/Fort_Wayne",
            "America/Fortaleza",
            "America/Glace_Bay",
            "America/Godthab",
            "America/Goose_Bay",
            "America/Grand_Turk",
            "America/Grenada",
            "America/Guadeloupe",
            "America/Guatemala",
            "America/Guayaquil",
            "America/Guyana",
            "America/Halifax",
            "America/Havana",
            "America/Hermosillo",
            "America/Indiana/Indianapolis",
            "America/Indiana/Knox",
            "America/Indiana/Marengo",
            "America/Indiana/Petersburg",
            "America/Indiana/Tell_City",
            "America/Indiana/Vevay",
            "America/Indiana/Vincennes",
            "America/Indiana/Winamac",
            "America/Indianapolis",
            "America/Inuvik",
            "America/Iqaluit",
            "America/Jamaica",
            "America/Jujuy",
            "America/Juneau",
            "America/Kentucky/Louisville",
            "America/Kentucky/Monticello",
            "America/Knox_IN",
            "America/Kralendijk",
            "America/La_Paz",
            "America/Lima",
            "America/Los_Angeles",
            "America/Louisville",
            "America/Lower_Princes",
            "America/Maceio",
            "America/Managua",
            "America/Manaus",
            "America/Marigot",
            "America/Martinique",
            "America/Matamoros",
```

```
            "America/Mazatlan",
            "America/Mendoza",
            "America/Menominee",
            "America/Merida",
            "America/Metlakatla",
            "America/Mexico_City",
            "America/Miquelon",
            "America/Moncton",
            "America/Monterrey",
            "America/Montevideo",
            "America/Montreal",
            "America/Montserrat",
            "America/Nassau",
            "America/New_York",
            "America/Nipigon",
            "America/Nome",
            "America/Noronha",
            "America/North_Dakota/Beulah",
            "America/North_Dakota/Center",
            "America/North_Dakota/New_Salem",
            "America/Nuuk",
            "America/Ojinaga",
            "America/Panama",
            "America/Pangnirtung",
            "America/Paramaribo",
            "America/Phoenix",
            "America/Port-au-Prince",
            "America/Port_of_Spain",
            "America/Porto_Acre",
            "America/Porto_Velho",
            "America/Puerto_Rico",
            "America/Punta_Arenas",
            "America/Rainy_River",
            "America/Rankin_Inlet",
            "America/Recife",
            "America/Regina",
            "America/Resolute",
            "America/Rio_Branco",
            "America/Rosario",
            "America/Santa_Isabel",
            "America/Santarem",
            "America/Santiago",
            "America/Santo_Domingo",
            "America/Sao_Paulo",
```

```
            "America/Scoresbysund",
            "America/Shiprock",
            "America/Sitka",
            "America/St_Barthelemy",
            "America/St_Johns",
            "America/St_Kitts",
            "America/St_Lucia",
            "America/St_Thomas",
            "America/St_Vincent",
            "America/Swift_Current",
            "America/Tegucigalpa",
            "America/Thule",
            "America/Thunder_Bay",
            "America/Tijuana",
            "America/Toronto",
            "America/Tortola",
            "America/Vancouver",
            "America/Virgin",
            "America/Whitehorse",
            "America/Winnipeg",
            "America/Yakutat",
            "America/Yellowknife",
            "Antarctica/Casey",
            "Antarctica/Davis",
            "Antarctica/DumontDUrville",
            "Antarctica/Macquarie",
            "Antarctica/Mawson",
            "Antarctica/McMurdo",
            "Antarctica/Palmer",
            "Antarctica/Rothera",
            "Antarctica/South_Pole",
            "Antarctica/Syowa",
            "Antarctica/Troll",
            "Antarctica/Vostok",
            "Arctic/Longyearbyen",
            "Asia/Aden",
            "Asia/Almaty",
            "Asia/Amman",
            "Asia/Anadyr",
            "Asia/Aqtau",
            "Asia/Aqtobe",
            "Asia/Ashgabat",
            "Asia/Ashkhabad",
            "Asia/Atyrau",
```

```
            "Asia/Baghdad",
            "Asia/Bahrain",
            "Asia/Baku",
            "Asia/Bangkok",
            "Asia/Barnaul",
            "Asia/Beirut",
            "Asia/Bishkek",
            "Asia/Brunei",
            "Asia/Calcutta",
            "Asia/Chita",
            "Asia/Choibalsan",
            "Asia/Chongqing",
            "Asia/Chungking",
            "Asia/Colombo",
            "Asia/Dacca",
            "Asia/Damascus",
            "Asia/Dhaka",
            "Asia/Dili",
            "Asia/Dubai",
            "Asia/Dushanbe",
            "Asia/Famagusta",
            "Asia/Gaza",
            "Asia/Harbin",
            "Asia/Hebron",
            "Asia/Ho_Chi_Minh",
            "Asia/Hong_Kong",
            "Asia/Hovd",
            "Asia/Irkutsk",
            "Asia/Istanbul",
            "Asia/Jakarta",
            "Asia/Jayapura",
            "Asia/Jerusalem",
            "Asia/Kabul",
            "Asia/Kamchatka",
            "Asia/Karachi",
            "Asia/Kashgar",
            "Asia/Kathmandu",
            "Asia/Katmandu",
            "Asia/Khandyga",
            "Asia/Kolkata",
            "Asia/Krasnoyarsk",
            "Asia/Kuala_Lumpur",
            "Asia/Kuching",
            "Asia/Kuwait",
```

```
            "Asia/Macao",
            "Asia/Macau",
            "Asia/Magadan",
            "Asia/Makassar",
            "Asia/Manila",
            "Asia/Muscat",
            "Asia/Nicosia",
            "Asia/Novokuznetsk",
            "Asia/Novosibirsk",
            "Asia/Omsk",
            "Asia/Oral",
            "Asia/Phnom_Penh",
            "Asia/Pontianak",
            "Asia/Pyongyang",
            "Asia/Qatar",
            "Asia/Qostanay",
            "Asia/Qyzylorda",
            "Asia/Rangoon",
            "Asia/Riyadh",
            "Asia/Saigon",
            "Asia/Sakhalin",
            "Asia/Samarkand",
            "Asia/Seoul",
            "Asia/Shanghai",
            "Asia/Singapore",
            "Asia/Srednekolymsk",
            "Asia/Taipei",
            "Asia/Tashkent",
            "Asia/Tbilisi",
            "Asia/Tehran",
            "Asia/Tel_Aviv",
            "Asia/Thimbu",
            "Asia/Thimphu",
            "Asia/Tokyo",
            "Asia/Tomsk",
            "Asia/Ujung_Pandang",
            "Asia/Ulaanbaatar",
            "Asia/Ulan_Bator",
            "Asia/Urumqi",
            "Asia/Ust-Nera",
            "Asia/Vientiane",
            "Asia/Vladivostok",
            "Asia/Yakutsk",
            "Asia/Yangon",
```

```
            "Asia/Yekaterinburg",
            "Asia/Yerevan",
            "Atlantic/Azores",
            "Atlantic/Bermuda",
            "Atlantic/Canary",
            "Atlantic/Cape_Verde",
            "Atlantic/Faeroe",
            "Atlantic/Faroe",
            "Atlantic/Jan_Mayen",
            "Atlantic/Madeira",
            "Atlantic/Reykjavik",
            "Atlantic/South_Georgia",
            "Atlantic/St_Helena",
            "Atlantic/Stanley",
            "Australia/ACT",
            "Australia/Adelaide",
            "Australia/Brisbane",
            "Australia/Broken_Hill",
            "Australia/Canberra",
            "Australia/Currie",
            "Australia/Darwin",
            "Australia/Eucla",
            "Australia/Hobart",
            "Australia/LHI",
            "Australia/Lindeman",
            "Australia/Lord_Howe",
            "Australia/Melbourne",
            "Australia/NSW",
            "Australia/North",
            "Australia/Perth",
            "Australia/Queensland",
            "Australia/South",
            "Australia/Sydney",
            "Australia/Tasmania",
            "Australia/Victoria",
            "Australia/West",
            "Australia/Yancowinna",
            "Brazil/Acre",
            "Brazil/DeNoronha",
            "Brazil/East",
            "Brazil/West",
            "CET",
            "CST6CDT",
            "Canada/Atlantic",
```

```
            "Canada/Central",
            "Canada/Eastern",
            "Canada/Mountain",
            "Canada/Newfoundland",
            "Canada/Pacific",
            "Canada/Saskatchewan",
            "Canada/Yukon",
            "Chile/Continental",
            "Chile/EasterIsland",
            "Cuba",
            "EET",
            "EST5EDT",
            "Egypt",
            "Eire",
            "Etc/GMT",
            "Etc/GMT+0",
            "Etc/GMT+1",
            "Etc/GMT+10",
            "Etc/GMT+11",
            "Etc/GMT+12",
            "Etc/GMT+2",
            "Etc/GMT+3",
            "Etc/GMT+4",
            "Etc/GMT+5",
            "Etc/GMT+6",
            "Etc/GMT+7",
            "Etc/GMT+8",
            "Etc/GMT+9",
            "Etc/GMT-0",
            "Etc/GMT-1",
            "Etc/GMT-10",
            "Etc/GMT-11",
            "Etc/GMT-12",
            "Etc/GMT-13",
            "Etc/GMT-14",
            "Etc/GMT-2",
            "Etc/GMT-3",
            "Etc/GMT-4",
            "Etc/GMT-5",
            "Etc/GMT-6",
            "Etc/GMT-7",
            "Etc/GMT-8",
            "Etc/GMT-9",
            "Etc/GMT0",
```

```
            "Etc/Greenwich",
            "Etc/UCT",
            "Etc/UTC",
            "Etc/Universal",
            "Etc/Zulu",
            "Europe/Amsterdam",
            "Europe/Andorra",
            "Europe/Astrakhan",
            "Europe/Athens",
            "Europe/Belfast",
            "Europe/Belgrade",
            "Europe/Berlin",
            "Europe/Bratislava",
            "Europe/Brussels",
            "Europe/Bucharest",
            "Europe/Budapest",
            "Europe/Busingen",
            "Europe/Chisinau",
            "Europe/Copenhagen",
            "Europe/Dublin",
            "Europe/Gibraltar",
            "Europe/Guernsey",
            "Europe/Helsinki",
            "Europe/Isle_of_Man",
            "Europe/Istanbul",
            "Europe/Jersey",
            "Europe/Kaliningrad",
            "Europe/Kiev",
            "Europe/Kirov",
            "Europe/Kyiv",
            "Europe/Lisbon",
            "Europe/Ljubljana",
            "Europe/London",
            "Europe/Luxembourg",
            "Europe/Madrid",
            "Europe/Malta",
            "Europe/Mariehamn",
            "Europe/Minsk",
            "Europe/Monaco",
            "Europe/Moscow",
            "Europe/Nicosia",
            "Europe/Oslo",
            "Europe/Paris",
            "Europe/Podgorica",
```

```
            "Europe/Prague",
            "Europe/Riga",
            "Europe/Rome",
            "Europe/Samara",
            "Europe/San_Marino",
            "Europe/Sarajevo",
            "Europe/Saratov",
            "Europe/Simferopol",
            "Europe/Skopje",
            "Europe/Sofia",
            "Europe/Stockholm",
            "Europe/Tallinn",
            "Europe/Tirane",
            "Europe/Tiraspol",
            "Europe/Ulyanovsk",
            "Europe/Uzhgorod",
            "Europe/Vaduz",
            "Europe/Vatican",
            "Europe/Vienna",
            "Europe/Vilnius",
            "Europe/Volgograd",
            "Europe/Warsaw",
            "Europe/Zagreb",
            "Europe/Zaporozhye",
            "Europe/Zurich",
            "GB",
            "GB-Eire",
            "GMT",
            "GMT0",
            "Greenwich",
            "Hongkong",
            "Iceland",
            "Indian/Antananarivo",
            "Indian/Chagos",
            "Indian/Christmas",
            "Indian/Cocos",
            "Indian/Comoro",
            "Indian/Kerguelen",
            "Indian/Mahe",
            "Indian/Maldives",
            "Indian/Mauritius",
            "Indian/Mayotte",
            "Indian/Reunion",
            "Iran",
```

```
            "Israel",
            "Jamaica",
            "Japan",
            "Kwajalein",
            "Libya",
            "MET",
            "MST7MDT",
            "Mexico/BajaNorte",
            "Mexico/BajaSur",
            "Mexico/General",
            "NZ",
            "NZ-CHAT",
            "Navajo",
            "PRC",
            "PST8PDT",
            "Pacific/Apia",
            "Pacific/Auckland",
            "Pacific/Bougainville",
            "Pacific/Chatham",
            "Pacific/Chuuk",
            "Pacific/Easter",
            "Pacific/Efate",
            "Pacific/Enderbury",
            "Pacific/Fakaofo",
            "Pacific/Fiji",
            "Pacific/Funafuti",
            "Pacific/Galapagos",
            "Pacific/Gambier",
            "Pacific/Guadalcanal",
            "Pacific/Guam",
            "Pacific/Honolulu",
            "Pacific/Johnston",
            "Pacific/Kanton",
            "Pacific/Kiritimati",
            "Pacific/Kosrae",
            "Pacific/Kwajalein",
            "Pacific/Majuro",
            "Pacific/Marquesas",
            "Pacific/Midway",
            "Pacific/Nauru",
            "Pacific/Niue",
            "Pacific/Norfolk",
            "Pacific/Noumea",
            "Pacific/Pago_Pago",
```

```
            "Pacific/Palau",
            "Pacific/Pitcairn",
            "Pacific/Pohnpei",
            "Pacific/Ponape",
            "Pacific/Port_Moresby",
            "Pacific/Rarotonga",
            "Pacific/Saipan",
            "Pacific/Samoa",
            "Pacific/Tahiti",
            "Pacific/Tarawa",
            "Pacific/Tongatapu",
            "Pacific/Truk",
            "Pacific/Wake",
            "Pacific/Wallis",
            "Pacific/Yap",
            "Poland",
            "Portugal",
            "ROK",
            "Singapore",
            "SystemV/AST4",
            "SystemV/AST4ADT",
            "SystemV/CST6",
            "SystemV/CST6CDT",
            "SystemV/EST5",
            "SystemV/EST5EDT",
            "SystemV/HST10",
            "SystemV/MST7",
            "SystemV/MST7MDT",
            "SystemV/PST8",
            "SystemV/PST8PDT",
            "SystemV/YST9",
            "SystemV/YST9YDT",
            "Turkey",
            "UCT",
            "US/Alaska",
            "US/Aleutian",
            "US/Arizona",
            "US/Central",
            "US/East-Indiana",
            "US/Eastern",
            "US/Hawaii",
            "US/Indiana-Starke",
            "US/Michigan",
            "US/Mountain",
```

```
            "US/Pacific",
            "US/Samoa",
            "UTC",
            "Universal",
            "W-SU",
            "WET",
            "Zulu",
            "EST",
            "HST",
            "MST",
            "ACT",
            "AET",
            "AGT",
            "ART",
            "AST",
            "BET",
            "BST",
            "CAT",
            "CNT",
            "CST",
            "CTT",
            "EAT",
            "ECT",
            "IET",
            "IST",
            "JST",
            "MIT",
            "NET",
            "NST",
            "PLT",
            "PNT",
            "PRT",
            "PST",
            "SST",
            "VST"
          ]
        },
        "pageRootPaths":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
```

```
        },
        "assetRootPaths":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "crawlAssets":
        {
          "type": "boolean"
        },
        "crawlPages":
        {
          "type": "boolean"
        },
        "pagePathInclusionPatterns":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "pagePathExclusionPatterns":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "pageNameInclusionPatterns":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "pageNameExclusionPatterns":
        {
          "type": "array",
```

```
          "items":
          {
            "type": "string"
          }
        },
        "assetPathInclusionPatterns":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "assetPathExclusionPatterns":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "assetTypeInclusionPatterns":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "assetTypeExclusionPatterns":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "assetNameInclusionPatterns":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
```

```
        },
        "assetNameExclusionPatterns":
        {
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "pageComponents": {
          "type": "array",
          "items": {
            "type": "object"
            }
        },
        "contentFragmentVariations": {
          "type": "array",
          "items": {
            "type": "object"
          }
        },
        "cugExemptedPrincipals": {
          "type": "array",
          "items": {
            "type": "string"
          }
        }
      },
      "required":
      []
    },
    "type": {
      "type": "string",
      "pattern": "AEM"
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
```

```
      ]
    },
    "secretArn": {
      "type": "string",
      "minLength": 20,
      "maxLength": 2048
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "syncMode",
    "additionalProperties",
    "secretArn",
    "type"
  ]
}
```

Amazon FSx (Windows) 模板架构

您可以将包含数据来源架构的 JSON 作为 TemplateConfiguration 对象的一部分。您可以在连接配置或存储库端点详细信息中提供文件系统 ID。您还必须将数据源的类型指定为FSX、身份验证凭证的密钥以及其他必要的配置。然后，当您 CreateDataSource 时，您可以将 TEMPLATE 指定为 Type。

您可以使用本开发者指南中提供的模板。请参阅 Amazon FSx (Windows) JSON 架构。

下表描述了 Amazon FSx (Windows) JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。

配置	描述
fileSystemId	Amazon FSx 文件系统的标识符。您可以在控制台的 “文件系统” Amazon FSx 控制面板上找到您的文件系统 ID。
fileSystemType	Amazon FSx 文件系统类型。要Windows File Server用作您的文件系统类型，请指定`WINDOWS`。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
全部	将 Amazon FSx 数据源中文件的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。
isCrawlAcl	`true`如果您有 ACL 并想将其用于访问控制，则可以抓取文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
inclusionPatterns	用于在 Amazon FSx 数据源中包含某些文件的正则表达式模式列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
exclusionPatterns	用于排除 Amazon FSx 数据源中某些文件的正则表达式模式列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配排除和包含模式，则以排除模式为优先，该文件不会包含在索引中。

配置	描述
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • `FORCED_FULL_CRAWL` 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • `FULL_CRAWL` 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
type	数据来源的类型。对于 Windows 文件系统数据源，请指定`FSX`。

Amazon FSx (Windows) JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "fileSystemId": {
              "type": "string",
              "pattern": "fs-.*"
```

```
            },
            "fileSystemType": {
              "type": "string",
              "pattern": "WINDOWS"
            }
          },
          "required": ["fileSystemId", "fileSystemType"]
        }
      }
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "All": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": ["STRING", "STRING_LIST", "DATE"]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
```

```
            }
          },
          "required": ["fieldMappings"]
        }
      },
      "required": ["All"]
    },
    "additionalProperties": {
      "type": "object",
      "properties": {
        "isCrawlAcl": {
          "type": "boolean"
        },
        "exclusionPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        }
      },
      "required": []
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL"
      ]
    },
    "type" : {
      "type" : "string",
      "pattern": "FSX"
    }
  },
  "version": {
```

```
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "syncMode",
    "enableIdentityCrawler",
    "additionalProperties",
    "type"
  ]
}
```

Amazon FSx (NetApp ONTAP) 模板架构

您可以将包含数据来源架构的 JSON 作为 TemplateConfiguration 对象的一部分。您可以在连接配置或存储库端点详细信息中提供文件系统 ID 和存储虚拟机 (SVM)。您还必须将数据源的类型指定为FSXONTAP、身份验证凭证的密钥以及其他必要的配置。然后，当您 CreateDataSource 时，您可以将 TEMPLATE 指定为 Type。

您可以使用本开发者指南中提供的模板。请参阅 Amazon FSx (NetApp ONTAP) JSON 架构。

下表描述了 Amazon FSx (NetApp ONTAP) JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
fileSystemId	Amazon FSx 文件系统的标识符。您可以在控制台的 “文件系统” Amazon FSx 控制面板上找到您的文件系统 ID。有关如何在 Amazon FSx 控制台中为 NetApp ONTAP 创建文件系统的信息，请参阅《FSx for ONTAP 用户指南》中的 NetApp ONTAP 入门指南。

配置	描述
fileSystemType	Amazon FSx 文件系统类型。要NetApp ONTAP用作您的文件系统类型，请指定ONTAP。
svMid	用于 Amazon FSx 文件系统的存储虚拟机 (SVM) 的标识符。NetApp ONTAP您可以前往控制台中的 “文件系统” Amazon FSx 控制面板，选择您的文件系统 ID，然后选择 “存储虚拟机”，找到您的 SVM ID。有关如何在 Amazon FSx 控制台中为创建文件系统的信息NetApp ONTAP，请参阅《FSx for ONTAP 用户指南》中的 NetApp ONTAP 入门指南。
协议类型	无论你使用适用于 Windows 的通用互联网文件系统 (CIFS) 协议，还是使用适用于 Linux 的网络文件系统 (NFS) 协议。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
file	将 Amazon FSx 数据源中文件的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。数据源字段名称必须存在于文件的自定义元数据中。
additionalProperties	数据来源中内容的其他配置选项。
crawlAcl	`true`如果您有 ACL 并想将其用于访问控制，则可以抓取文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

配置	描述
inclusionPatterns	用于在 Amazon FSx 数据源中包含某些文件的正则表达式模式列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
exclusionPatterns	用于排除 Amazon FSx 数据源中某些文件的正则表达式模式列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配排除和包含模式，则以排除模式为优先，该文件不会包含在索引中。
type	数据来源的类型。对于NetApp ONTAP文件系统数据源，请指定FSXONTAP。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

配置	描述

secretArn	包含连接到您的文件系统所需的键值对的 AWS Secrets Manager 密钥的 Amazon 资源名称 (ARN)。 Amazon FSx 密钥必须包含具有以下键的 JSON 结构：
	
	{
	"username": " *user@corp.example.com* ",
	"password": " *password*"
	}
	
	如果您对 Amazon FSx 文件系统使用 NFS 协议，则密钥将存储在 JSON 结构中，其中包含以下密钥：
	
	{
	"leftId": "*left ID*",
	"rightId": " *right ID*",
	"preSharedKey": " *pre-shared key* "
	}

Amazon FSx (NetApp ONTAP) JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "fileSystemId": {
              "type": "string",
                "pattern": "^(fs-[0-9a-f]{8,21})$"
            },
```

```
            "fileSystemType": {
              "type": "string",
              "enum": ["ONTAP"]
            },
            "svmId": {
              "type": "string",
              "pattern": "^(svm-[0-9a-f]{17,21})$"
            },
            "protocolType": {
              "type": "string",
              "enum": [
                "CIFS",
                "NFS"
              ]
            }
          },
          "required": [
            "fileSystemId",
            "fileSystemType"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "file": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string",
                      "pattern": "^([a-zA-Z_]{1,20})$"
                    },
                    "indexFieldType": {
                      "type": "string",
```

```
                    "enum": [
                      "STRING",
                      "STRING_LIST",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string",
                    "pattern": "^([a-zA-Z_]{1,20})$"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ],
            "maxItems": 50
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
      "file"
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "crawlAcl": {
        "type": "boolean"
      },
      "inclusionPatterns": {
        "type": "array",
        "items": {
```

```
            "type": "string",
            "maxLength": 30
          },
          "maxItems": 100
        },
        "exclusionPatterns": {
          "type": "array",
          "items": {
            "type": "string",
            "maxLength": 30
          },
          "maxItems": 100
        }
      }
    },
    "type": {
      "type": "string",
      "pattern": "FSXONTAP"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL"
      ]
    },
    "secretArn": {
      "type": "string",
      "pattern": "arn:aws:secretsmanager:.*"
    }
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "additionalProperties",
    "secretArn",
    "type"
  ]
}
```

Alfresco 模板架构

您可以将包含数据来源架构的 JSON 作为 TemplateConfiguration 对象的一部分。您需要提供 Alfresco 站点 ID、存储库 URL、用户界面 URL、身份验证类型，您是使用云还是本地部署，以及要爬取的内容类型。您可以将其作为连接配置或存储库端点详细信息的一部分提供。还要将数据来源的类型指定为 ALFRESCO、身份验证凭证的密钥以及其他必要的配置。然后，当您 CreateDataSource 时，您可以将 TEMPLATE 指定为 Type。

您可以使用本开发者指南中提供的模板。请参阅 Alfresco JSON 架构。

下表描述了 Alfresco JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
siteId	Alfresco 站点的标识符。
repoUrl	您的 Alfresco 存储库的 URL。您可以向 Alfresco 管理员获取存储库 URL。例如，如果您使用 Alfresco Cloud（PaaS），则存储库 URL 可能是 https://company.alfrescocloud.com。或者，如果您使用 Alfresco On-Premises，则存储库 URL 可能是 https://company-alfresco-instance.company-domain.suffix:port。
webAppUrl	您的 Alfresco 用户界面的 URL。您可以向 Alfresco 管理员获取 Alfresco 用户界面 URL。例如，用户界面 URL 可能是 https://example.com。
repositoryAdditionalProperties	用于连接存储库/数据来源端点的其他属性。
authType	您使用的身份验证类型，可以是 OAuth2 或 Basic。
type (deployment)	您使用的 Alfresco 的类型，可以是 PAAS 或 ON-PREM。

配置	描述
crawlType	您要爬取的内容类型，可以是 ASPECT（Alfresco 中标有“方面”的内容）、SITE_ID（特定 Alfresco 网站内的内容）或 ALL_SITES （所有 Alfresco 网站上的内容）。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• 文档 • comment	将您的 Alfresco 文档和注释的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。
aspectName	要编制索引的特定“方面”的名称。
aspectProperties	要编制索引的特定“方面”内容属性的列表。
enableFineGrained控制	如果为 true，要爬取“方面”。
isCrawlComment	true搜寻评论。
• inclusionFileName图案 • inclusionFileType图案 • inclusionFilePath图案	用于在 Alfresco 数据来源中包含某些文件的正则表达式模式的列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
• exclusionFileName图案 • exclusionFileType图案 • exclusionFilePath图案	用于在 Alfresco 数据来源中排除某些文件的正则表达式模式的列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
type	数据来源的类型。指定 ALFRESCO 作为数据来源类型。

配置	描述

secretArn	AWS Secrets Manager 密钥的 Amazon 资源名称 (ARN)，其中包含连接到您的所需的键值对。Alfresco密钥必须包含具有以下键的 JSON 结构：
	
	如果使用基本身份验证：
	
	`{ "username": " user name", "password": " password" }`
	
	如果使用 OAuth 2.0 身份验证：
	
	`{ "clientId": " client ID", "clientSecret": " client secret", "tokenUrl": " token URL" }`

syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

配置	描述
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。
版本	当前支持的此模板的版本。

Alfresco JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "siteId": {
              "type": "string"
            },
            "repoUrl": {
              "type": "string"
            },
            "webAppUrl": {
              "type": "string"
            },
            "repositoryAdditionalProperties": {
              "type": "object",
              "properties": {
                "authType": {
                  "type": "string",
                  "enum": [
                    "OAuth2",
                    "Basic"
                  ]
```

```
              },
              "type": {
                "type": "string",
                "enum": [
                  "PAAS",
                  "ON_PREM"
                ]
              },
              "crawlType": {
                "type": "string",
                "enum": [
                  "ASPECT",
                  "SITE_ID",
                  "ALL_SITES"
                ]
              }
            }
          }
        }
      }
    },
    "required": [
      "repositoryEndpointMetadata"
    ]
  },
  "repositoryConfigurations": {
    "type": "object",
    "properties": {
      "document": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": {
              "anyOf": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
```

```
                    "STRING",
                    "DATE",
                    "STRING_LIST",
                    "LONG"
                  ]
                },
                "dataSourceFieldName": {
                  "type": "string"
                },
                "dateFieldFormat": {
                  "type": "string",
                  "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                }
              },
              "required": [
                "indexFieldName",
                "indexFieldType",
                "dataSourceFieldName"
              ]
            }
          ]
        }
      }
    },
    "required": [
      "fieldMappings"
    ]
  },
  "comment": {
    "type": "object",
    "properties": {
      "fieldMappings": {
        "type": "array",
        "items": {
          "anyOf": [
            {
              "type": "object",
              "properties": {
                "indexFieldName": {
                  "type": "string"
                },
                "indexFieldType": {
                  "type": "string",
                  "enum": [
```

```
                        "STRING",
                        "DATE",
                        "STRING_LIST",
                        "LONG"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    }
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "aspectName": {
        "type": "string"
      },
      "aspectProperties": {
        "type": "array"
      },
      "enableFineGrainedControl": {
        "type": "boolean"
      },
      "isCrawlComment": {
        "type": "boolean"
```

```
        },
        "inclusionFileNamePatterns": {
          "type": "array"
        },
        "exclusionFileNamePatterns": {
          "type": "array"
        },
        "inclusionFileTypePatterns": {
          "type": "array"
        },
        "exclusionFileTypePatterns": {
          "type": "array"
        },
        "inclusionFilePathPatterns": {
          "type": "array"
        },
        "exclusionFilePathPatterns": {
          "type": "array"
        }
      }
    },
    "type": {
      "type": "string",
      "pattern": "ALFRESCO"
    },
    "secretArn": {
      "type": "string",
      "minLength": 20,
      "maxLength": 2048
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL"
      ]
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "version": {
      "type": "string",
      "anyOf": [
        {
```

```
          "pattern": "1.0.0"
        }
      ]
    }
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "additionalProperties",
    "type",
    "secretArn"
  ]
}
```

Aurora (MySQL) 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、mysql 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Aurora (MySQL) JSON 架构。

下表描述了 Aurora (MySQL) JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。 • dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。

配置	描述
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。

配置	描述
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

Aurora (MySQL) JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
```

```
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string"
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        }
      },
      "required": [
      ]
    },
    "additionalProperties": {
      "type": "object",
      "properties": {
        "primaryKey": {
          "type": "string"
```

```
        },
        "titleColumn": {
          "type": "string"
        },
        "bodyColumn": {
          "type": "string"
        },
        "sqlQuery": {
          "type": "string",
          "not": {
            "pattern": ";+"
          }
        },
        "timestampColumn": {
          "type": "string"
        },
        "timestampFormat": {
          "type": "string"
        },
        "timezone": {
          "type": "string"
        },
        "changeDetectingColumns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
```

```
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Aurora (PostgreSQL) 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、postgresql 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Aurora (PostgreSQL) JSON 架构。

下表描述了 (PostgreSQL Aurora) JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。 • dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。

配置	描述
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

Aurora (PostgreSQL) JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
```

```
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
```

```
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string"
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "primaryKey": {
        "type": "string"
      },
      "titleColumn": {
        "type": "string"
      },
      "bodyColumn": {
        "type": "string"
      },
      "sqlQuery": {
```

```
          "type": "string",
          "not": {
            "pattern": ";+"
          }
        },
        "timestampColumn": {
          "type": "string"
        },
        "timestampFormat": {
          "type": "string"
        },
        "timezone": {
          "type": "string"
        },
        "changeDetectingColumns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
```

```
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Amazon RDS （微软 SQL Server ）模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、sqlserver 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Amazon RDS （微软 SQL Server ）JSON 架构。

下表描述了 Amazon RDS （微软 SQL Server ）JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。

配置	描述
	• dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。

配置	描述
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

配置	描述
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

Amazon RDS（微软 SQL Server）JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
```

```
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string"
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
```

```
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "primaryKey": {
        "type": "string"
      },
      "titleColumn": {
        "type": "string"
      },
      "bodyColumn": {
        "type": "string"
      },
      "sqlQuery": {
        "type": "string",
        "not": {
          "pattern": ";+"
        }
      },
      "timestampColumn": {
        "type": "string"
      },
      "timestampFormat": {
        "type": "string"
      },
      "timezone": {
        "type": "string"
      },
      "changeDetectingColumns": {
        "type": "array",
        "items": {
          "type": "string"
        }
```

```
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
```

```
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Amazon RDS (MySQL) 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、mysql 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Amazon RDS (MySQL) JSON 架构。

下表描述了 Amazon RDS (MySQL) JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。 • dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。

配置	描述
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

Amazon RDS (MySQL) JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
```

```
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
```

```
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string"
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "primaryKey": {
        "type": "string"
      },
      "titleColumn": {
        "type": "string"
      },
      "bodyColumn": {
        "type": "string"
      },
      "sqlQuery": {
```

```
          "type": "string",
          "not": {
            "pattern": ";+"
          }
        },
        "timestampColumn": {
          "type": "string"
        },
        "timestampFormat": {
          "type": "string"
        },
        "timezone": {
          "type": "string"
        },
        "changeDetectingColumns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
```

```
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Amazon RDS (甲骨文) 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、oracle 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Amazon RDS (甲骨文) JSON 架构。

下表描述了 Amazon RDS (Oracle) JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。

配置	描述
	• dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。

配置	描述
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

配置	描述
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

Amazon RDS (甲骨文) JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
```

```
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string"
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
```

```
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "primaryKey": {
        "type": "string"
      },
      "titleColumn": {
        "type": "string"
      },
      "bodyColumn": {
        "type": "string"
      },
      "sqlQuery": {
        "type": "string",
        "not": {
          "pattern": ";+"
        }
      },
      "timestampColumn": {
        "type": "string"
      },
      "timestampFormat": {
        "type": "string"
      },
      "timezone": {
        "type": "string"
      },
      "changeDetectingColumns": {
        "type": "array",
        "items": {
          "type": "string"
        }
```

```
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
```

```
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Amazon RDS (PostgreSQL) 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、postgresql 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Amazon RDS (PostgreSQL) JSON 架构。

下表描述了 (PostgreSQL Amazon RDS) JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。 • dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。

配置	描述
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

Amazon RDS (PostgreSQL) JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
```

```
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
```

```
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string"
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "primaryKey": {
        "type": "string"
      },
      "titleColumn": {
        "type": "string"
      },
      "bodyColumn": {
        "type": "string"
      },
      "sqlQuery": {
```

```
          "type": "string",
          "not": {
            "pattern": ";+"
          }
        },
        "timestampColumn": {
          "type": "string"
        },
        "timestampFormat": {
          "type": "string"
        },
        "timezone": {
          "type": "string"
        },
        "changeDetectingColumns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
```

```
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Amazon S3 模板架构

您可以将包含数据来源架构的 JSON 作为模板配对的一部分。您可以将其作为连接配置或存储库端点详细信息的一部分提供 S3 存储桶的名称。还要将数据来源的类型指定为 S3，以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 S3 JSON 架构。

下表描述了 Amazon S3 JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
BucketName	您的 Amazon S3 存储桶的名称。

配置	描述
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
additionalProperties	数据来源中内容的其他配置选项
• inclusionPatterns • exclusionPatterns • inclusionPrefixes • exclusionPrefixes	用于在 Amazon S3 数据源中包含或排除特定文件的正则表达式模式列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
aclConfigurationFile路径	控制对 Amazon Kendra 索引中文档的访问权限的文件路径。
metadataFilesPrefix	存储桶中存放元数据文件的位置。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
type	数据来源的类型。指定 S3 作为数据来源类型。
版本	支持的模板的版本。

S3 JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
```

```
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "BucketName": {
              "type": "string"
            }
          },
          "required": [
            "BucketName"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    }
```

```
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
      "document"
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "inclusionPatterns": {
        "type": "array"
      },
      "exclusionPatterns": {
        "type": "array"
      },
      "inclusionPrefixes": {
        "type": "array"
      },
      "exclusionPrefixes": {
        "type": "array"
      },
      "aclConfigurationFilePath": {
        "type": "string"
      },
      "metadataFilesPrefix": {
        "type": "string"
      }
    }
  },
  "syncMode": {
    "type": "string",
```

```
      "enum": [
        "FULL_CRAWL",
        "FORCED_FULL_CRAWL"
      ]
    },
    "type": {
      "type": "string",
      "pattern": "S3"
    },
    "version": {
      "type": "string",
      "anyOf": [
        {
          "pattern": "1.0.0"
        }
      ]
    }
  },
  "required": [
    "connectionConfiguration",
    "type",
    "syncMode",
    "repositoryConfigurations"
  ]
}
```

Amazon Kendra Web Crawler 模板架构

您可以将包含数据来源架构的 JSON 作为 TemplateConfiguration 对象的一部分。

您可以将其作为连接配置或存储库端点详细信息的一部分提供种子或起点 URL，也可以提供站点地图 URL。与其手动列出所有 URL，不如提供存储种子网址列表的文本文件的 Amazon S3 存储桶路径或站点地图 XML 文件，您可以在 S3 中将这些文件组合成一个 ZIP 文件。

您还可以将数据源的类型指定为WEBCRAWLERV2、网站身份验证凭证和身份验证类型（如果您的网站需要身份验证）以及其他必要的配置。

然后，当您 CreateDataSource 时，您可以将 TEMPLATE 指定为 Type。

> **Important**
>
> 不支持 Web Crawler v2.0 连接器的创建。 AWS CloudFormation如果需要 AWS CloudFormation 支持，请使用 Web Crawler v1.0 连接器。

当选择要编制索引的网站时，您必须遵守 Amazon 可接受使用政策以及所有其他 Amazon 条款。请记住，您只能使用 Amazon Kendra Web Crawler 来索引自己的网页或您有权编制索引的网页。要了解如何阻止 Amazon Kendra Web 爬网程序将您的网站编入索引，请参阅为 Amazon Kendra Web 爬网程序配置 *robots.txt* 文件。

您可以使用本开发者指南中提供的模板。请参阅 Amazon Kendra 网络爬虫 JSON 架构。

下表描述了 Amazon Kendra Web Crawler JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
siteMapUrls	要爬取的网站站点地图 URL 的列表。您最多可以列出三个站点地图 URL。
s3 SeedUrl	存储种子 URL 或启动 URL 列表的文本文件的 S3 路径。例如，s3://bucket-name/directory/。文本文件中的每个 URL 都必须对单行进行格式化。在一个文件中最多可以列出 100 个种子 URL。
s3 SiteMapUrl	站点地图 XML 文件的 S3 路径。例如，s3://bucket-name/directory/。您最多可以列出三个站点地图 XML 文件。您可以将多个站点地图文件组合成一个 ZIP 文件，然后将 ZIP 文件存储在存储 Amazon S3 桶中。
seedUrlConnections	您想要爬取的网站的种子或起点 URL 的列表。您最多可以列出 100 个种子 URL。

配置	描述
seedUrl	种子或起点 URL。
身份验证	如果您的网站需要相同的身份验证，则为身份验证类型，否则指定 NoAuthentication 。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• webPage • attachment	将网页和网页文件的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。例如，HTML 网页标题标签可以映射到 _document_title 索引字段。有关更多信息，请参阅映射数据来源字段。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
additionalProperties	数据来源中内容的其他配置选项。
rateLimit	每分钟爬取的每个网站主机的最大 URL 数量。
maxFileSize	要爬取的网页或附件的最大大小（以 MB 为单位）。
crawlDepth	从种子 URL 爬取的层数。例如，种子 URL 页面的深度为 1，在该页面上同时爬取的所有超链接的深度都是 2。

配置	描述
maxLinksPer网址	爬取网站时要包含的网页 URL 的最大数量。每个网页都有一个数字。在爬取网站的网页时，网页链接的任何 URL 也会被爬取。按显示顺序爬取网页上的 URL。
crawlSubDomain	`true`，爬取包含子域的网站域。例如，如果种子 URL 是“abc.example.com”，则还会爬取“a.abc.example.com”和“b.abc.example.com”。如果您未设置`crawlSubDomain` 或 `crawlAllDomain` `true`，则 Amazon Kendra 只会抓取您要抓取的网站的域名。
crawlAllDomain	`true`，爬取有子域的网站域和网页链接到的其他域。如果您未设置`crawlSubDomain` 或 `crawlAllDomain` `true`，则 Amazon Kendra 只会抓取您要抓取的网站的域名。
honorRobots	`true`，遵循您想要爬取的网站的 robots.txt 指令。这些指令控制 Amazon Kendra Web Crawler 如何抓取网站，无论是 Amazon Kendra 只能抓取特定内容还是不能抓取任何内容。
crawlAttachments	`true`，爬取网页链接到的文件。
• 包含网址 CrawlPatterns • 包含网址 IndexPatterns	正则表达式模式的列表，以便包含爬取某些 URL，并为这些 URL 网页上的任何超链接编制索引。与模式匹配的 URL 将包含在索引中。与模式不匹配的 URL 将从索引中排除。如果 URL 同时匹配包含和排除模式，则以排除模式为优先，并且该 URL/网站的网页不会包含在索引中。

配置	描述
• 排除网址 CrawlPatterns • 排除网址 IndexPatterns	正则表达式模式的列表，以便排除爬取某些 URL，并为这些 URL 网页上的任何超链接编制索引。与模式匹配的 URL 将从索引中排除。与模式不匹配的 URL 将包含在索引中。如果 URL 同时匹配包含和排除模式，则以排除模式为优先，并且该 URL/网站的网页不会包含在索引中。
inclusionFileIndex图案	正则表达式模式的列表，用于包含某些网页文件。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
exclusionFileIndex图案	正则表达式模式的列表，用于排除某些网页文件。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
proxy	通过 Web 代理连接到内部网站所需的配置信息。
host	您想要通过用于连接内部网站的代理服务器的主机名。例如，https://a.example.com/page1.html 的主机名是“a.example.com”。
port	您想要用于连接内部网站的代理服务器的端口号。例如，443 是 HTTPS 的标准端口。
secretArn（代理）	如果需要网络代理凭据才能连接到网站主机，则可以创建一个存储凭据的 AWS Secrets Manager 密钥。为密钥提供 Amazon 资源名称（ARN）。

配置	描述
type	数据来源的类型。指定 `WEBCRAWLERV2` 作为数据来源类型。
secretArn	您的网站需要身份验证才能访问网站时使用的 AWS Secrets Manager 密钥的 Amazon 资源名称 (ARN)。网站的身份验证凭证存储在包含 JSON 键值对的密钥中。 如果您使用基本或 NTML/Kerberos 身份验证，请输入用户名和密码。密钥中的 JSON 键必须是 `userName` 和 `password`。NTLM 身份验证协议包括密码哈希，Kerberos 身份验证协议包括密码加密。 如果您使用 SAML 或表单身份验证，请输入用户名和密码，在用户名字段中输入 XPath（如果使用 SAML，则输入用户名按钮），在密码字段和按钮中输入 XPath，以及登录页面 URL。密钥中的 JSON 键必须是 `userName`、`password`、`userNameFieldXpath` 、`userNameButtonXpath` 、`passwordFieldXpath` 、`passwordButtonXpath` 和 `loginPageUrl` 。您可以使用 Web 浏览器的开发者工具找到元素的 XPaths（XML 路径语言）。XPaths 通常遵循以下格式：`//tagname[@Attribute='Value']` . Amazon Kendra 还会检查密钥中包含的端点信息（种子 URL）是否与您的数据源终端节点配置详细信息中指定的端点信息相同。
版本	当前支持的此模板的版本。

Amazon Kendra 网络爬虫 JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "siteMapUrls": {
              "type": "array",
              "items":{
                "type": "string",
                "pattern": "https://.*"
              }
            },
            "s3SeedUrl": {
              "type": "string",
              "pattern": "s3:.*"
            },
            "s3SiteMapUrl": {
              "type": "string",
              "pattern": "s3:.*"
            },
            "seedUrlConnections": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "seedUrl":{
                      "type": "string",
                      "pattern": "https://.*"
                    }
                  },
                  "required": [
                    "seedUrl"
                  ]
                }
              ]
            },
```

```
            "authentication": {
              "type": "string",
              "enum": [
                "NoAuthentication",
                "BasicAuth",
                "NTLM_Kerberos",
                "Form",
                "SAML"
              ]
            }
          }
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "webPage": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING",
                        "DATE",
                        "LONG"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
```

```
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "attachment": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
```

```
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    }
  },
  "syncMode": {
    "type": "string",
    "enum": [
      "FORCED_FULL_CRAWL",
      "FULL_CRAWL"
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "rateLimit": {
        "type": "string",
        "default": "300"
      },
      "maxFileSize": {
        "type": "string",
        "default": "50"
      },
      "crawlDepth": {
        "type": "string",
        "default": "2"
      },
      "maxLinksPerUrl": {
        "type": "string",
        "default": "100"
      },
      "crawlSubDomain": {
        "type": "boolean",
```

```
          "default": false
        },
        "crawlAllDomain": {
          "type": "boolean",
          "default": false
        },
        "honorRobots": {
          "type": "boolean",
          "default": false
        },
        "crawlAttachments": {
          "type": "boolean",
          "default": false
        },
        "inclusionURLCrawlPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionURLCrawlPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionURLIndexPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionURLIndexPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionFileIndexPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
```

```
        "exclusionFileIndexPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "proxy": {
          "type": "object",
          "properties": {
            "host": {
              "type": "string"
            },
            "port": {
              "type": "string"
            },
            "secretArn": {
              "type": "string",
              "minLength": 20,
              "maxLength": 2048
            }
          }
        }
      },
      "required": [
        "rateLimit",
        "maxFileSize",
        "crawlDepth",
        "crawlSubDomain",
        "crawlAllDomain",
        "maxLinksPerUrl",
        "honorRobots"
      ]
    },
    "type": {
      "type": "string",
      "pattern": "WEBCRAWLERV2"
    },
    "secretArn": {
      "type": "string",
      "minLength": 20,
      "maxLength": 2048
    }
  },
  "version": {
```

```
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "syncMode",
    "type",
    "additionalProperties"
  ]
}
```

Confluence 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您可以在连接配置或存储库端点详细信息中提供 Confluence 主机 URL、托管方法和身份验证类型。还要将数据来源的类型指定为 CONFLUENCEV2、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Confluence JSON 架构。

下表描述了 Confluence JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
hostUrl	您的 Confluence 实例的 URL。例如，*https://example.confluence.com* 。
type	您的 Confluence 实例的托管方法，可以是 SAAS 和 ON_PREM。

配置	描述
authType	您的 Confluence 实例的身份验证方法，可以是 `Basic`、`OAuth2` 或 `Personal-token` 。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• space • 页 • blog • comment • attachment	将您的 Confluence 空间、页面、博客、评论和附件的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。您的 Confluence 自定义元数据中必须有 Confluence 数据来源字段名称。
additionalProperties	数据来源中内容的其他配置选项。
isCrawlAcl	`true`如果您有 ACL 并想将其用于访问控制，则可以抓取文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
fieldForUser我是	指定`email`是否要使用用户电子邮件作为用户 ID。 `email`默认使用，并且是目前唯一支持的用户 ID 类型。

配置	描述
• inclusionSpaceKey过滤器 • exclusionSpaceKey过滤器 • pageTitleRegEX • blogTitleRegEX • commentTitleRegEX • attachmentTitleRegEX • inclusionFileType图案 • exclusionFileType图案 • inclusionUrlPatterns • exclusionUrlPatterns	用于在 Confluence 数据来源中排除某些文件的正则表达式模式的列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
proxyHost	您使用的 Web 代理的主机名，不带 `http://`或`https://`协议。
proxyPort	主机 URL 传输协议使用的端口号。必须是介于 0 和 65535 之间的数值。
• isCrawlPersonal空间 • isCrawlArchived空间 • isCrawlArchived页面 • isCrawlPage • isCrawlBlog • isCrawlPage评论 • isCrawlPage附件 • isCrawlBlog评论 • isCrawlBlog附件	`true`抓取 Confluence 个人空间、页面、博客、页面评论、页面附件、博客评论和博客附件中的文件。
maxFileSizeInMegaBytes	指定 Amazon Kendra 可以抓取的文件大小限制（以 MB 为单位）。 Amazon Kendra 仅抓取您定义的大小限制内的文件。默认文件大小为 50MB。最大文件大小应大于 0MB 且小于或等于 50MB。

配置	描述
type	数据来源的类型。指定 CONFLUENCEV2 作为数据来源类型。
enableIdentityCrawler	true使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretARN	包含连接您的 Confluence 所需的键值对的 AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)。有关这些键值对的信息，请参阅 Confluence 的连接说明。
版本	当前支持的此模板的版本。

Confluence JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
```

```
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "hostUrl": {
              "type": "string",
              "pattern": "https:.*"
            },
            "type": {
              "type": "string",
              "enum": [
                "SAAS",
                "ON_PREM"
              ]
            },
            "authType": {
              "type": "string",
              "enum": [
                "Basic",
                "OAuth2",
                "Personal-token"
              ]
            }
          },
          "required": [
            "hostUrl",
            "type",
            "authType"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "space": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
```

```
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "STRING_LIST",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "page": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
```

```
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "STRING_LIST",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "blog": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
```

```
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "STRING_LIST",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "comment": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
```

```
                "indexFieldType": {
                  "type": "string",
                  "enum": [
                    "STRING",
                    "STRING_LIST",
                    "DATE",
                    "LONG"
                  ]
                },
                "dataSourceFieldName": {
                  "type": "string"
                },
                "dateFieldFormat": {
                  "type": "string",
                  "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                }
              },
              "required": [
                "indexFieldName",
                "indexFieldType",
                "dataSourceFieldName"
              ]
            }
          ]
        }
      },
      "required": [
        "fieldMappings"
      ]
    },
    "attachment": {
      "type": "object",
      "properties": {
        "fieldMappings": {
          "type": "array",
          "items": [
            {
              "type": "object",
              "properties": {
                "indexFieldName": {
                  "type": "string"
                },
                "indexFieldType": {
                  "type": "string",
```

```
                    "enum": [
                      "STRING",
                      "STRING_LIST",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    }
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "usersAclS3FilePath": {
        "type": "string"
      },
      "isCrawlAcl": {
        "type": "boolean"
      },
      "fieldForUserId": {
        "type": "string"
      },
      "inclusionSpaceKeyFilter": {
        "type": "array",
```

```
          "items": {
            "type": "string"
          }
        },
        "exclusionSpaceKeyFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "pageTitleRegEX": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "blogTitleRegEX": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "commentTitleRegEX": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "attachmentTitleRegEX": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "isCrawlPersonalSpace": {
          "type": "boolean"
        },
        "isCrawlArchivedSpace": {
          "type": "boolean"
        },
        "isCrawlArchivedPage": {
          "type": "boolean"
        },
        "isCrawlPage": {
```

```
          "type": "boolean"
        },
        "isCrawlBlog": {
          "type": "boolean"
        },
        "isCrawlPageComment": {
          "type": "boolean"
        },
        "isCrawlPageAttachment": {
          "type": "boolean"
        },
        "isCrawlBlogComment": {
          "type": "boolean"
        },
        "isCrawlBlogAttachment": {
          "type": "boolean"
        },
        "maxFileSizeInMegaBytes":  {
          "type":"string"
        },
        "inclusionFileTypePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionFileTypePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionUrlPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionUrlPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
```

```
        "proxyHost": {
          "type": "string"
        },
        "proxyPort": {
          "type": "string"
        }
      },
      "required": []
    },
    "type": {
      "type": "string",
      "pattern": "CONFLUENCEV2"
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FULL_CRAWL",
        "FORCED_FULL_CRAWL"
      ]
    },
    "secretArn": {
      "type": "string",
      "minLength": 20,
      "maxLength": 2048
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "syncMode",
    "additionalProperties",
    "secretArn",
    "type"
```

```
  ]
}
```

Dropbox 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您提供 Dropbox 应用程序键、应用程序密钥和访问令牌作为存储身份验证凭证的密钥的一部分。还要将数据来源的类型指定为 DROPBOX、要使用的访问令牌的类型（临时或永久）以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Dropbox JSON 模式。

下表描述了 Dropbox JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。此数据来源未在 `repositoryEndpointMetadata` 中指定端点。相反，连接信息包含在您提供的 AWS Secrets Manager 密钥中`secretArn` 。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• file • paper • papert • shortcut	映射您的 Dropbox 文件、Dropbox Paper 的属性或字段名称的对象列表，以及 Amazon Kendra 索引字段名称的快捷方式。有关更多信息，请参阅映射数据来源字段。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • `FORCED_FULL_CRAWL` 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • `FULL_CRAWL` 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。

配置	描述
	Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
enableIdentityCrawler	true使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。
secretARN	包含连接您的 Dropbox 所需的键值对的 AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须包含具有以下键的 JSON 结构： `{ "appKey": "Dropbox app key", "appSecret": " Dropbox app secret", "accesstoken": " temporary access token or refresh access token" }`
additionalProperties	数据来源中内容的其他配置选项。
isCrawlAcl	true如果您有 ACL 并想将其用于访问控制，则可以抓取文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

配置	描述
• inclusionFileName图案 • inclusionFileType图案	用于在 Dropbox 数据来源中包含某些文件名和类型的正则表达式模式的列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
• exclusionFileName图案 • exclusionFileType图案	用于在 Dropbox 数据来源中排除某些文件名和类型的正则表达式模式的列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配排除和包含模式，则以排除模式为优先，该文件不会包含在索引中。
• crawlFile • crawlPaper • crawlPapert • crawlShortcut	`true`抓取 Dropbox、Dropbox Paper 文档、Dropbox Paper 模板和存储在 Dropbox 中的网页快捷方式中的文件。
type	数据来源的类型。指定 `DROPBOX` 作为数据来源类型。
tokenType	指定您的访问令牌类型：永久或临时访问令牌。建议您创建在 Dropbox 中永不过期的刷新访问令牌，而不是依赖在 4 小时后过期的一次性访问令牌。您可以在 Dropbox 开发者控制台中创建应用程序和刷新访问令牌，并在密钥中提供访问令牌。
版本	当前支持的此模板的版本。

Dropbox JSON 模式

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
```

```
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
          }
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "file": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": {
                "anyOf": [
                  {
                    "type": "object",
                    "properties": {
                      "indexFieldName": {
                        "type": "string"
                      },
                      "indexFieldType": {
                        "type": "string",
                        "enum": [
                          "STRING",
                          "STRING_LIST",
                          "LONG",
                          "DATE"
                        ]
                      },
                      "dataSourceFieldName": {
                        "type": "string"
                      },
                      "dateFieldFormat": {
```

```
                    "type": "string",
                    "pattern": "dd-MM-yyyy HH:mm:ss"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        }
      },
      "required": [
        "fieldMappings"
      ]
    },
    "paper": {
      "type": "object",
      "properties": {
        "fieldMappings": {
          "type": "array",
          "items": {
            "anyOf": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "STRING_LIST",
                      "LONG",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
```

```
                    "type": "string",
                    "pattern": "dd-MM-yyyy HH:mm:ss"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        }
      },
      "required": [
        "fieldMappings"
      ]
    },
    "papert": {
      "type": "object",
      "properties": {
        "fieldMappings": {
          "type": "array",
          "items": {
            "anyOf": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "STRING_LIST",
                      "LONG",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
```

```
                    "type": "string",
                    "pattern": "dd-MM-yyyy HH:mm:ss"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        }
      },
      "required": [
        "fieldMappings"
      ]
    },
    "shortcut": {
      "type": "object",
      "properties": {
        "fieldMappings": {
          "type": "array",
          "items": {
            "anyOf": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "STRING_LIST",
                      "LONG",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
```

```
                    "type": "string",
                    "pattern": "dd-MM-yyyy HH:mm:ss"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        }
      },
      "required": [
        "fieldMappings"
      ]
    }
  }
},
"syncMode": {
  "type": "string",
  "enum": [
    "FULL_CRAWL",
    "FORCED_FULL_CRAWL",
    "CHANGE_LOG"
  ]
},
"enableIdentityCrawler": {
  "type": "boolean"
},
"secretArn": {
  "type": "string"
},
"additionalProperties": {
  "type": "object",
  "properties": {
    "isCrawlAcl": {
      "type": "boolean"
    },
    "inclusionFileNamePatterns": {
      "type": "array"
    },
    "exclusionFileNamePatterns": {
```

```
          "type": "array"
        },
        "inclusionFileTypePatterns": {
          "type": "array"
        },
        "exclusionFileTypePatterns": {
          "type": "array"
        },
        "crawlFile": {
          "type": "boolean"
        },
        "crawlPaper": {
          "type": "boolean"
        },
        "crawlPapert": {
          "type": "boolean"
        },
        "crawlShortcut": {
          "type": "boolean"
        }
      }
    },
    "type": {
      "type": "string",
      "pattern": "DROPBOX"
    },
    "tokenType": {
      "type": "string",
      "enum": [
        "PERMANENT",
        "TEMPORARY"
      ]
    },
    "version": {
      "type": "string",
      "anyOf": [
        {
          "pattern": "1.0.0"
        }
      ]
    }
  },
  "additionalProperties": false,
  "required": [
```

```
    "connectionConfiguration",
    "repositoryConfigurations",
    "additionalProperties",
    "syncMode",
    "enableIdentityCrawler",
    "secretArn",
    "type",
    "tokenType"
  ]
}
```

Drupal 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您可以在连接配置或存储库端点详细信息中提供 Drupal 主机 URL 和身份验证类型。还要将数据来源的类型指定为 DRUPAL、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Drupal JSON 架构。

下表描述了 Drupal JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
hostUrl	您的 Drupal 网站的主机 URL。例如，*https://<hostname>/<drupalsitename>* 。
repositoryConfigurations	数据来源内容的配置信息。
• content • comment • attachment	映射 Drupal 文件的属性或字段名称的对象的列表。有关更多信息，请参阅映射数据来源字段。您的 Drupal 自定义元数据中必须有 Drupal 数据来源字段名称。
additionalProperties	数据来源中内容的其他配置选项。

配置	描述
• inclusionFileName图案 • articleTitleInclusion图案 • pageTitleInclusion图案 • customContentTitleInclusionPatterns • basicBlockTitleInclusionPatterns • customBlockTitleInclusionPatterns	用于在 Drupal 数据来源中包含某些文件的正则表达式模式的列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
• exclusionFileName图案 • articleTitleExclusion图案 • pageTitleExclusion图案 • customContentTitleExclusionPatterns • basicBlockTitleExclusionPatterns • customBlockTitleExclusionPatterns	用于在 Drupal 数据来源中排除某些文件的正则表达式模式的列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配排除和包含模式，则以排除模式为优先，该文件不会包含在索引中。
contentDefinitions • contentType • fieldDefinition • isCrawlComments • isCrawlFiles • isCrawlArticle • isCrawlBasic页面 • isCrawlBasic屏蔽 • isCrawlCustomContentTypesList	指定要爬取的内容类型以及是否爬取所选内容类型的评论和附件。
type	数据来源的类型。指定 DRUPAL 作为数据来源类型。
authType	您使用的身份验证类型，可以是 BASIC-AUTH 或 OAUTH2。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。

配置	描述

secretARN	包含连接您的 Drupal 所需的键值对的 AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须包含具有以下键的 JSON 结构：
	
	如果使用基本身份验证：
	
	`{ "username": "user name", "passwords": "password" }`
	
	如果使用 OAuth 2.0 身份验证：
	
	`{ "username": "user name", "password": "password" , "clientId": "client id", "clientSecret": "client secret" }`

版本	当前支持的此模板的版本。

Drupal JSON 架构

```
{
 "$schema": "http://json-schema.org/draft-04/schema#",
 "type": "object",
 "properties": {
  "connectionConfiguration": {
   "type": "object",
   "properties": {
    "repositoryEndpointMetadata": {
     "type": "object",
     "properties": {
      "hostUrl": {
       "type": "string",
       "pattern": "https:.*"
```

```
      }
     },
     "required": [
      "hostUrl"
     ]
    }
   },
   "required": [
    "repositoryEndpointMetadata"
   ]
  },
  "repositoryConfigurations": {
   "type": "object",
   "properties": {
    "content": {
     "type": "object",
     "properties": {
      "fieldMappings": {
       "type": "array",
       "items": [
        {
         "type": "object",
         "properties": {
          "indexFieldName": {
           "type": "string"
          },
          "indexFieldType": {
           "type": "string",
           "enum": [
            "STRING",
            "DATE"
           ]
          },
          "dataSourceFieldName": {
           "type": "string"
          },
          "dateFieldFormat": {
           "type": "string",
           "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
          }
         },
         "required": [
          "indexFieldName",
          "indexFieldType",
```

```
          "dataSourceFieldName"
         ]
        }
       ]
      }
     },
     "required": [
      "fieldMappings"
     ]
    },
    "comment": {
     "type": "object",
     "properties": {
      "fieldMappings": {
       "type": "array",
       "items": [
        {
         "type": "object",
         "properties": {
          "indexFieldName": {
           "type": "string"
          },
          "indexFieldType": {
           "type": "string",
           "enum": [
            "STRING",
            "DATE"
           ]
          },
          "dataSourceFieldName": {
           "type": "string"
          },
          "dateFieldFormat": {
           "type": "string",
           "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
          }
         },
         "required": [
          "indexFieldName",
          "indexFieldType",
          "dataSourceFieldName"
         ]
        }
       ]
```

```
      }
     },
     "required": [
      "fieldMappings"
     ]
    },
    "attachment": {
     "type": "object",
     "properties": {
      "fieldMappings": {
       "type": "array",
       "items": [
        {
         "type": "object",
         "properties": {
          "indexFieldName": {
           "type": "string"
          },
          "indexFieldType": {
           "type": "string",
           "enum": [
            "STRING",
            "DATE"
           ]
          },
          "dataSourceFieldName": {
           "type": "string"
          },
          "dateFieldFormat": {
           "type": "string",
           "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
          }
         },
         "required": [
          "indexFieldName",
          "indexFieldType",
          "dataSourceFieldName"
         ]
        }
       ]
      }
     },
     "required": [
      "fieldMappings"
```

```
     ]
    }
   }
  },
  "additionalProperties": {
   "type": "object",
   "properties": {
    "isCrawlArticle": {
     "type": "boolean"
    },
    "isCrawlBasicPage": {
     "type": "boolean"
    },
    "isCrawlBasicBlock": {
     "type": "boolean"
    },
    "crawlCustomContentTypesList": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "crawlCustomBlockTypesList": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "filePath": {
     "anyOf": [
      {
       "type": "string",
       "pattern": "s3:.*"
      },
      {
       "type": "string",
       "pattern": ""
      }
     ]
    },
    "inclusionFileNamePatterns": {
     "type": "array",
     "items": {
      "type": "string"
```

```
      }
     },
     "exclusionFileNamePatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "articleTitleInclusionPatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "articleTitleExclusionPatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "pageTitleInclusionPatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "pageTitleExclusionPatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "customContentTitleInclusionPatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "customContentTitleExclusionPatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
```

```
    "basicBlockTitleInclusionPatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "basicBlockTitleExclusionPatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "customBlockTitleInclusionPatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "customBlockTitleExclusionPatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "contentDefinitions": {
     "type": "array",
     "items": {
      "properties": {
       "contentType": {
        "type": "string"
       },
       "fieldDefinition": {
        "type": "array",
        "items": [
         {
          "type": "object",
          "properties": {
           "machineName": {
            "type": "string"
           },
           "type": {
            "type": "string"
           }
          },
```

```
          "required": [
           "machineName",
           "type"
          ]
         }
        ]
       },
       "isCrawlComments": {
        "type": "boolean"
       },
       "isCrawlFiles": {
        "type": "boolean"
       }
      }
     },
     "required": [
      "contentType",
      "fieldDefinition",
      "isCrawlComments",
      "isCrawlFiles"
     ]
    }
   },
   "required": []
  },
  "type": {
   "type": "string",
   "pattern": "DRUPAL"
  },
  "authType": {
   "type": "string",
   "enum": [
    "BASIC-AUTH",
    "OAUTH2"
   ]
  },
  "syncMode": {
   "type": "string",
   "enum": [
    "FORCED_FULL_CRAWL",
    "FULL_CRAWL",
    "CHANGE_LOG"
   ]
  },
```

```
  "enableIdentityCrawler": {
   "type": "boolean"
  },
  "secretArn": {
   "type": "string",
   "minLength": 20,
   "maxLength": 2048
  }
 },
 "version": {
  "type": "string",
  "anyOf": [
   {
    "pattern": "1.0.0"
   }
  ]
 },
 "required": [
  "connectionConfiguration",
  "repositoryConfigurations",
  "syncMode",
  "additionalProperties",
  "secretArn",
  "type"
 ]
}
```

GitHub 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。在连接配置或存储库终端节点详细信息中，您需要提供 GitHub 主机 URL、组织名称以及您是使用 GitHub 云端还是 GitHub 本地部署。还要将数据来源的类型指定为 GITHUB、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 GitHub JSON 模式。

下表描述了 GitHub JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。

配置	描述
repositoryEndpointMetadata	数据来源的端点信息。
type	将类型指定为SAAS或ON_PREMISE 。
hostUrl	GitHub 主机网址。例如，如果您使用 GitHub SaaS/企业云：。https://api.github.com或者，如果您使用 GitHub本地/企业服务器：。https://on-prem-host-url/api/v3/
organizationName	登录到 GitHub 桌面并转到个人资料图片下拉列表下的 “您的组织” 时，可以找到您的组织名称。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• ghRepor • ghComm • ghIssueDocument • ghIssueComment • ghIssueAttachment • GHPR 文档 • GHPR评论 • ghpr附件	将 GitHub 内容的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。
isCrawlAcl	true如果您有 ACL 并想将其用于访问控制，则可以抓取文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问和搜索哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

配置	描述
fieldForUser我是	指定要用于 ACL 搜索的用户 ID 的类型。指定是email要使用用户电子邮件作为用户 ID，还是username要使用用户名作为用户 ID。如果您未指定选项，email则默认使用该选项。
存储库筛选器	要索引的特定存储库和分支名称的列表。
抓取存储库	true抓取存储库。
crawlRepositoryDocuments	true以抓取存储库文档。
crawlisSu	true抓取问题。
crawlIssueComment	true搜寻问题评论。
crawlIssueComment附件	true搜寻议题评论附件。
crawlPullRequest	true来抓取拉取请求。
crawlPullRequest评论	true抓取拉取请求评论。
crawlPullRequestCommentAttachment	true抓取拉取请求评论附件。
• inclusionFolderName图案 • inclusionFileType图案 • inclusionFileName图案	用于在 GitHub数据源中包含某些内容的正则表达式模式列表。与模式匹配的内容将包含在索引中。与模式不匹配的内容将从索引中排除。如果任何内容同时匹配包含模式和排除模式，则排除模式优先，并且该内容不会包含在索引中。
• exclusionFolderName图案 • exclusionFileType图案 • exclusionFileName图案	用于排除 GitHub数据源中某些内容的正则表达式模式列表。与模式匹配的内容将从索引中排除。与模式不匹配的内容将包含在索引中。如果任何内容同时匹配包含模式和排除模式，则排除模式优先，并且该内容不会包含在索引中。
type	数据来源的类型。指定 GITHUB 作为数据来源类型。

配置	描述
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • `FORCED_FULL_CRAWL` 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • `FULL_CRAWL` 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • `CHANGE_LOG` 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	AWS Secrets Manager 密钥的 Amazon 资源名称 (ARN)，其中包含连接到您的所需的键值对。 GitHub密钥必须包含具有以下键的 JSON 结构： `{ "personalToken": " token" }`
版本	当前支持的此模板的版本。

GitHub JSON 模式

以下是 GitHub JSON 架构：

```
{
    "$schema": "http://json-schema.org/draft-04/schema#",
    "type": "object",
    "properties": {
        "connectionConfiguration": {
            "type": "object",
            "properties": {
                "repositoryEndpointMetadata": {
                    "type": "object",
                    "properties": {
                        "type": {
                            "type": "string"
                        },
                        "hostUrl": {
                            "type": "string",
                            "pattern": "https://.*"
                        },
                        "organizationName": {
                            "type": "string"
                        }
                    },
                    "required": [
                        "type",
                        "hostUrl",
                        "organizationName"
                    ]
                }
            },
            "required": [
                "repositoryEndpointMetadata"
            ]
        },
        "repositoryConfigurations": {
            "type": "object",
            "properties": {
                "ghRepository": {
                    "type": "object",
                    "properties": {
                        "fieldMappings": {
                            "type": "array",
```

```
                        "items": [
                            {
                                "type": "object",
                                "properties": {
                                    "indexFieldName": {
                                        "type": "string"
                                    },
                                    "indexFieldType": {
                                        "type": "string",
                                        "enum": [
                                            "STRING",
                                            "STRING_LIST",
                                            "DATE"
                                        ]
                                    },
                                    "dataSourceFieldName": {
                                        "type": "string"
                                    },
                                    "dateFieldFormat": {
                                        "type": "string",
                                        "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                                    }
                                },
                                "required": [
                                    "indexFieldName",
                                    "indexFieldType",
                                    "dataSourceFieldName"
                                ]
                            }
                        ]
                    }
                },
                "required": [
                    "fieldMappings"
                ]
            },
            "ghCommit": {
                "type": "object",
                "properties": {
                    "fieldMappings": {
                        "type": "array",
                        "items": [
                            {
                                "type": "object",
```

```
                                "properties": {
                                    "indexFieldName": {
                                        "type": "string"
                                    },
                                    "indexFieldType": {
                                        "type": "string",
                                        "enum": [
                                            "STRING",
                                            "STRING_LIST",
                                            "DATE"
                                        ]
                                    },
                                    "dataSourceFieldName": {
                                        "type": "string"
                                    },
                                    "dateFieldFormat": {
                                        "type": "string",
                                        "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                                    }
                                },
                                "required": [
                                    "indexFieldName",
                                    "indexFieldType",
                                    "dataSourceFieldName"
                                ]
                            }
                        ]
                    }
                },
                "required": [
                    "fieldMappings"
                ]
            },
            "ghIssueDocument": {
                "type": "object",
                "properties": {
                    "fieldMappings": {
                        "type": "array",
                        "items": [
                            {
                                "type": "object",
                                "properties": {
                                    "indexFieldName": {
                                        "type": "string"
```

```
                                },
                                "indexFieldType": {
                                    "type": "string",
                                    "enum": [
                                        "STRING",
                                        "STRING_LIST",
                                        "DATE"
                                    ]
                                },
                                "dataSourceFieldName": {
                                    "type": "string"
                                },
                                "dateFieldFormat": {
                                    "type": "string",
                                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                                }
                            },
                            "required": [
                                "indexFieldName",
                                "indexFieldType",
                                "dataSourceFieldName"
                            ]
                        }
                    ]
                }
            },
            "required": [
                "fieldMappings"
            ]
        },
        "ghIssueComment": {
            "type": "object",
            "properties": {
                "fieldMappings": {
                    "type": "array",
                    "items": [
                        {
                            "type": "object",
                            "properties": {
                                "indexFieldName": {
                                    "type": "string"
                                },
                                "indexFieldType": {
                                    "type": "string",
```

```
                                    "enum": [
                                        "STRING",
                                        "STRING_LIST",
                                        "DATE"
                                    ]
                                },
                                "dataSourceFieldName": {
                                    "type": "string"
                                },
                                "dateFieldFormat": {
                                    "type": "string",
                                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                                }
                            },
                            "required": [
                                "indexFieldName",
                                "indexFieldType",
                                "dataSourceFieldName"
                            ]
                        }
                    ]
                }
            },
            "required": [
                "fieldMappings"
            ]
        },
        "ghIssueAttachment": {
            "type": "object",
            "properties": {
                "fieldMappings": {
                    "type": "array",
                    "items": [
                        {
                            "type": "object",
                            "properties": {
                                "indexFieldName": {
                                    "type": "string"
                                },
                                "indexFieldType": {
                                    "type": "string",
                                    "enum": [
                                        "STRING",
                                        "STRING_LIST",
```

```
                                    "DATE"
                                ]
                            },
                            "dataSourceFieldName": {
                                "type": "string"
                            },
                            "dateFieldFormat": {
                                "type": "string",
                                "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                            }
                        },
                        "required": [
                            "indexFieldName",
                            "indexFieldType",
                            "dataSourceFieldName"
                        ]
                    }
                ]
            }
        },
        "required": [
            "fieldMappings"
        ]
    },
    "ghPRDocument": {
        "type": "object",
        "properties": {
            "fieldMappings": {
                "type": "array",
                "items": [
                    {
                        "type": "object",
                        "properties": {
                            "indexFieldName": {
                                "type": "string"
                            },
                            "indexFieldType": {
                                "type": "string",
                                "enum": [
                                    "STRING",
                                    "STRING_LIST",
                                    "DATE"
                                ]
                            },
```

```
                                "dataSourceFieldName": {
                                    "type": "string"
                                },
                                "dateFieldFormat": {
                                    "type": "string",
                                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                                }
                            },
                            "required": [
                                "indexFieldName",
                                "indexFieldType",
                                "dataSourceFieldName"
                            ]
                        }
                    ]
                }
            },
            "required": [
                "fieldMappings"
            ]
        },
        "ghPRComment": {
            "type": "object",
            "properties": {
                "fieldMappings": {
                    "type": "array",
                    "items": [
                        {
                            "type": "object",
                            "properties": {
                                "indexFieldName": {
                                    "type": "string"
                                },
                                "indexFieldType": {
                                    "type": "string",
                                    "enum": [
                                        "STRING",
                                        "STRING_LIST",
                                        "DATE"
                                    ]
                                },
                                "dataSourceFieldName": {
                                    "type": "string"
                                },
```

```
                                "dateFieldFormat": {
                                    "type": "string",
                                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                                }
                            },
                            "required": [
                                "indexFieldName",
                                "indexFieldType",
                                "dataSourceFieldName"
                            ]
                        }
                    ]
                }
            },
            "required": [
                "fieldMappings"
            ]
        },
        "ghPRAttachment": {
            "type": "object",
            "properties": {
                "fieldMappings": {
                    "type": "array",
                    "items": [
                        {
                            "type": "object",
                            "properties": {
                                "indexFieldName": {
                                    "type": "string"
                                },
                                "indexFieldType": {
                                    "type": "string",
                                    "enum": [
                                        "STRING",
                                        "STRING_LIST",
                                        "DATE"
                                    ]
                                },
                                "dataSourceFieldName": {
                                    "type": "string"
                                },
                                "dateFieldFormat": {
                                    "type": "string",
                                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
```

```
                                    }
                                },
                                "required": [
                                    "indexFieldName",
                                    "indexFieldType",
                                    "dataSourceFieldName"
                                ]
                            }
                        ]
                    }
                },
                "required": [
                    "fieldMappings"
                ]
            }
        }
    },
    "additionalProperties": {
        "type": "object",
        "properties": {
            "isCrawlAcl": {
                "type": "boolean"
            },
            "fieldForUserId": {
                "type": "string"
            },
            "crawlRepository": {
                "type": "boolean"
            },
            "crawlRepositoryDocuments": {
                "type": "boolean"
            },
            "crawlIssue": {
                "type": "boolean"
            },
            "crawlIssueComment": {
                "type": "boolean"
            },
            "crawlIssueCommentAttachment": {
                "type": "boolean"
            },
            "crawlPullRequest": {
                "type": "boolean"
            },
```

```
                "crawlPullRequestComment": {
                    "type": "boolean"
                },
                "crawlPullRequestCommentAttachment": {
                    "type": "boolean"
                },
                "repositoryFilter": {
                    "type": "array",
                    "items": [
                        {
                            "type": "object",
                            "properties": {
                                "repositoryName": {
                                    "type": "string"
                                },
                                "branchNameList": {
                                    "type": "array",
                                    "items": {
                                        "type": "string"
                                    }
                                }
                            }
                        }
                    ]
                },
                "inclusionFolderNamePatterns": {
                    "type": "array",
                    "items": {
                        "type": "string"
                    }
                },
                "inclusionFileTypePatterns": {
                    "type": "array",
                    "items": {
                        "type": "string"
                    }
                },
                "inclusionFileNamePatterns": {
                    "type": "array",
                    "items": {
                        "type": "string"
                    }
                },
                "exclusionFolderNamePatterns": {
```

```
                    "type": "array",
                    "items": {
                        "type": "string"
                    }
                },
                "exclusionFileTypePatterns": {
                    "type": "array",
                    "items": {
                        "type": "string"
                    }
                },
                "exclusionFileNamePatterns": {
                    "type": "array",
                    "items": {
                        "type": "string"
                    }
                }
            },
            "required": []
        },
        "type": {
            "type": "string",
            "pattern": "GITHUB"
        },
        "syncMode": {
            "type": "string",
            "enum": [
                "FULL_CRAWL",
                "FORCED_FULL_CRAWL",
                "CHANGE_LOG"
            ]
        },
        "enableIdentityCrawler": {
            "type": "boolean"
        },
        "secretArn": {
            "type": "string",
            "minLength": 20,
            "maxLength": 2048
        }
    },
    "version": {
        "type": "string",
        "anyOf": [
```

```
            {
                "pattern": "1.0.0"
            }
        ]
    },
    "required": [
        "connectionConfiguration",
        "repositoryConfigurations",
        "syncMode",
        "additionalProperties",
        "enableIdentityCrawler"
    ]
}
```

Gmail 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 GMAIL、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Gmail JSON 架构。

下表描述了 Gmail JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。此数据来源未在 `repositoryEndpointMetadata` 中指定端点。相反，连接信息包含在您提供的 AWS Secrets Manager 密钥中`secretArn` 。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
• 消息 • attachments	将您的 Gmail 邮件和附件的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。

配置	描述
additionalProperties	数据来源中内容的其他配置选项。
• inclusionLabelName图案 • exclusionLabelName图案 • inclusionAttachmentType图案 • exclusionAttachmentType图案 • inclusionAttachmentName图案 • exclusionAttachmentName图案 • inclusionSubjectFilter • exclusionSubjectFilter • isSubjectAnd • inclusionFromFilter • exclusionFromFilter • inclusionToFilter • exclusionToFilter • inclusionCcFilter • exclusionCcFilter • inclusionBccFilter • exclusionBccFilter	用于在 Gmail 数据来源中包含或排除特定主题名称的正则表达式模式的列表。与模式匹配的文件将包含在索引中。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
beforeDateFilter	指定包含特定日期之前的邮件和附件。
afterDateFilter	指定包含特定日期之后的邮件和附件。
isCrawlAttachment	一个布尔值，用于选择是否要爬取附件。自动爬取邮件。
type	数据来源的类型。指定 `GMAIL` 作为数据来源类型。
shouldCrawlDraft消息	一个布尔值，用于选择是否要爬取邮件草稿。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 **Important** 由于没有用于更新永久删除的 Gmail 邮件的 API，因此任何新的、修改过的或删除的内容都会同步： • 不会从您的 Amazon Kendra 索引中移除从 Gmail 中永久删除的邮件 • 无法同步 Gmail 电子邮件标签中的更改 要将您的 Gmail 数据源标签更改和永久删除的电子邮件同步到您的 Amazon Kendra 索引，您必须定期进行全面抓取。

配置	描述
secretARN	Secrets Manager 密钥的 Amazon 资源名称（ARN），该密钥包含连接到 Gmail 所需的键-值对。密钥必须包含具有以下键的 JSON 结构： `{ "adminAccountEmailId": " service account email", "clientEmailId": " user account email", "privateKey": " private key" }`
版本	当前支持的此模板的版本。

Gmail JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
      }
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "message": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
```

```
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": ["STRING", "STRING_LIST", "DATE"]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        }
      },
      "attachments": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": ["STRING"]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  }
                },
                "required": [
```

```
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                ]
              }
            ]
          }
        }
      }
    },
    "required": []
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "inclusionLabelNamePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "exclusionLabelNamePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "inclusionAttachmentTypePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "exclusionAttachmentTypePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "inclusionAttachmentNamePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
```

```
        },
        "exclusionAttachmentNamePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionSubjectFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionSubjectFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "isSubjectAnd": {
          "type": "boolean"
        },
        "inclusionFromFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionFromFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionToFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionToFilter": {
          "type": "array",
          "items": {
            "type": "string"
```

```
        }
      },
      "inclusionCcFilter": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "exclusionCcFilter": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "inclusionBccFilter": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "exclusionBccFilter": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "beforeDateFilter": {
        "anyOf": [
          {
            "type": "string",
            "pattern": "^[0-9]{4}-[0-9]{2}-[0-9]{2}T[0-9]{2}:[0-9]{2}:[0-9]{2}Z$"
          },
          {
            "type": "string",
            "pattern": ""
          }
        ]
      },
      "afterDateFilter": {
        "anyOf": [
          {
            "type": "string",
            "pattern": "^[0-9]{4}-[0-9]{2}-[0-9]{2}T[0-9]{2}:[0-9]{2}:[0-9]{2}Z$"
          },
```

```
            {
              "type": "string",
              "pattern": ""
            }
          ]
        },
        "isCrawlAttachment": {
          "type": "boolean"
        },
        "shouldCrawlDraftMessages": {
          "type": "boolean"
        }
      },
      "required": [
        "isCrawlAttachment",
        "shouldCrawlDraftMessages"
      ]
    },
    "type" : {
      "type" : "string",
      "pattern": "GMAIL"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL"
      ]
    },
    "secretArn": {
      "type": "string"
    },
    "version": {
      "type": "string",
      "anyOf": [
        {
          "pattern": "1.0.0"
        }
      ]
    }
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
```

```
    "additionalProperties",
    "syncMode",
    "secretArn",
    "type"
  ]
}
```

Google Drive 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 GOOGLEDRIVE2、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Google Drive JSON 架构。

下表描述了 Google 云端硬盘 JSON 架构的参数。

配置	描述
connectionConfiguration	数据来源的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。此数据来源未指定端点。您可以选择身份验证类型：serviceAccount 和 OAuth2。连接信息包含在您提供的 AWS Secrets Manager 密钥中secretArn 。
authType	根据您的使用案例，选择 serviceAccount 或 OAuth2。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• file • comment	将 Google Drive 的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象的列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项
• maxFileSizeInMegaBytes	指定 Amazon Kendra 应抓取的文件大小限制（以 MB 为单位）。

配置	描述
• iscrawlComment	`true`抓取您的 Google 云端硬盘数据源中的评论。
• isCrawlMyDriveAndSharedWithMe	`true`在您的 Google 云端硬盘数据源中抓取 MyDrive 并与我共享云端硬盘。
• isCrawlShared驱动器	`true`在您的 Google 云端硬盘数据源中抓取共享云端硬盘。
isCrawlAcl	`true`如果您有 ACL 并想将其用于访问控制，则可以抓取文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问和搜索哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
• excludeUserAccounts • excludeSharedDrives • excludeMimeTypes • exclusionFileType图案 • exclusionFileName图案 • exclusionFilePath过滤器	用于在 Google Drive 数据来源中排除某些文件的正则表达式模式的列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配排除和包含模式，则以排除模式为优先，该文件不会包含在索引中。
• includeUserAccounts • includeSharedDrives • includeMimeTypes • inclusionFileType图案 • inclusionFileName图案 • inclusionFilePath过滤器	用于在 Google Drive 数据来源中包含某些文件的正则表达式模式的列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
type	数据来源的类型。指定 `GOOGLEDRIVEV2` 作为数据来源类型。

配置	描述
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

配置	描述

secretARN	包含连接您的 Google 云端硬盘所需的键值对的 AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须包含具有以下键的 JSON 结构：
	
	，如果使用 Google 服务账户身份验证：
	
	`{ "clientEmail": " user account email", "adminAccountEmail": " service account email", "privateKey": " private key" }`
	
	如果使用 OAuth 2.0 身份验证：
	
	`{ "clientID": " OAuth client ID", "clientSecret": " client secret", "refreshToken": " refresh token" }`

版本	当前支持的此模板的版本。

Google Drive JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "authType": {
```

```
            "type": "string",
            "enum": [
              "serviceAccount",
              "OAuth2"
            ]
          }
        },
        "required": [
          "authType"
        ]
      }
    },
    "required": [
      "repositoryEndpointMetadata"
    ]
  },
  "repositoryConfigurations": {
    "type": "object",
    "properties": {
      "file": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "DATE",
                      "STRING_LIST",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
```

```
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "comment": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "DATE",
                      "STRING_LIST"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
```

```
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    }
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "maxFileSizeInMegaBytes": {
        "type": "string"
      },
      "isCrawlComment": {
        "type": "boolean"
      },
      "isCrawlMyDriveAndSharedWithMe": {
        "type": "boolean"
      },
      "isCrawlSharedDrives": {
        "type": "boolean"
      },
      "isCrawlAcl": {
        "type": "boolean"
      },
      "excludeUserAccounts": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "excludeSharedDrives": {
        "type": "array",
        "items": {
          "type": "string"
```

```
        }
      },
      "excludeMimeTypes": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "includeUserAccounts": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "includeSharedDrives": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "includeMimeTypes": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "includeTargetAudienceGroup": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "inclusionFileTypePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "inclusionFileNamePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
```

```
        "exclusionFileTypePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionFileNamePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionFilePathFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionFilePathFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        }
      }
    },
    "type": {
      "type": "string",
      "pattern": "GOOGLEDRIVEV2"
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string",
      "minLength": 20,
```

```
      "maxLength": 2048
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "syncMode",
    "additionalProperties",
    "secretArn",
    "type"
  ]
}
```

IBM DB2 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、db2 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 IBM DB2 JSON 架构。

下表描述了 IBM DB2 JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。 • dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。

配置	描述
	• dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容

配置	描述
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

配置	描述
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

IBM DB2 JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
```

```
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string"
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
```

```
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "primaryKey": {
        "type": "string"
      },
      "titleColumn": {
        "type": "string"
      },
      "bodyColumn": {
        "type": "string"
      },
      "sqlQuery": {
        "type": "string",
        "not": {
          "pattern": ";+"
        }
      },
      "timestampColumn": {
        "type": "string"
      },
      "timestampFormat": {
        "type": "string"
      },
      "timezone": {
        "type": "string"
      },
      "changeDetectingColumns": {
        "type": "array",
        "items": {
          "type": "string"
        }
```

```
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
```

```
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Microsoft Exchange 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您可以将租户 ID 作为连接配置或存储库端点详细信息的一部分提供。还要将数据来源的类型指定为MSEXCHANGE、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Microsoft Exchange JSON 架构。

下表描述了微软 Exchange JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
tenantId	Microsoft 365 租户 ID。您可以在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中找到您的租户 ID。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• email • attachment • 日历 • 联系人 • notes	将你的 Microsoft Exchange 数据源的属性或字段名称映射到 Amazon Kendra 索引字段的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项
inclusionPatterns	用于在 Microsoft Exchange 数据来源中包含某些文件的正则表达式模式的列表。与模式匹配的

配置	描述
	文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
exclusionPatterns	用于在 Microsoft Exchange 数据来源中排除某些文件的正则表达式模式的列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配排除和包含模式，则以排除模式为优先，该文件不会包含在索引中。
• inclusionUsersList • inclusionUsersFile姓名 • inclusionDomainUsers	用于在 Microsoft Exchange 数据来源中包含某些用户的正则表达式模式的列表。与模式匹配的用户将包含在索引中。与模式不匹配的用户将从索引中排除。如果用户同时匹配包含和排除模式，则以排除模式为优先，该用户不会包含在索引中。
• exclusionUsersList • exclusionUsersFile姓名 • exclusionDomainUsers	用于在 Microsoft Exchange 数据来源中排除某些用户和用户文件的正则表达式模式的列表。与模式匹配的用户将从索引中排除。与模式不匹配的用户将包含在索引中。如果用户同时匹配排除和包含模式，则以排除模式为优先，该用户不会包含在索引中。
s3bucketName	S3 存储桶的名称（如果要使用）。
• crawlCalendar • crawlNotes • crawlContacts • crawlFolderAcl	`true`抓取这些类型的内容和访问控制信息你的 Microsoft Exchange 数据源。
startCalendarDate时间	您可以为日历内容配置特定的开始日期时间。
endCalendarDate时间	您可以为日历内容配置特定的结束日期时间。

配置	描述
subject	您可以为邮件内容配置特定的主题行。
emailFrom	您可以为“发件人”或发件人邮件内容配置特定的电子邮件。
emailTo	您可以为“收件人”或收件人邮件内容配置特定的电子邮件。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
type	数据来源的类型。指定 MSEXCHANGE 作为数据来源类型。
secretARN	包含连接微软 Exchange 所需的键值对的 AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)。这包括您的客户端 ID 和在 Azure 门户中创建 OAuth 应用程序时生成的客户端密钥。
版本	当前支持的此模板的版本。

Microsoft Exchange JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "tenantId": {
              "type": "string",
              "pattern": "^[0-9a-f]{8}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]
{12}$",
              "minLength": 36,
              "maxLength": 36
            }
          },
          "required": ["tenantId"]
        }
      }
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "email": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": ["STRING", "STRING_LIST", "DATE"]
                    },
                    "dataSourceFieldName": {
```

```
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "attachment": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": ["STRING", "DATE","LONG"]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
```

```
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "calendar": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": ["STRING", "STRING_LIST", "DATE"]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
```

```
            }
          },
          "required": [
            "fieldMappings"
          ]
        },
        "contacts": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": ["STRING", "STRING_LIST", "DATE"]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        },
        "notes": {
```

```
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": ["STRING", "DATE"]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        }
      },
      "required": ["email"
      ]
    },
    "additionalProperties": {
      "type": "object",
      "properties": {
        "inclusionPatterns": {
```

```
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionUsersList": {
          "type": "array",
          "items": {
            "type": "string",
            "format": "email"
          }
        },
        "exclusionUsersList": {
          "type": "array",
          "items": {
            "type": "string",
            "format": "email"
          }
        },
        "s3bucketName": {
          "type": "string"
        },
        "inclusionUsersFileName": {
          "type": "string"
        },
        "exclusionUsersFileName": {
          "type": "string"
        },
        "inclusionDomainUsers": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionDomainUsers": {
          "type": "array",
          "items": {
            "type": "string"
```

```
        }
      },
      "crawlCalendar": {
        "type": "boolean"
      },
      "crawlNotes": {
        "type": "boolean"
      },
      "crawlContacts": {
        "type": "boolean"
      },
      "crawlFolderAcl": {
        "type": "boolean"
      },
      "startCalendarDateTime": {
        "anyOf": [
          {
            "type": "string",
            "pattern": "^[0-9]{4}-[0-9]{2}-[0-9]{2}T[0-9]{2}:[0-9]{2}:[0-9]{2}Z$"
          },
          {
            "type": "string",
            "pattern": ""
          }
        ]
      },
      "endCalendarDateTime": {
        "anyOf": [
          {
          "type": "string",
          "pattern": "^[0-9]{4}-[0-9]{2}-[0-9]{2}T[0-9]{2}:[0-9]{2}:[0-9]{2}Z$"
          },
          {
            "type": "string",
            "pattern": ""
          }
        ]
      },
      "subject": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
```

```
        "emailFrom": {
          "type": "array",
          "items": {
            "type": "string",
            "format": "email"
          }
        },
        "emailTo": {
          "type": "array",
          "items": {
            "type": "string",
            "format": "email"
          }
        }
      },
      "required": [
      ]
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "type" : {
      "type" : "string",
      "pattern": "MSEXCHANGE"
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
```

```
    "repositoryConfigurations",
    "syncMode",
    "additionalProperties",
    "secretArn",
    "type"
  ]
}
```

微软 OneDrive 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您可以将 租户 ID 作为连接配置或存储库端点详细信息的一部分提供。还要将数据来源的类型指定为ONEDRIVEV2、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 微软 OneDrive JSON 架构。

下表描述了微软 OneDrive JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
tenantId	Microsoft 365 租户 ID。您可以在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中找到您的租户 ID。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
file	将你的 Microsoft OneDrive 文件的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项
• userNameFilter • userFilterPath	您可以选择为特定文件、 OneNote 章节、 OneNote 页面编制索引，并按用户名进行筛选。

配置	描述
• inclusionFileType图案 • exclusionFileType图案 • inclusionFileName图案 • exclusionFileName图案 • inclusionFilePath图案 • exclusionFilePath图案 • inclusionOneNoteSectionNamePatterns • exclusionOneNoteSectionNamePatterns • inclusionOneNotePageNamePatterns • exclusionOneNotepageNamePatterns	
isUserNameonS3	`true` 提供存储在 Amazon S3的文件中的用户名列表。
type	数据来源的类型。指定 `ONEDRIVEV2` 作为数据来源类型。
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。
type	数据来源的类型。指定 `ONEDRIVEV2` 作为数据来源类型。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretARN	包含连接微软所需的键值对的 AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)。OneDrive密钥必须包含具有以下键的 JSON 结构： `{ "clientId": " client ID", "clientSecret": " client secret" }`
版本	当前支持的此模板的版本。

微软 OneDrive JSON 架构

```
{
 "$schema": "http://json-schema.org/draft-04/schema#",
 "type": "object",
 "properties": {
  "connectionConfiguration": {
   "type": "object",
```

```
  "properties": {
   "repositoryEndpointMetadata": {
    "type": "object",
    "properties": {
     "tenantId": {
      "type": "string",
      "pattern": "^[0-9a-f]{8}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{12}$",
      "minLength": 36,
      "maxLength": 36
     }
    },
    "required": [
     "tenantId"
    ]
   }
  },
  "required": [
   "repositoryEndpointMetadata"
  ]
 },
 "repositoryConfigurations": {
  "type": "object",
  "properties": {
   "file": {
    "type": "object",
    "properties": {
     "fieldMappings": {
      "type": "array",
      "items": [
       {
        "type": "object",
        "properties": {
         "indexFieldName": {
          "type": "string"
         },
         "indexFieldType": {
          "type": "string",
          "enum": [
           "STRING",
           "STRING_LIST",
           "DATE",
           "LONG"
          ]
         },
```

```
          "dataSourceFieldName": {
           "type": "string"
          },
          "dateFieldFormat": {
           "type": "string",
           "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
          }
         },
         "required": [
          "indexFieldName",
          "indexFieldType",
          "dataSourceFieldName"
         ]
        }
       ]
      }
     },
     "required": [
      "fieldMappings"
     ]
    }
   }
  },
  "additionalProperties": {
   "type": "object",
   "properties": {
    "userNameFilter": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "userFilterPath": {
     "type": "string"
    },
    "isUserNameOnS3": {
     "type": "boolean"
    },
    "inclusionFileTypePatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
```

```
    "exclusionFileTypePatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "inclusionFileNamePatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "exclusionFileNamePatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "inclusionFilePathPatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "exclusionFilePathPatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "inclusionOneNoteSectionNamePatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "exclusionOneNoteSectionNamePatterns": {
     "type": "array",
     "items": {
      "type": "string"
     }
    },
    "inclusionOneNotePageNamePatterns": {
     "type": "array",
```

```
      "items": {
       "type": "string"
      }
     },
     "exclusionOneNotePageNamePatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     }
    },
    "required": []
   },

   "enableIdentityCrawler": {
    "type": "boolean"
   },
   "type": {
    "type": "string",
    "pattern": "ONEDRIVEV2"
   },
   "syncMode": {
    "type": "string",
    "enum": [
     "FULL_CRAWL",
     "FORCED_FULL_CRAWL",
     "CHANGE_LOG"
    ]
   },
   "secretArn": {
    "type": "string",
    "minLength": 20,
    "maxLength": 2048
   }
  },
  "version": {
   "type": "string",
   "anyOf": [
    {
     "pattern": "1.0.0"
    }
   ]
  },
  "required": [
```

```
  "connectionConfiguration",
  "repositoryConfigurations",
  "syncMode",
  "additionalProperties",
  "secretArn",
  "type"
 ]
}
```

微软 SharePoint 模板架构

您将包含包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。在连接配置或存储库端点详细信息中，您可以提供 SharePoint 站点 URL/URL、域以及租户 ID（如果需要）。还要将数据来源的类型指定为 SHAREPOINTV2、身份验证凭证的密钥以及其他必要的配置。然后在呼叫时指定TEMPLATE为 “类型” CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 SharePoint JSON 模式。

下表描述了微软 SharePoint JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息
repositoryEndpointMetadata	数据来源的端点信息
tenantId	您 SharePoint 账户的租户 ID。
域	您 SharePoint 账户的域名。
siteUrls	您 SharePoint 账户的主机网址。
repositoryAdditionalProperties	用于连接存储库/数据来源端点的其他属性。
s3bucketName	存储 Azure AD 自签名 X.509 证书的 Amazon S3 存储桶的名称。
s3certificateName	存储在存储桶中的 Azure AD 自签名 X.509 证书的名称。 Amazon S3
authType	您使用的身份验证类型，是OAuth2、OAuth2Cer

配置	描述
	tificate 、OAuth2App 、Basic、OAuth2_RefreshToke n NTLM、或Kerberos。
版本	您使用的 SharePoint 版本，Server无论是Online。
onPremVersion	您使用的 SharePoint 服务器版本，是201320162019、或SubscriptionEdition 。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• 事件 • 页 • file • link • attachment • comment	将 SharePoint内容的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。

配置	描述
• eventTitleFilterRegEx • pageTitleFilterRegEx • linkTitleFilterRegEx • inclusionFilePath • exclusionFilePath • inclusionFileType图案 • exclusionFileType图案 • inclusionFileName图案 • exclusionFileName图案 • inclusionOneNoteSectionNamePatterns • exclusionOneNoteSectionNamePatterns • inclusionOneNotePageNamePatterns • exclusionOneNotePageNamePatterns	用于在 SharePoint 数据源中包含/排除某些内容的正则表达式模式列表。与包含模式相匹配的内容项目包含在索引中。与包含模式不匹配的内容项将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
• crawlFile • crawlPages • crawlEvents • crawlComments • crawlLinks • crawlAttachments	`true`来抓取这些类型的内容。
crawlAcl	`true`如果您有 ACL 并想将其用于访问控制，则可以抓取文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问和搜索哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
fieldForUser我是	指定是`email`要使用用户电子邮件作为用户 ID，还是`userPrincipalName` 要使用用户名作为用户 ID。如果您未指定选项，`email`则默认使用该选项。

配置	描述
aclConfiguration	指定ACLWithLDAPEmailFmt ACLWithManualEmailFmt 、或ACLWithUsernameFmtM 。
emailDomain	电子邮件的域名。例如，*“amazon.com”* 。
• isCrawlLocalGroupMapping • isCrawlAdGroupMapping	true以搜寻群组映射信息。
proxyHost	您使用的网络代理的主机名，不带 http://或 https://协议。
proxyPort	主机 URL 传输协议使用的端口号。必须是介于 0 和 65535 之间的数值。
type	指定 SHAREPOINTV2 作为数据来源类型。
enableIdentityCrawler	true使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretARN	AWS Secrets Manager 密钥的 Amazon 资源名称 (ARN)，其中包含连接到您的所需的键值对。SharePoint有关这些键值对的信息，请参阅 On SharePoint line 和 SharePoint Server 的连接说明。
版本	当前支持的此模板的版本。

SharePoint JSON 模式

```
{
 "$schema": "http://json-schema.org/draft-04/schema#",
 "type": "object",
 "properties": {
  "connectionConfiguration": {
   "type": "object",
   "properties": {
    "repositoryEndpointMetadata": {
     "type": "object",
     "properties": {
```

```
      "tenantId": {
       "type": "string",
       "pattern": "^[0-9a-f]{8}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{12}$",
       "minLength": 36,
       "maxLength": 36
      },
      "domain": {
       "type": "string"
      },
      "siteUrls": {
       "type": "array",
       "items": {
        "type": "string",
        "pattern": "https://.*"
       }
      },
      "repositoryAdditionalProperties": {
       "type": "object",
       "properties": {
        "s3bucketName": {
         "type": "string"
        },
        "s3certificateName": {
         "type": "string"
        },
        "authType": {
         "type": "string",
         "enum": [
          "OAuth2",
          "OAuth2Certificate",
          "OAuth2App",
          "Basic",
          "OAuth2_RefreshToken",
          "NTLM",
          "Kerberos"
         ]
        },
        "version": {
         "type": "string",
         "enum": [
          "Server",
          "Online"
         ]
        },
```

```
        "onPremVersion": {
         "type": "string",
         "enum": [
          "",
          "2013",
          "2016",
          "2019",
          "SubscriptionEdition"
         ]
        }
       },
       "required": [
        "authType",
        "version"
       ]
      }
     },
     "required": [
      "siteUrls",
      "domain",
      "repositoryAdditionalProperties"
     ]
    }
   },
   "required": [
    "repositoryEndpointMetadata"
   ]
  },
  "repositoryConfigurations": {
   "type": "object",
   "properties": {
    "event": {
     "type": "object",
     "properties": {
      "fieldMappings": {
       "type": "array",
       "items": [
        {
         "type": "object",
         "properties": {
          "indexFieldName": {
           "type": "string"
          },
          "indexFieldType": {
```

```
          "type": "string",
          "enum": [
           "STRING",
           "STRING_LIST",
           "DATE"
          ]
         },
         "dataSourceFieldName": {
          "type": "string"
         },
         "dateFieldFormat": {
          "type": "string",
          "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
         }
        },
        "required": [
         "indexFieldName",
         "indexFieldType",
         "dataSourceFieldName"
        ]
       }
      ]
     }
    },
    "required": [
     "fieldMappings"
    ]
   },
   "page": {
    "type": "object",
    "properties": {
     "fieldMappings": {
      "type": "array",
      "items": [
       {
        "type": "object",
        "properties": {
         "indexFieldName": {
          "type": "string"
         },
         "indexFieldType": {
          "type": "string",
          "enum": [
           "STRING",
```

```
            "DATE",
            "LONG"
           ]
          },
          "dataSourceFieldName": {
           "type": "string"
          },
          "dateFieldFormat": {
           "type": "string",
           "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
          }
         },
         "required": [
          "indexFieldName",
          "indexFieldType",
          "dataSourceFieldName"
         ]
        }
       ]
      }
     },
     "required": [
      "fieldMappings"
     ]
    },
    "file": {
     "type": "object",
     "properties": {
      "fieldMappings": {
       "type": "array",
       "items": [
        {
         "type": "object",
         "properties": {
          "indexFieldName": {
           "type": "string"
          },
          "indexFieldType": {
           "type": "string",
           "enum": [
            "STRING",
            "DATE",
            "LONG"
           ]
```

```
          },
          "dataSourceFieldName": {
           "type": "string"
          },
          "dateFieldFormat": {
           "type": "string",
           "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
          }
         },
         "required": [
          "indexFieldName",
          "indexFieldType",
          "dataSourceFieldName"
         ]
        }
       ]
      }
     },
     "required": [
      "fieldMappings"
     ]
    },
    "link": {
     "type": "object",
     "properties": {
      "fieldMappings": {
       "type": "array",
       "items": [
        {
         "type": "object",
         "properties": {
          "indexFieldName": {
           "type": "string"
          },
          "indexFieldType": {
           "type": "string",
           "enum": [
            "STRING",
            "STRING_LIST",
            "DATE"
           ]
          },
          "dataSourceFieldName": {
           "type": "string"
```

```
          },
          "dateFieldFormat": {
           "type": "string",
           "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
          }
         },
         "required": [
          "indexFieldName",
          "indexFieldType",
          "dataSourceFieldName"
         ]
        }
       ]
      }
     },
     "required": [
      "fieldMappings"
     ]
    },
    "attachment": {
     "type": "object",
     "properties": {
      "fieldMappings": {
       "type": "array",
       "items": [
        {
         "type": "object",
         "properties": {
          "indexFieldName": {
           "type": "string"
          },
          "indexFieldType": {
           "type": "string",
           "enum": [
            "STRING",
            "STRING_LIST",
            "DATE"
           ]
          },
          "dataSourceFieldName": {
           "type": "string"
          },
          "dateFieldFormat": {
           "type": "string",
```

```
          "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
         }
        },
        "required": [
         "indexFieldName",
         "indexFieldType",
         "dataSourceFieldName"
        ]
       }
      ]
     }
    },
    "required": [
     "fieldMappings"
    ]
   },
   "comment": {
    "type": "object",
    "properties": {
     "fieldMappings": {
      "type": "array",
      "items": [
       {
        "type": "object",
        "properties": {
         "indexFieldName": {
          "type": "string"
         },
         "indexFieldType": {
          "type": "string",
          "enum": [
           "STRING",
           "STRING_LIST",
           "DATE"
          ]
         },
         "dataSourceFieldName": {
          "type": "string"
         },
         "dateFieldFormat": {
          "type": "string",
          "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
         }
        },
```

```
          "required": [
           "indexFieldName",
           "indexFieldType",
           "dataSourceFieldName"
          ]
         }
        ]
       }
      },
      "required": [
       "fieldMappings"
      ]
     }
    }
   },
   "additionalProperties": {
    "type": "object",
    "properties": {
     "eventTitleFilterRegEx": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "pageTitleFilterRegEx": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "linkTitleFilterRegEx": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "inclusionFilePath": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "exclusionFilePath": {
      "type": "array",
```

```
      "items": {
       "type": "string"
      }
     },
     "inclusionFileTypePatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "exclusionFileTypePatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "inclusionFileNamePatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "exclusionFileNamePatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "inclusionOneNoteSectionNamePatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "exclusionOneNoteSectionNamePatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "inclusionOneNotePageNamePatterns": {
      "type": "array",
      "items": {
       "type": "string"
```

```
      }
     },
     "exclusionOneNotePageNamePatterns": {
      "type": "array",
      "items": {
       "type": "string"
      }
     },
     "crawlFiles": {
      "type": "boolean"
     },
     "crawlPages": {
      "type": "boolean"
     },
     "crawlEvents": {
      "type": "boolean"
     },
     "crawlComments": {
      "type": "boolean"
     },
     "crawlLinks": {
      "type": "boolean"
     },
     "crawlAttachments": {
      "type": "boolean"
     },
     "crawlListData": {
      "type": "boolean"
     },
     "crawlAcl": {
      "type": "boolean"
     },
     "fieldForUserId": {
      "type": "string"
     },
     "aclConfiguration": {
      "type": "string",
      "enum": [
       "ACLWithLDAPEmailFmt",
       "ACLWithManualEmailFmt",
       "ACLWithUsernameFmt"
      ]
     },
     "emailDomain": {
```

```
      "type": "string"
     },
     "isCrawlLocalGroupMapping": {
      "type": "boolean"
     },
     "isCrawlAdGroupMapping": {
      "type": "boolean"
     },
     "proxyHost": {
      "type": "string"
     },
     "proxyPort": {
      "type": "string"
     }
    },
    "required": [
    ]
   },
   "type": {
    "type": "string",
    "pattern": "SHAREPOINTV2"
   },
   "enableIdentityCrawler": {
    "type": "boolean"
   },
   "syncMode": {
    "type": "string",
    "enum": [
     "FULL_CRAWL",
     "FORCED_FULL_CRAWL",
     "CHANGE_LOG"
    ]
   },
   "secretArn": {
    "type": "string",
    "minLength": 20,
    "maxLength": 2048
   }
  },
  "version": {
   "type": "string",
   "anyOf": [
    {
     "pattern": "1.0.0"
```

```
    }
   ]
  },
  "required": [
   "connectionConfiguration",
   "repositoryConfigurations",
   "enableIdentityCrawler",
   "syncMode",
   "additionalProperties",
   "secretArn",
   "type"
  ]
}
```

Microsoft SQL Server 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、sqlserver 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Microsoft SQL Server JSON 架构。

下表描述了 Micorosft SQL Server JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。 • dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。

配置	描述
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。

配置	描述
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

Microsoft SQL Server JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
```

```
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string"
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        }
      },
      "required": [
      ]
    },
    "additionalProperties": {
      "type": "object",
      "properties": {
        "primaryKey": {
          "type": "string"
```

```
        },
        "titleColumn": {
          "type": "string"
        },
        "bodyColumn": {
          "type": "string"
        },
        "sqlQuery": {
          "type": "string",
          "not": {
            "pattern": ";+"
          }
        },
        "timestampColumn": {
          "type": "string"
        },
        "timestampFormat": {
          "type": "string"
        },
        "timezone": {
          "type": "string"
        },
        "changeDetectingColumns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
```

```
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Microsoft Teams 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您可以将租户 ID 作为连接配置或存储库端点详细信息的一部分提供。还要将数据来源的类型指定为 MSTEAMS、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Microsoft Teams JSON 架构。

下表描述了微软 Teams JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
tenantId	Microsoft 365 租户 ID。您可以在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中找到您的租户 ID。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• chatMessage • chatAttachment • channelPost • channelWiki • channelAttachment • meetingChat • meetingFile • meetingNote • calendarMeeting	将你的 Microsoft Teams 内容的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。
paymentModel	指定用于您的 Microsoft Teams 数据来源的支付模式类型。A 型支付模式仅限于需要安全合规的许可和支付模式。B 型支付模式适用于不需要安全合规的许可和支付模式。
• inclusionTeamName过滤器 • inclusionChannelName过滤器 • inclusionFileName图案 • inclusionFileType图案 • inclusionUserEmail过滤器 • inclusionOneNoteSectionNamePatterns	用于在 Microsoft Teams 数据来源中包含某些内容的正则表达式模式的列表。与模式匹配的内容将包含在索引中。与模式不匹配的内容将从索引中排除。如果内容同时匹配包含和排除模式，则以排除模式为优先，该内容不会包含在索引中。

配置	描述
• inclusionOneNotePageNamePatterns	
• exclusionTeamName过滤器 • exclusionChannelName过滤器 • exclusionFileName图案 • exclusionFileType图案 • exclusionUserEmail过滤器 • exclusionOneNoteSectionNamePatterns • exclusionOneNotePageNamePatterns	用于在 Microsoft Teams 数据来源中排除某些内容的正则表达式模式的列表。与模式匹配的内容将从索引中排除。与模式不匹配的内容将包含在索引中。如果内容同时匹配包含和排除模式，则以排除模式为优先，该内容不会包含在索引中。
• isCrawlChat消息 • isCrawlChat附件 • isCrawlChannel帖子 • isCrawlChannel附件 • isCrawlChannel维基 • isCrawlCalendar会议 • isCrawlMeeting聊天 • isCrawlMeeting文件 • isCrawlMeeting注意	`true`在你的 Microsoft Teams 数据源中抓取这些类型的内容。
startCalendarDate时间	您可以为日历内容配置特定的开始日期时间。
endCalendarDate时间	您可以为日历内容配置特定的结束日期时间。
type	数据来源的类型。指定 `MSTEAMS` 作为数据来源类型。
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)，其中包含连接你的 Microsoft Teams 所需的键值对。这包括您的客户端 ID 和在 Azure 门户中创建 OAuth 应用程序时生成的客户端密钥。
版本	当前支持的此模板的版本。

Microsoft Teams JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
```

```
            "tenantId": {
              "type": "string",
              "pattern": "^[0-9a-f]{8}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]{4}-[0-9a-f]
{12}$",
              "minLength": 36,
              "maxLength": 36
            }
          },
          "required": [
            "tenantId"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "chatMessage": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING",
                        "STRING_LIST",
                        "DATE"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
```

```
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "chatAttachment": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
```

```
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "channelPost": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "STRING_LIST",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
```

```
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        },
        "channelWiki": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING",
                        "DATE",
                        "LONG"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
```

```
            }
          ]
        }
      },
      "required": [
        "fieldMappings"
      ]
    },
    "channelAttachment": {
      "type": "object",
      "properties": {
        "fieldMappings": {
          "type": "array",
          "items": [
            {
              "type": "object",
              "properties": {
                "indexFieldName": {
                  "type": "string"
                },
                "indexFieldType": {
                  "type": "string",
                  "enum": [
                    "STRING",
                    "DATE",
                    "LONG"
                  ]
                },
                "dataSourceFieldName": {
                  "type": "string"
                },
                "dateFieldFormat": {
                  "type": "string",
                  "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                }
              },
              "required": [
                "indexFieldName",
                "indexFieldType",
                "dataSourceFieldName"
              ]
            }
          ]
        }
```

```
      },
      "required": [
        "fieldMappings"
      ]
    },
    "meetingChat": {
      "type": "object",
      "properties": {
        "fieldMappings": {
          "type": "array",
          "items": [
            {
              "type": "object",
              "properties": {
                "indexFieldName": {
                  "type": "string"
                },
                "indexFieldType": {
                  "type": "string",
                  "enum": [
                    "STRING",
                    "STRING_LIST",
                    "DATE"
                  ]
                },
                "dataSourceFieldName": {
                  "type": "string"
                },
                "dateFieldFormat": {
                  "type": "string",
                  "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                }
              },
              "required": [
                "indexFieldName",
                "indexFieldType",
                "dataSourceFieldName"
              ]
            }
          ]
        }
      },
      "required": [
        "fieldMappings"
```

```
          ]
        },
        "meetingFile": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING",
                        "DATE",
                        "LONG"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        },
        "meetingNote": {
```

```
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING",
                        "DATE"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        },
        "calendarMeeting": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
```

```
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    }
  },
   "additionalProperties": {
    "type": "object",
    "properties": {
      "paymentModel": {
        "type": "string",
        "enum": [
          "A",
```

```
            "B",
            "Evaluation Mode"
          ]
        },
        "inclusionTeamNameFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionTeamNameFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionChannelNameFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionChannelNameFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionFileNamePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionFileNamePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionFileTypePatterns": {
          "type": "array",
          "items": {
            "type": "string"
```

```
        }
      },
      "exclusionFileTypePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "inclusionUserEmailFilter": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "inclusionOneNoteSectionNamePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "exclusionOneNoteSectionNamePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "inclusionOneNotePageNamePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "exclusionOneNotePageNamePatterns": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "isCrawlChatMessage": {
        "type": "boolean"
      },
      "isCrawlChatAttachment": {
        "type": "boolean"
      },
```

```
        "isCrawlChannelPost": {
          "type": "boolean"
        },
        "isCrawlChannelAttachment": {
          "type": "boolean"
        },
        "isCrawlChannelWiki": {
          "type": "boolean"
        },
        "isCrawlCalendarMeeting": {
          "type": "boolean"
        },
        "isCrawlMeetingChat": {
          "type": "boolean"
        },
        "isCrawlMeetingFile": {
          "type": "boolean"
        },
        "isCrawlMeetingNote": {
          "type": "boolean"
        },
        "startCalendarDateTime": {
          "anyOf": [
            {
              "type": "string",
              "pattern": "^[0-9]{4}-[0-9]{2}-[0-9]{2}T[0-9]{2}:[0-9]{2}:[0-9]{2}Z$"
            },
            {
              "type": "string",
              "pattern": ""
            }
          ]
        },
        "endCalendarDateTime": {
          "anyOf": [
            {
              "type": "string",
              "pattern": "^[0-9]{4}-[0-9]{2}-[0-9]{2}T[0-9]{2}:[0-9]{2}:[0-9]{2}Z$"
            },
            {
              "type": "string",
              "pattern": ""
            }
          ]
```

```
        }
      },
      "required": []
    },
    "type": {
      "type": "string",
      "pattern": "MSTEAMS"
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string",
      "minLength": 20,
      "maxLength": 2048
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "syncMode",
    "additionalProperties",
    "secretArn",
    "type"
  ]
}
```

Microsoft Yammer 模板架构

您将包含包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 YAMMER、身份验证凭证的密钥以及其他必要的配置。然后在呼叫时指定TEMPLATE为 “类型” CreateDataSource。

您可以使用本开发者指南中提供的模板。

下表描述了微软 Yammer JSON 架构的参数。

配置	描述
connectionConfiguration	数据来源的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。此数据来源未在 `repositoryEndpointMetadata` 中指定端点。相反，连接信息包含在您提供的 AWS Secrets Manager 密钥中`secretArn` 。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• community • 用户 • 消息 • attachment	将 Microsoft Yammer 内容的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象的列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项
inclusionPatterns	用于在 Microsoft Yammer 数据来源中包含某些文件的正则表达式模式的列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。
exclusionPatterns	用于在 Microsoft Yammer 数据来源中排除某些文件的正则表达式模式的列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含

配置	描述
	在索引中。如果文件同时匹配排除和包含模式，则以排除模式为优先，该文件不会包含在索引中。
sinceDate	您可以选择配置一个 `sinceDate` 参数，以便 Microsoft Yammer 连接器根据特定的 `sinceDate` 爬取内容。
communityNameFilter	您可以选择将特定的社区内容编入索引。
• isCrawlMessage • isCrawlAttachment • isCrawlPrivate消息	`true`抓取消息、邮件附件和私人消息。
type	指定 `YAMMER` 作为数据来源类型。
secretARN	AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)，其中包含连接你的 Microsoft Yammer 所需的键值对。这包括您的客户端 ID 和在 Azure 门户中创建 OAuth 应用程序时生成的 Microsoft Yammer 用户名和密码、客户端 ID、客户端密钥。
useChangeLog	`true`使用 Microsoft Yammer 更改日志来确定索引中哪些文档需要更新。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
enableIdentityCrawler	true使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。

Microsoft Yammer JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
          }
```

```
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "community": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": {
                "anyOf": [
                  {
                    "type": "object",
                    "properties": {
                      "indexFieldName": {
                        "type": "string"
                      },
                      "indexFieldType": {
                        "type": "string",
                        "enum": [
                          "STRING",
                          "DATE"
                        ]
                      },
                      "dataSourceFieldName": {
                        "type": "string"
                      },
                      "dateFieldFormat": {
                        "type": "string",
                        "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                      }
                    },
                    "required": [
                      "indexFieldName",
                      "indexFieldType",
                      "dataSourceFieldName"
                    ]
                  }
                ]
```

```
            }
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "user": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": {
              "anyOf": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING",
                        "DATE"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          }
```

```
        },
        "required": [
          "fieldMappings"
        ]
      },
      "message": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": {
              "anyOf": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING",
                        "DATE"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          }
        },
        "required": [
```

```
            "fieldMappings"
          ]
        },
        "attachment": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": {
                "anyOf": [
                  {
                    "type": "object",
                    "properties": {
                      "indexFieldName": {
                        "type": "string"
                      },
                      "indexFieldType": {
                        "type": "string",
                        "enum": [
                          "STRING",
                          "DATE"
                        ]
                      },
                      "dataSourceFieldName": {
                        "type": "string"
                      },
                      "dateFieldFormat": {
                        "type": "string",
                        "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                      }
                    },
                    "required": [
                      "indexFieldName",
                      "indexFieldType",
                      "dataSourceFieldName"
                    ]
                  }
                ]
              }
            }
          },
          "required": [
            "fieldMappings"
          ]
```

```
        }
      }
    },
    "additionalProperties": {
      "type": "object",
      "properties": {
        "inclusionPatterns": {
          "type": "array"
        },
        "exclusionPatterns": {
          "type": "array"
        },
        "sinceDate": {
          "type": "string",
          "pattern": "^(19|2[0-9])[0-9]{2}-(0[1-9]|1[012])-(0[1-9]|[12][0-9]|
3[01])T(0[0-9]|1[0-9]|2[0-3]):([0-5][0-9]):([0-5][0-9])((\\+|-)(0[0-9]|1[0-9]|2[0-3]):
([0-5][0-9]))?$"
        },
        "communityNameFilter": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "isCrawlMessage": {
          "type": "boolean"
        },
        "isCrawlAttachment": {
          "type": "boolean"
        },
        "isCrawlPrivateMessage": {
          "type": "boolean"
        }
      },
      "required": [
        "sinceDate"
      ]
    },
    "type": {
      "type": "string",
      "pattern": "YAMMER"
    },
    "secretArn": {
      "type": "string",
```

```
      "minLength": 20,
      "maxLength": 2048
    },
    "useChangeLog": {
      "type": "string",
      "enum": [
        "true",
        "false"
      ]
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "version": {
      "type": "string",
      "anyOf": [
        {
          "pattern": "1.0.0"
        }
      ]
    }
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "additionalProperties",
    "type",
    "secretArn",
    "syncMode"
  ]
}
```

MySQL 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、mysql 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 MySQL JSON 架构。

下表描述了 MySQL JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。 • dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。

配置	描述
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

MySQL JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
```

```
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
```

```
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string"
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "primaryKey": {
        "type": "string"
      },
      "titleColumn": {
        "type": "string"
      },
      "bodyColumn": {
        "type": "string"
      },
      "sqlQuery": {
```

```
          "type": "string",
          "not": {
            "pattern": ";+"
          }
        },
        "timestampColumn": {
          "type": "string"
        },
        "timestampFormat": {
          "type": "string"
        },
        "timezone": {
          "type": "string"
        },
        "changeDetectingColumns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
```

```
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Oracle Database 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、oracle 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Oracle Database JSON 架构。

下表介绍了 Oracle 数据库 JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。

配置	描述
	• dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。

配置	描述
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

配置	描述
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

Oracle Database JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
```

```
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string"
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
```

```
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "primaryKey": {
        "type": "string"
      },
      "titleColumn": {
        "type": "string"
      },
      "bodyColumn": {
        "type": "string"
      },
      "sqlQuery": {
        "type": "string",
        "not": {
          "pattern": ";+"
        }
      },
      "timestampColumn": {
        "type": "string"
      },
      "timestampFormat": {
        "type": "string"
      },
      "timezone": {
        "type": "string"
      },
      "changeDetectingColumns": {
        "type": "array",
        "items": {
          "type": "string"
        }
```

```
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
```

```
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

PostgreSQL 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。将数据来源的类型指定为 JDBC、postgresql 的数据库类型、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 PostgreSQL JSON 架构。

下表描述了 PostgreSQL JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	连接数据来源所需的配置信息。 • dbtype-您使用的 Java 数据库的类型，无论是、、mysqldb2、postgresql 还是。oracle sqlserver • dbHost - 数据库主机名。 • dbPort - 数据库端口。 • dbInstance - 数据库实例。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。指定数据来源的类型和密钥 ARN。
文档	将数据库内容的属性或字段名映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。
additionalProperties	数据来源中内容的其他配置选项。用于在数据库数据来源中包含或排除特定内容。

配置	描述
primaryKey	提供数据库表的主键。这将标识数据库中的表。
titleColumn	提供数据库表中文档标题列的名称。
bodyColumn	提供数据库表中文档标题列的名称。
sqlQuery	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
timestampColumn	输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
timestampFormat	输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。
timezone	输入列的名称，该列包含要搜索的内容的时区。
changeDetectingColumns	输入 Amazon Kendra 将用于检测内容更改的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容
allowedUsersColumns	输入包含允许访问内容的用户 ID 的列的名称。
allowedGroupsColumn	输入包含允许访问内容的用户 ID 的列的名称。
sourceURIColumn	输入包含要编制索引的源 URL 的列的名称。
isSslEnabled	输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
type	数据来源的类型。指定 JDBC 作为数据来源类型。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
secretArn	Secrets Manager 密钥的 Amazon 资源名称（ARN），其中包含用于连接到数据库的用户名和密码。密钥必须包含具有以下键的 JSON 结构： `{ "user name": "database user name", "password": " password" }`
版本	当前支持的此模板的版本。

PostgreSQL JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
```

```
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "dbType": {
              "type": "string",
              "enum": [
                "mysql",
                "db2",
                "postgresql",
                "oracle",
                "sqlserver"
              ]
            },
            "dbHost": {
              "type": "string"
            },
            "dbPort": {
              "type": "string"
            },
            "dbInstance": {
              "type": "string"
            }
          },
          "required": [
            "dbType",
            "dbHost",
            "dbPort",
            "dbInstance"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "document": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
```

```
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string"
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    },
    "required": [
    ]
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "primaryKey": {
        "type": "string"
      },
      "titleColumn": {
        "type": "string"
      },
      "bodyColumn": {
        "type": "string"
      },
      "sqlQuery": {
```

```
          "type": "string",
          "not": {
            "pattern": ";+"
          }
        },
        "timestampColumn": {
          "type": "string"
        },
        "timestampFormat": {
          "type": "string"
        },
        "timezone": {
          "type": "string"
        },
        "changeDetectingColumns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "allowedUsersColumn": {
          "type": "string"
        },
        "allowedGroupsColumn": {
          "type": "string"
        },
        "sourceURIColumn": {
          "type": "string"
        },
        "isSslEnabled": {
          "type": "boolean"
        }
      },
      "required": ["primaryKey", "titleColumn", "bodyColumn", "sqlQuery"]
    },
    "type" : {
      "type" : "string",
      "pattern": "JDBC"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL",
```

```
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
      "connectionConfiguration",
      "repositoryConfigurations",
      "syncMode",
      "additionalProperties",
      "secretArn",
      "type"
  ]
}
```

Salesforce 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您可以将 Salesforc 主机 URL 作为连接配置或存储库端点详细信息的一部分提供。还要将数据来源的类型指定为 SALESFORCEV2、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Salesforce JSON 架构。

下表描述了 Salesforce JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。

配置	描述
hostUrl	要编制索引的 Salesforce 实例的 URL。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• account • contact • 市场活动 • 案例 • product • lead • contract • partner • 配置文件 • idea • pricebook • task • solution • attachment • 用户 • 文档 • knowledgeArticles • 组 • opportunity • chatter • customEntity	将您的 Salesforce 实体的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。有关更多信息，请参阅映射数据来源字段。

配置	描述

secretARN	包含连接您的 Salesforce 所需的键值对的 AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须包含具有以下键的 JSON 结构：
	
	{
	"authenticationUrl": " *OAUTH endpoint that Amazon Kendra connects to get an OAUTH token*",
	"consumerKey": " *Application public key generated when you created your Salesforce application* ",
	"consumerSecret": " *Application private key generated when you created your Salesforce application* ",
	"password": " *Password associated with the user logging in to the Salesforce instance* ",
	"securityToken": " *Token associated with the user account logging in to the Salesforce instance* ",
	"username": " *User name of the user logging in to the Salesforce instance*"
	}

additionalProperties	数据来源中内容的其他配置选项

配置	描述
• accountFilter • contactFilter • caseFilter • campaignFilter • contractFilter • groupFilter • leadFilter • productFilter • opportunityFilter • partnerFilter • pricebookFilter • ideaFilter • profileFilter • taskFilter • solutionFilter • userFilter • chatterFilter • documentFilter • knowledgeArticleFilter • customEntities	一组字符串，用于指定要筛选的实体。

配置	描述
inclusionPatterns • inclusionDocumentFileTypePatterns • inclusionDocumentFileNamePatterns • inclusionAccountFileTypePatterns • inclusionCampaignFileTypePatterns • inclusionDocumentFileNamePatterns • inclusionCampaignFileNamePatterns • inclusionCaseFileTypePatterns • inclusionCaseFileNamePatterns • inclusionContactFileTypePatterns • inclusionContractFileNamePatterns • inclusionLeadFileTypePatterns • inclusionLeadFileNamePatterns • inclusionOpportunityFileTypePatterns • inclusionOpportunityFileNamePatterns • inclusionSolutionFileTypePatterns • inclusionSolutionFileNamePatterns • inclusionTaskFileTypePatterns • inclusionTaskFileNamePatterns • inclusionGroupFileTypePatterns • inclusionGroupFileNamePatterns • inclusionChatterFileTypePatterns • inclusionChatterFileNamePatterns • inclusionCustomEntityFileTypePatterns • inclusionCustomEntityFileNamePatterns	用于在 Salesforce 数据来源中包含某些文件的正则表达式模式的列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。

配置	描述
exclusionPatterns • exclusionDocumentFileTypePatterns • exclusionDocumentFileNamePatterns • exclusionAccountFileTypePatterns • exclusionCampaignFileTypePatterns • exclusionCampaignFileNamePatterns • exclusionCaseFileTypePatterns • exclusionCaseFileNamePatterns • exclusionContactFileTypePatterns • exclusionContractFileNamePatterns • exclusionLeadFileTypePatterns • exclusionLeadFileNamePatterns • exclusionOpportunityFileTypePatterns • exclusionOpportunityFileNamePatterns • exclusionSolutionFileTypePatterns • exclusionSolutionFileNamePatterns • exclusionTaskFileTypePatterns • exclusionTaskFileNamePatterns • exclusionGroupFileTypePatterns • exclusionGroupFileNamePatterns • exclusionChatterFileTypePatterns • exclusionChatterFileNamePatterns • exclusionCustomEntityFileTypePatterns • exclusionCustomEntityFileNamePatterns	用于在 Salesforce 数据来源中排除某些文件的正则表达式模式的列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配排除和包含模式，则以排除模式为优先，该文件不会包含在索引中。

配置	描述
• isCrawlAccount • isCrawlContact • isCrawlCase • isCrawlCampaign • isCrawlProduct • isCrawlLead • isCrawlContract • isCrawlPartner • isCrawlProfile • isCrawlIdea • isCrawlPricebook • isCrawlDocument • crawlSharedDocument • isCrawlGroup • isCrawlOpportunity • isCrawlChatter • isCrawlUser • isCrawlSolution • isCrawlTask • isCrawlAccount附件 • isCrawlContact附件 • isCrawlCase附件 • isCrawlCampaign附件 • isCrawlLead附件 • isCrawlContract附件 • isCrawlGroup附件 • isCrawlOpportunity附件 • isCrawlChatter附件 • isCrawlSolution附件	`true`在你的 Salesforce 账户中抓取这些类型的文件。

配置	描述

• isCrawlTask附件
• isCrawlCustomEntityAttachments
• isCrawlKnowledge文章
  • isCrawlDraft
  • isCrawlPublish
  • isCrawlArchived	

type	数据来源的类型。指定 SALESFORCEV2 作为数据来源类型。
enableIdentityCrawler	true使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • CHANGE_LOG 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
版本	当前支持的此模板的版本。

Salesforce JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties":
  {
    "connectionConfiguration": {
      "type": "object",
      "properties":
      {
        "repositoryEndpointMetadata":
        {
          "type": "object",
          "properties":
          {
            "hostUrl":
            {
              "type": "string",
              "pattern": "https:.*"
            }
          },
          "required":
          [
            "hostUrl"
          ]
        }
      },
      "required":
      [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties":
      {
        "account":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
```

```
            "type": "array",
            "items":
            [
              {
                "type": "object",
                "properties":
                {
                  "indexFieldName":
                  {
                    "type": "string"
                  },
                  "indexFieldType":
                  {
                    "type": "string",
                    "enum":
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName":
                  {
                    "type": "string"
                  },
                  "dateFieldFormat":
                  {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required":
                [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
```

```
            "fieldMappings"
          ]
        },
        "contact":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
                      [
                        "STRING",
                        "STRING_LIST",
                        "DATE"
                      ]
                    },
                    "dataSourceFieldName":
                    {
                      "type": "string"
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required":
                  [
                    "indexFieldName",
```

```
                    "indexFieldType",
                    "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
      },
      "campaign":
      {
        "type": "object",
        "properties":
        {
          "fieldMappings":
          {
            "type": "array",
            "items":
            [
              {
                "type": "object",
                "properties":
                {
                  "indexFieldName":
                  {
                    "type": "string"
                  },
                  "indexFieldType":
                  {
                    "type": "string",
                    "enum":
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName":
                  {
                    "type": "string"
```

```
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required":
                  [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required":
          [
            "fieldMappings"
          ]
        },
        "case":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
```

```
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName":
                  {
                    "type": "string"
                  },
                  "dateFieldFormat":
                  {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required":
                [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
      },
      "product":
      {
        "type": "object",
        "properties":
        {
          "fieldMappings":
          {
            "type": "array",
            "items":
            [
              {
                "type": "object",
                "properties":
```

```
                {
                  "indexFieldName":
                  {
                    "type": "string"
                  },
                  "indexFieldType":
                  {
                    "type": "string",
                    "enum":
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName":
                  {
                    "type": "string"
                  },
                  "dateFieldFormat":
                  {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required":
                [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
      },
      "lead":
      {
        "type": "object",
        "properties":
```

```
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
                      [
                        "STRING",
                        "STRING_LIST",
                        "DATE",
                        "LONG"
                      ]
                    },
                    "dataSourceFieldName":
                    {
                      "type": "string"
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required":
                  [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
```

```
          },
          "required":
          [
            "fieldMappings"
          ]
        },
        "contract":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
                      [
                        "STRING",
                        "STRING_LIST",
                        "DATE"
                      ]
                    },
                    "dataSourceFieldName":
                    {
                      "type": "string"
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
```

```
                "required":
                [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
      },
      "partner":
      {
        "type": "object",
        "properties":
        {
          "fieldMappings":
          {
            "type": "array",
            "items":
            [
              {
                "type": "object",
                "properties":
                {
                  "indexFieldName":
                  {
                    "type": "string"
                  },
                  "indexFieldType":
                  {
                    "type": "string",
                    "enum":
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName":
```

```
                {
                  "type": "string"
                },
                "dateFieldFormat":
                {
                  "type": "string",
                  "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                }
              },
              "required":
              [
                "indexFieldName",
                "indexFieldType",
                "dataSourceFieldName"
              ]
            }
          ]
        }
      },
      "required":
      [
        "fieldMappings"
      ]
    },
    "profile":
    {
      "type": "object",
      "properties":
      {
        "fieldMappings":
        {
          "type": "array",
          "items":
          [
            {
              "type": "object",
              "properties":
              {
                "indexFieldName":
                {
                  "type": "string"
                },
                "indexFieldType":
                {
```

```
                    "type": "string",
                    "enum":
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName":
                  {
                    "type": "string"
                  },
                  "dateFieldFormat":
                  {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required":
                [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
      },
      "idea":
      {
        "type": "object",
        "properties":
        {
          "fieldMappings":
          {
            "type": "array",
            "items":
            [
              {
```

```
                "type": "object",
                "properties":
                {
                  "indexFieldName":
                  {
                    "type": "string"
                  },
                  "indexFieldType":
                  {
                    "type": "string",
                    "enum":
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE",
                      "LONG"
                    ]
                  },
                  "dataSourceFieldName":
                  {
                    "type": "string"
                  },
                  "dateFieldFormat":
                  {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required":
                [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
      },
      "pricebook":
```

```
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
                      [
                        "STRING",
                        "STRING_LIST",
                        "DATE"
                      ]
                    },
                    "dataSourceFieldName":
                    {
                      "type": "string"
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required":
                  [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
```

```
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
      },
      "task":
      {
        "type": "object",
        "properties":
        {
          "fieldMappings":
          {
            "type": "array",
            "items":
            [
              {
                "type": "object",
                "properties":
                {
                  "indexFieldName":
                  {
                    "type": "string"
                  },
                  "indexFieldType":
                  {
                    "type": "string",
                    "enum":
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName":
                  {
                    "type": "string"
                  },
                  "dateFieldFormat":
                  {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
```

```
                    }
                  },
                  "required":
                  [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required":
          [
            "fieldMappings"
          ]
        },
        "solution":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
                      [
                        "STRING",
                        "STRING_LIST",
                        "DATE"
                      ]
```

```
                    },
                    "dataSourceFieldName":
                    {
                      "type": "string"
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required":
                  [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required":
          [
            "fieldMappings"
          ]
        },
        "attachment":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
```

```
                "indexFieldType":
                {
                  "type": "string",
                  "enum":
                  [
                    "STRING",
                    "STRING_LIST",
                    "DATE",
                    "LONG"
                  ]
                },
                "dataSourceFieldName":
                {
                  "type": "string"
                },
                "dateFieldFormat":
                {
                  "type": "string",
                  "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                }
              },
              "required":
              [
                "indexFieldName",
                "indexFieldType",
                "dataSourceFieldName"
              ]
            }
          ]
        }
      },
      "required":
      [
        "fieldMappings"
      ]
    },
    "user":
    {
      "type": "object",
      "properties":
      {
        "fieldMappings":
        {
          "type": "array",
```

```
            "items":
            [
              {
                "type": "object",
                "properties":
                {
                  "indexFieldName":
                  {
                    "type": "string"
                  },
                  "indexFieldType":
                  {
                    "type": "string",
                    "enum":
                    [
                      "STRING",
                      "STRING_LIST",
                      "DATE"
                    ]
                  },
                  "dataSourceFieldName":
                  {
                    "type": "string"
                  },
                  "dateFieldFormat":
                  {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required":
                [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required":
        [
          "fieldMappings"
        ]
```

```
        },
        "document":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
                      [
                        "STRING",
                        "STRING_LIST",
                        "DATE",
                        "LONG"
                      ]
                    },
                    "dataSourceFieldName":
                    {
                      "type": "string"
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required":
                  [
                    "indexFieldName",
                    "indexFieldType",
```

```
                "dataSourceFieldName"
              ]
            }
          ]
        }
      },
      "required":
      [
        "fieldMappings"
      ]
    },
    "knowledgeArticles":
    {
      "type": "object",
      "properties":
      {
        "fieldMappings":
        {
          "type": "array",
          "items":
          [
            {
              "type": "object",
              "properties":
              {
                "indexFieldName":
                {
                  "type": "string"
                },
                "indexFieldType":
                {
                  "type": "string",
                  "enum":
                  [
                    "STRING",
                    "STRING_LIST",
                    "DATE"
                  ]
                },
                "dataSourceFieldName":
                {
                  "type": "string"
                },
                "dateFieldFormat":
```

```
                {
                  "type": "string",
                  "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                }
              },
              "required":
              [
                "indexFieldName",
                "indexFieldType",
                "dataSourceFieldName"
              ]
            }
          ]
        }
      },
      "required":
      [
        "fieldMappings"
      ]
    },
    "group":
    {
      "type": "object",
      "properties":
      {
        "fieldMappings":
        {
          "type": "array",
          "items":
          [
            {
              "type": "object",
              "properties":
              {
                "indexFieldName":
                {
                  "type": "string"
                },
                "indexFieldType":
                {
                  "type": "string",
                  "enum":
                  [
                    "STRING",
```

```
                    "STRING_LIST",
                    "DATE"
                  ]
                },
                "dataSourceFieldName":
                {
                  "type": "string"
                },
                "dateFieldFormat":
                {
                  "type": "string",
                  "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                }
              },
              "required":
              [
                "indexFieldName",
                "indexFieldType",
                "dataSourceFieldName"
              ]
            }
          ]
        }
      },
      "required":
      [
        "fieldMappings"
      ]
    },
    "opportunity":
    {
      "type": "object",
      "properties":
      {
        "fieldMappings":
        {
          "type": "array",
          "items":
          [
            {
              "type": "object",
              "properties":
              {
                "indexFieldName":
```

```
                {
                  "type": "string"
                },
                "indexFieldType":
                {
                  "type": "string",
                  "enum":
                  [
                    "STRING",
                    "STRING_LIST",
                    "DATE",
                    "LONG"
                  ]
                },
                "dataSourceFieldName":
                {
                  "type": "string"
                },
                "dateFieldFormat":
                {
                  "type": "string",
                  "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                }
              },
              "required":
              [
                "indexFieldName",
                "indexFieldType",
                "dataSourceFieldName"
              ]
            }
          ]
        }
      },
      "required":
      [
        "fieldMappings"
      ]
    },
    "chatter":
    {
      "type": "object",
      "properties":
      {
```

```
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
                      [
                        "STRING",
                        "STRING_LIST",
                        "DATE"
                      ]
                    },
                    "dataSourceFieldName":
                    {
                      "type": "string"
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required":
                  [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required":
```

```
          [
            "fieldMappings"
          ]
        },
        "customEntity":
        {
          "type": "object",
          "properties":
          {
            "fieldMappings":
            {
              "type": "array",
              "items":
              [
                {
                  "type": "object",
                  "properties":
                  {
                    "indexFieldName":
                    {
                      "type": "string"
                    },
                    "indexFieldType":
                    {
                      "type": "string",
                      "enum":
                      [
                        "STRING",
                        "STRING_LIST",
                        "DATE"
                      ]
                    },
                    "dataSourceFieldName":
                    {
                      "type": "string"
                    },
                    "dateFieldFormat":
                    {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required":
                  [
```

```
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required":
          [
            "fieldMappings"
          ]
        }
      }
    },
    "additionalProperties": {
      "type": "object",
      "properties":
      {
        "accountFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "contactFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "caseFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "campaignFilter":{
          "type": "array",
          "items":
          {
```

```
            "type": "string"
          }
        },
        "contractFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "groupFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "leadFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "productFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "opportunityFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "partnerFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
```

```
        },
        "pricebookFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "ideaFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "profileFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "taskFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "solutionFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "userFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "chatterFilter":{
```

```
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "documentFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "knowledgeArticleFilter":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "customEntities":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "isCrawlAccount": {
          "type": "boolean"
        },
        "isCrawlContact": {
          "type": "boolean"
        },
        "isCrawlCase": {
          "type": "boolean"
        },
        "isCrawlCampaign": {
          "type": "boolean"
        },
        "isCrawlProduct": {
          "type": "boolean"
        },
        "isCrawlLead": {
          "type": "boolean"
```

```
        },
        "isCrawlContract": {
          "type": "boolean"
        },
        "isCrawlPartner": {
          "type": "boolean"
        },
        "isCrawlProfile": {
          "type": "boolean"
        },
        "isCrawlIdea": {
          "type": "boolean"
        },
        "isCrawlPricebook": {
          "type": "boolean"
        },
        "isCrawlDocument": {
          "type": "boolean"
        },
        "crawlSharedDocument": {
          "type": "boolean"
        },
        "isCrawlGroup": {
          "type": "boolean"
        },
        "isCrawlOpportunity": {
          "type": "boolean"
        },
        "isCrawlChatter": {
          "type": "boolean"
        },
        "isCrawlUser": {
          "type": "boolean"
        },
        "isCrawlSolution":{
          "type": "boolean"
        },
        "isCrawlTask":{
          "type": "boolean"
        },

        "isCrawlAccountAttachments": {
          "type": "boolean"
        },
```

```
        "isCrawlContactAttachments": {
          "type": "boolean"
        },
        "isCrawlCaseAttachments": {
          "type": "boolean"
        },
        "isCrawlCampaignAttachments": {
          "type": "boolean"
        },
        "isCrawlLeadAttachments": {
          "type": "boolean"
        },
        "isCrawlContractAttachments": {
          "type": "boolean"
        },
        "isCrawlGroupAttachments": {
          "type": "boolean"
        },
        "isCrawlOpportunityAttachments": {
          "type": "boolean"
        },
        "isCrawlChatterAttachments": {
          "type": "boolean"
        },
        "isCrawlSolutionAttachments":{
          "type": "boolean"
        },
        "isCrawlTaskAttachments":{
          "type": "boolean"
        },
        "isCrawlCustomEntityAttachments":{
          "type": "boolean"
        },
        "isCrawlKnowledgeArticles": {
          "type": "object",
          "properties":
          {
            "isCrawlDraft": {
              "type": "boolean"
            },
            "isCrawlPublish": {
              "type": "boolean"
            },
            "isCrawlArchived": {
```

```
            "type": "boolean"
          }
        }
      },
      "inclusionDocumentFileTypePatterns":{
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "exclusionDocumentFileTypePatterns": {
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "inclusionDocumentFileNamePatterns":{
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "exclusionDocumentFileNamePatterns": {
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "inclusionAccountFileTypePatterns": {
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "exclusionAccountFileTypePatterns": {
        "type": "array",
        "items":
        {
          "type": "string"
```

```
        }
      },
      "inclusionAccountFileNamePatterns":{
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "exclusionAccountFileNamePatterns":{
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "inclusionCampaignFileTypePatterns": {
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "exclusionCampaignFileTypePatterns": {
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "inclusionCampaignFileNamePatterns":{
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
      "exclusionCampaignFileNamePatterns":{
        "type": "array",
        "items":
        {
          "type": "string"
        }
      },
```

```
        "inclusionCaseFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionCaseFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionCaseFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionCaseFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionContactFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionContactFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionContactFileNamePatterns":{
          "type": "array",
```

```
          "items":
          {
            "type": "string"
          }
        },
        "exclusionContactFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionContractFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionContractFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionContractFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionContractFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionLeadFileTypePatterns":{
          "type": "array",
          "items":
          {
```

```
            "type": "string"
          }
        },
        "exclusionLeadFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionLeadFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionLeadFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionOpportunityFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionOpportunityFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionOpportunityFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
```

```
        },
        "exclusionOpportunityFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionSolutionFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionSolutionFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionSolutionFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionSolutionFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionTaskFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionTaskFileTypePatterns":{
```

```
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionTaskFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionTaskFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionGroupFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionGroupFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionGroupFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionGroupFileNamePatterns":{
          "type": "array",
          "items":
```

```
          {
            "type": "string"
          }
        },
        "inclusionChatterFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionChatterFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionChatterFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionChatterFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "inclusionCustomEntityFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionCustomEntityFileTypePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
```

```
          }
        },
        "inclusionCustomEntityFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        },
        "exclusionCustomEntityFileNamePatterns":{
          "type": "array",
          "items":
          {
            "type": "string"
          }
        }
      },
      "required":
      []
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "type": {
      "type": "string",
      "pattern": "SALESFORCEV2"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FULL_CRAWL",
        "FORCED_FULL_CRAWL",
        "CHANGE_LOG"
      ]
    },
    "secretArn": {
      "type": "string",
      "minLength": 20,
      "maxLength": 2048
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
```

```
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "syncMode",
    "additionalProperties",
    "secretArn",
    "type"
  ]
}
```

ServiceNow 模板架构

您可以将包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您可以在连接配置或存储库端点详细信息中提供 ServiceNow 主机 URL、身份验证类型和实例版本。还要将数据来源的类型指定为 SERVICENOWV2、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 ServiceNow JSON 模式。

下表描述了 ServiceNow JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
hostUrl	ServiceNow 主机网址。例如，*your-domain.service-now.com* 。
authType	您使用的身份验证类型，可以是 basicAuth 或 OAuth2。
servicenowInstanceVersion	您使用的 ServiceNow 版本。您可以在Tokyo、SandiegoRome、和之间进行选择Others。

配置	描述
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• knowledgeArticle • attachment • serviceCatalog • incident	将 ServiceNow知识文章、附件、服务目录和事件的属性或字段名称映射到 Amazon Kendra 索引字段名的对象列表。有关更多信息，请参阅映射数据来源字段。ServiceNow 数据源字段名称必须存在于您的 ServiceNow 自定义元数据中。
其他属性	数据来源中内容的其他配置选项。
maxFileSizeInMegaBytes	指定 Amazon Kendra 将抓取的文件大小限制（以 MB 为单位）。Amazon Kendra 只会抓取您定义的大小限制内的文件。默认文件大小为 50MB。最大文件大小应大于 0MB 且小于或等于 50MB。
• knowledgeArticleFilter • incidentQueryFilter • serviceCatalogQuery过滤器 • knowledgeArticleTitleRegExp • serviceCatalogTitleRegExp • incidentTitleReg经验值 • inclusionFileType图案 • exclusionFileType图案 • inclusionFileName图案 • exclusionFileName图案 • incidentStateType	用于在 ServiceNow 数据源中包含和/或排除某些文件的正则表达式模式列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。

配置	描述
• isCrawlKnowledge文章 • isCrawlKnowledgeArticleAttachment • includePublicArticles只有 • isCrawlService目录 • isCrawlServiceCatalogAttachment • isCrawlActiveServiceCatalog • isCrawlInactiveServiceCatalog • isCrawlIncident • isCrawlIncident附件 • isCrawlActive事件 • isCrawlInactive事件 • ApplyACL ForKnowledgeArticle • ApplyACL ForServiceCatalog • ApplyACL ForIncident	`true`搜寻 ServiceNow 知识文章、服务目录、事件和附件。
type	数据来源的类型。指定 `SERVICENOWV2` 作为数据来源类型。
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。

配置	描述
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • FORCED_FULL_CRAWL 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • FULL_CRAWL 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

secretARN	AWS Secrets Manager 密钥的 Amazon 资源名称 (ARN)，其中包含连接到您的所需的键值对。 ServiceNow密钥必须包含具有以下键的 JSON 结构：
	`{ "username": " user name", "password": " password" }`
	如果您使用 OAuth 2 身份验证，则密钥必须包含具有以下键的 JSON 结构：
	`{ "username": " user name", "password": " password", "clientId": " client id", "clientSecret": " client secret" }`

版本	当前支持的此模板的版本。

ServiceNow JSON 模式

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
```

```
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "hostUrl": {
              "type": "string",
              "pattern": "^(?!(^(https?|ftp|file):\/\/))[a-z0-9-]+(.service-
now.com|.servicenowservices.com)$",
              "minLength": 1,
              "maxLength": 2048
            },
            "authType": {
              "type": "string",
              "enum": [
                "basicAuth",
                "OAuth2"
              ]
            },
            "servicenowInstanceVersion": {
              "type": "string",
              "enum": [
                "Tokyo",
                "Sandiego",
                "Rome",
                "Others"
                ]
            }
          },
          "required": [
            "hostUrl",
            "authType",
            "servicenowInstanceVersion"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
```

```
      "properties": {
        "knowledgeArticle": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING",
                        "DATE",
                        "STRING_LIST"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        },
        "attachment": {
          "type": "object",
```

```
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": [
                        "STRING",
                        "LONG",
                        "DATE",
                        "STRING_LIST"
                      ]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        },
        "serviceCatalog": {
          "type": "object",
          "properties": {
            "fieldMappings": {
```

```
            "type": "array",
            "items": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "DATE",
                      "STRING_LIST"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "incident": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": [
              {
```

```
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": [
                      "STRING",
                      "DATE",
                      "STRING_LIST"
                    ]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        },
        "required": [
          "fieldMappings"
        ]
      }
    }
  },
  "additionalProperties": {
    "type": "object",
    "properties": {
      "maxFileSizeInMegaBytes": {
        "type": "string"
      },
      "isCrawlKnowledgeArticle": {
        "type": "boolean"
```

```
        },
        "isCrawlKnowledgeArticleAttachment": {
          "type": "boolean"
        },
        "includePublicArticlesOnly": {
          "type": "boolean"
        },
        "knowledgeArticleFilter": {
          "type": "string"
        },
        "incidentQueryFilter": {
          "type": "string"
        },
        "serviceCatalogQueryFilter": {
          "type": "string"
        },
        "isCrawlServiceCatalog": {
          "type": "boolean"
        },
        "isCrawlServiceCatalogAttachment": {
          "type": "boolean"
        },
        "isCrawlActiveServiceCatalog": {
          "type": "boolean"
        },
        "isCrawlInactiveServiceCatalog": {
          "type": "boolean"
        },
        "isCrawlIncident": {
          "type": "boolean"
        },
        "isCrawlIncidentAttachment": {
          "type": "boolean"
        },
        "isCrawlActiveIncident": {
          "type": "boolean"
        },
        "isCrawlInactiveIncident": {
          "type": "boolean"
        },
        "applyACLForKnowledgeArticle": {
          "type": "boolean"
        },
        "applyACLForServiceCatalog": {
```

```
          "type": "boolean"
        },
        "applyACLForIncident": {
          "type": "boolean"
        },
        "incidentStateType": {
          "type": "array",
          "items": {
            "type": "string",
            "enum": [
              "Open",
              "Open - Unassigned",
              "Resolved",
              "All"
            ]
          }
        },
        "knowledgeArticleTitleRegExp": {
          "type": "string"
        },
        "serviceCatalogTitleRegExp": {
          "type": "string"
        },
        "incidentTitleRegExp": {
          "type": "string"
        },
        "inclusionFileTypePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "exclusionFileTypePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionFileNamePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
```

```
        "exclusionFileNamePatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        }
      },
      "required": []
    },
    "type": {
      "type": "string",
      "pattern": "SERVICENOWV2"
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "syncMode": {
      "type": "string",
      "enum": [
        "FORCED_FULL_CRAWL",
        "FULL_CRAWL"
      ]
    },
    "secretArn": {
      "type": "string",
      "minLength": 20,
      "maxLength": 2048
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "syncMode",
    "additionalProperties",
    "secretArn",
    "type"
```

```
  ]
}
```

Slack 模板架构

您将包含包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您可以将主机 URL 作为连接配置或存储库端点详细信息的一部分提供。还要将数据来源的类型指定为 SLACK、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 松弛 JSON 架构。

下表描述了 Slack JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。
teamID	你从 Slack 主页网址中复制的 Slack 团队编号。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
全部	将Slack内容的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象列表。
additionalProperties	数据来源中内容的其他配置选项。
inclusionPatterns	用于在Slack数据源中包含特定内容的正则表达式模式列表。与模式匹配的内容将包含在索引中。与模式不匹配的内容将从索引中排除。如果任何内容同时匹配包含模式和排除模式，则排除模式优先，并且该内容不会包含在索引中。
exclusionPatterns	用于排除Slack数据源中特定内容的正则表达式模式列表。与模式匹配的内容将从索引中排除。与模式不匹配的内容包含在索引中。如果任何内容同时匹配包含模式和排除模式，则排除模式优先，并且该内容不会包含在索引中。

配置	描述
crawlBotMessages	`true`抓取机器人消息。
排除已存档	`true`以排除对存档邮件的抓取。
对话类型	您要索引的对话类型`PUBLIC_CHANNEL` ，是否为`PRIVATE_CHANNEL` 、`GROUP_MESSAGE` 和`DIRECT_MESSAGE` 。
频道过滤器	您要索引的频道类型`private_channel` 是否为`public_channel` 。
sinceDate	您可以选择配置`sinceDate` 参数，以便Slack连接器根据特定`sinceDate` 参数抓取内容。
回顾	您可以选择配置一个`lookBack`参数，以便连接器在上次Slack连接器同步之前的指定小时数内抓取已更新或已删除的内容。
syncMode	指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。您可以选择： • `FORCED_FULL_CRAWL` 对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。 • `FULL_CRAWL` 每次数据源与索引同步时，仅对新的、修改过的和删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。 • `CHANGE_LOG` 每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
type	数据来源的类型。指定 `SLACK` 作为数据来源类型。

配置	描述
enableIdentityCrawler	`true`使用 Amazon Kendra身份搜寻器同步有权访问某些文档的用户和群组的身份/主体信息。如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则也可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息。
secretArn	AWS Secrets Manager 密钥的 Amazon 资源名称 (ARN)，其中包含连接到您的所需的键值对。Slack密钥必须包含具有以下键的 JSON 结构： `{ "slackToken": " token" }`
版本	当前支持的此模板的版本。

松弛 JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
          "properties": {
            "teamId": {
              "type": "string"
            }
          },
          "required": ["teamId"]
        }
      }
```

```
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "All": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": ["STRING", "STRING_LIST", "DATE","LONG"]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "yyyy-MM-dd'T'HH:mm:ss'Z'"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          },
          "required": [
            "fieldMappings"
          ]
        }
      },
      "required": [
      ]
```

```
    },
    "additionalProperties": {
      "type": "object",
      "properties": {
        "exclusionPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "inclusionPatterns": {
          "type": "array",
          "items": {
            "type": "string"
          }
        },
        "crawlBotMessages": {
          "type": "boolean"
        },
        "excludeArchived": {
          "type": "boolean"
        },
        "conversationType": {
          "type": "array",
          "items": {
            "type": "string",
            "enum": [
              "PUBLIC_CHANNEL",
              "PRIVATE_CHANNEL",
              "GROUP_MESSAGE",
              "DIRECT_MESSAGE"
            ]
          }
        },
        "channelFilter": {
            "type": "object",
            "properties": {
              "private_channel": {
                "type": "array",
                "items": {
                  "type": "string"
                }
              },
              "public_channel": {
```

```
              "type": "array",
              "items": {
                "type": "string"
              }
            }
        }
      },
      "channelIdFilter": {
        "type": "array",
        "items": {
          "type": "string"
        }
      },
      "sinceDate": {
        "anyOf": [
          {
            "type": "string",
            "pattern": "^[0-9]{4}-[0-9]{2}-[0-9]{2}T[0-9]{2}:[0-9]{2}:[0-9]{2}Z$"
          },
          {
            "type": "string",
            "pattern": ""
          }
        ]
      },
      "lookBack": {
        "type": "string",
        "pattern": "^[0-9]*$"
      }
    },
    "required": [
    ]
  },
  "syncMode": {
    "type": "string",
    "enum": [
      "FORCED_FULL_CRAWL",
      "FULL_CRAWL",
      "CHANGE_LOG"
    ]
  },
  "type" : {
    "type" : "string",
    "pattern": "SLACK"
```

```
    },
    "enableIdentityCrawler": {
      "type": "boolean"
    },
    "secretArn": {
      "type": "string"
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "syncMode",
    "additionalProperties",
    "secretArn",
    "type",
    "enableIdentityCrawler"
  ]
}
```

Zendesk 模板架构

您将包含包含数据源架构的 JSON 作为TemplateConfiguration对象的一部分。您可以将主机 URL 作为连接配置或存储库端点详细信息的一部分提供。还要将数据来源的类型指定为 ZENDESK、身份验证凭证的密钥以及其他必要的配置。然后，您可以将呼叫Type时间指定TEMPLATE为CreateDataSource。

您可以使用本开发者指南中提供的模板。请参阅 Zendesk JSON 架构。

下表描述了 Zendesk JSON 架构的参数。

配置	描述
connectionConfiguration	有关数据来源端点的配置信息。
repositoryEndpointMetadata	数据来源的端点信息。

配置	描述
hostURL	Zendesk 主机 URL。例如，https://yoursubdomain.zendesk.com。
repositoryConfigurations	数据来源内容的配置信息。例如，配置特定类型的内容和字段映射。
• ticket • ticketComment • ticketCommentAttachment • article • articleComment • articleAttachment • communityTopic • communityPostComment	将 Zendesk 工单的属性或字段名称映射到 Amazon Kendra 索引字段名称的对象的列表。有关更多信息，请参阅映射数据来源字段。
secretARN	包含连接到 Zendesk 所需的键值对的 AWS Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须包含具有以下键的 JSON 结构：主机 URL、客户端 ID、客户端密钥、用户名和密码。
additionalProperties	数据来源中内容的其他配置选项
organizationNameFilter	您可以选择为特定组织中存在的工单编制索引。
sinceDate	您可以选择配置一个 `sinceDate` 参数，以便 Zendesk 连接器根据特定的 `sinceDate` 爬取内容。
inclusionPatterns	用于在 Zendesk 数据来源中包含某些文件的正则表达式模式的列表。与模式匹配的文件将包含在索引中。与模式不匹配的文件将从索引中排除。如果文件同时匹配包含和排除模式，则以排除模式为优先，该文件不会包含在索引中。

配置	描述
exclusionPatterns	用于在 Zendesk 数据来源中排除某些文件的正则表达式模式的列表。与模式匹配的文件将从索引中排除。与模式不匹配的文件将包含在索引中。如果文件同时匹配排除和包含模式，则以排除模式为优先，该文件不会包含在索引中。
• isCrawlTicket • isCrawlTicket评论 • isCrawlTicketCommentAttachment • isCrawlArticle • isCrawlArticle评论 • isCrawlArticle附件 • isCrawlCommunity话题 • isCrawlCommunity帖子 • isCrawlCommunityPostComment	输入 “true” 可抓取这些类型的内容。
type	指定 ZENDESK 作为数据来源类型。
useChangeLog	输入 “true” 以使用 Zendesk 更改日志来确定索引中哪些文档需要更新。根据更改日志的大小，在 Zendesk 中扫描文档可能会更快。如果您是首次将 Zendesk 数据来源与索引同步，则会扫描所有文档。

Zendesk JSON 架构

```
{
  "$schema": "http://json-schema.org/draft-04/schema#",
  "type": "object",
  "properties": {
    "connectionConfiguration": {
      "type": "object",
      "properties": {
        "repositoryEndpointMetadata": {
          "type": "object",
```

```
          "properties": {
            "hostUrl": {
              "type": "string",
              "pattern": "https:.*"
            }
          },
          "required": [
            "hostUrl"
          ]
        }
      },
      "required": [
        "repositoryEndpointMetadata"
      ]
    },
    "repositoryConfigurations": {
      "type": "object",
      "properties": {
        "ticket": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": {
                "anyOf": [
                  {
                    "type": "object",
                    "properties": {
                      "indexFieldName": {
                        "type": "string"
                      },
                      "indexFieldType": {
                        "type": "string",
                        "enum": ["STRING", "STRING_LIST", "LONG", "DATE"]
                      },
                      "dataSourceFieldName": {
                        "type": "string"
                      },
                      "dateFieldFormat": {
                        "type": "string",
                        "pattern": "dd-MM-yyyy HH:mm:ss"
                      }
                    },
                    "required": [
```

```
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"

                ]
              }
            ]
          }
        }
      },
      "required": [
        "fieldMappings"
      ]
    },
    "ticketComment": {
      "type": "object",
      "properties": {
        "fieldMappings": {
          "type": "array",
          "items": {
            "anyOf": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": ["STRING", "STRING_LIST", "LONG", "DATE"]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "dd-MM-yyyy HH:mm:ss"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
```

```
                ]
              }
            ]
          }
        }
      },
      "required": [
        "fieldMappings"
      ]
    },
    "ticketCommentAttachment": {
      "type": "object",
      "properties": {
        "fieldMappings": {
          "type": "array",
          "items": {
            "anyOf": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": ["STRING", "STRING_LIST", "LONG", "DATE"]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "dd-MM-yyyy HH:mm:ss"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        }
```

```
      },
      "required": [
        "fieldMappings"
      ]
    },
    "article": {
      "type": "object",
      "properties": {
        "fieldMappings": {
          "type": "array",
          "items": {
            "anyOf": [
              {
                "type": "object",
                "properties": {
                  "indexFieldName": {
                    "type": "string"
                  },
                  "indexFieldType": {
                    "type": "string",
                    "enum": ["STRING", "STRING_LIST", "LONG", "DATE"]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "dd-MM-yyyy HH:mm:ss"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        }
      },
      "required": [
        "fieldMappings"
      ]
    },
```

```
        "communityPostComment": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
              "items": {
                "anyOf": [
                  {
                    "type": "object",
                    "properties": {
                      "indexFieldName": {
                        "type": "string"
                      },
                      "indexFieldType": {
                        "type": "string",
                        "enum": ["STRING", "STRING_LIST", "LONG", "DATE"]
                      },
                      "dataSourceFieldName": {
                        "type": "string"
                      },
                      "dateFieldFormat": {
                        "type": "string",
                        "pattern": "dd-MM-yyyy HH:mm:ss"
                      }
                    },
                    "required": [
                      "indexFieldName",
                      "indexFieldType",
                      "dataSourceFieldName"
                    ]
                  }
                ]
              }
            }
          },
          "required": [
            "fieldMappings"
          ]
        },
        "articleComment": {
          "type": "object",
          "properties": {
            "fieldMappings": {
              "type": "array",
```

```
            "items": {
              "anyOf": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": ["STRING", "STRING_LIST", "LONG", "DATE"]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "dd-MM-yyyy HH:mm:ss"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "articleAttachment": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": {
              "anyOf": [
                {
                  "type": "object",
                  "properties": {
```

```
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
                      "enum": ["STRING", "STRING_LIST", "LONG", "DATE"]
                    },
                    "dataSourceFieldName": {
                      "type": "string"
                    },
                    "dateFieldFormat": {
                      "type": "string",
                      "pattern": "dd-MM-yyyy HH:mm:ss"
                    }
                  },
                  "required": [
                    "indexFieldName",
                    "indexFieldType",
                    "dataSourceFieldName"
                  ]
                }
              ]
            }
          }
        },
        "required": [
          "fieldMappings"
        ]
      },
      "communityTopic": {
        "type": "object",
        "properties": {
          "fieldMappings": {
            "type": "array",
            "items": {
              "anyOf": [
                {
                  "type": "object",
                  "properties": {
                    "indexFieldName": {
                      "type": "string"
                    },
                    "indexFieldType": {
                      "type": "string",
```

```
                    "enum": ["STRING", "STRING_LIST", "LONG", "DATE"]
                  },
                  "dataSourceFieldName": {
                    "type": "string"
                  },
                  "dateFieldFormat": {
                    "type": "string",
                    "pattern": "dd-MM-yyyy HH:mm:ss"
                  }
                },
                "required": [
                  "indexFieldName",
                  "indexFieldType",
                  "dataSourceFieldName"
                ]
              }
            ]
          }
        }
      },
      "required": [
        "fieldMappings"
      ]
    }
  }
},
"secretArn": {
  "type": "string",
  "minLength": 20,
  "maxLength": 2048
},
"additionalProperties": {
  "type": "object",
  "properties": {
    "organizationNameFilter": {
      "type": "array"
    },
    "sinceDate": {
      "type": "string",
      "pattern": "^[0-9]{4}-[0-9]{2}-[0-9]{2} [0-9]{2}:[0-9]{2}:[0-9]{2}$"
    },
    "inclusionPatterns": {
      "type": "array"
    },
```

```
        "exclusionPatterns": {
          "type": "array"
        },
        "isCrawTicket": {
          "type": "string"
        },
        "isCrawTicketComment": {
          "type": "string"
        },
        "isCrawTicketCommentAttachment": {
          "type": "string"
        },
        "isCrawlArticle": {
          "type": "string"
        },
        "isCrawlArticleAttachment": {
          "type": "string"
        },
        "isCrawlArticleComment": {
          "type": "string"
        },
        "isCrawlCommunityTopic": {
          "type": "string"
        },
        "isCrawlCommunityPost": {
          "type": "string"
        },
        "isCrawlCommunityPostComment": {
          "type": "string"
        }
      }
    },
    "type": {
      "type": "string",
      "pattern": "ZENDESK"
    },
    "useChangeLog": {
      "type": "string",
      "enum": ["true", "false"]
    }
  },
  "version": {
    "type": "string",
    "anyOf": [
```

```
      {
        "pattern": "1.0.0"
      }
    ]
  },
  "additionalProperties": false,
  "required": [
    "connectionConfiguration",
    "repositoryConfigurations",
    "additionalProperties",
    "useChangeLog",
    "secretArn",
    "type"
  ]
}
```

Adobe Experience Manager

Adobe Experience Manager 是一个用于创建网站或移动应用程序内容的内容管理系统。您可以使用 Amazon Kendra 连接到您的页面Adobe Experience Manager和内容资产并为其编制索引。

Amazon Kendra 支持 Adobe Experience Manager (AEM) 作为云服务作者实例以及Adobe Experience Manager本地作者和发布实例。

您可以使用Amazon Kendra 控制台或 TemplateConfigurationAPI Amazon Kendra 连接到您的Adobe Experience Manager数据源。

有关 Amazon Kendra Adobe Experience Manager 数据源连接器的疑难解答，数据来源故障排除

主题

- 支持的特征
- 先决条件
- 连接说明

支持的特征

Adobe Experience Manager 数据来源连接器支持以下功能：

- 字段映射
- 用户访问控制

- 包含/排除筛选条件
- 完整内容和增量内容同步
- OAuth 2.0 和基本身份验证
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引Adobe Experience Manager数据源之前，请在Adobe Experience Manager和 AWS 帐户中进行这些更改。

在 Adobe Experience Manager 中，请确保：

- 可访问具有管理权限的账户或管理员用户。
- 已复制您的 Adobe Experience Manager 主机 URL。

> **Note**
>
> （本地/服务器）Amazon Kendra 会检查中 AWS Secrets Manager 包含的端点信息是否与数据源配置详细信息中指定的端点信息相同。这有助于防止出现混淆代理人问题，这是一个安全问题，即用户无权执行操作，但可以将 Amazon Kendra 作为代理来访问配置的密钥和执行操作。如果以后更改端点信息，则必须创建一个新密钥来同步此信息。

- 记下您的管理员用户名和密码的基本身份验证凭证。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 可选：在 Adobe Experience Manager (AEM) 中将 OAuth 2.0 凭据配置为云服务或 AEM 本地部署。如果您使用 AEM On-Premise，则凭证包括客户端 ID、客户端密钥和私有密钥。如果您使用 AEM 即云服务，则凭证包括客户端 ID、客户端密钥、私有密钥、组织 ID、技术账户 ID 和 Adobe Identity Management System（IMS）主机。有关如何为 AEM 即云服务生成这些凭证的更多信息，请参阅 Adobe Experience Manager 文档。对于 AEM 本地部署，Adobe Granite OAuth 2.0 服务器实现（com.adobe.granite.oauth.server）支持 AEM 中的 OAuth 2.0 服务器功能。

- 在 Adobe Experience Manager 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Adobe Experience Manager 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Adobe Experience Manager 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的Adobe Experience Manager数据源，您必须提供Adobe Experience Manager数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置 Amazon Kendra，Adobe Experience Manager请参阅先决条件。

Console

要连接 Amazon Kendra 到 Adobe Experience Manager

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。

2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Adobe Experience Manager 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Adobe Experience Manager 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 源 - 选择 AEM On-Premise 或 AEM 即云服务。

      输入您的 Adobe Experience Manager 主机 URL。例如，如果您使用 AEM On-Premise，则需要包含主机名和端口：https://hostname:port。或者，如果您使用 AEM 即云服务，则可以使用作者 URL：https://author-xxxxxx-xxxxxxx.adobeaemcloud.com。

   b. SSL 证书位置 - 输入您存储在 Amazon S3 存储桶中的 SSL 证书的路径。您可以通过安全 SSL 连接使用它连接到 AEM On-Premise。

   c. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

   d. 身份验证 - 选择基本身份验证或 OAuth 2.0 身份验证。然后选择现有 AWS Secrets Manager 密钥或创建新密钥来存储您的Adobe Experience Manager凭据。如果您选择创建新密钥，则会打开一个 AWS Secrets Manager 秘密窗口。

      如果选择基本身份验证，请输入密钥的名称、Adobe Experience Manager 站点用户名和密码。用户必须具有管理员权限或是管理员用户。

如果您选择 OAuth 2.0 身份验证并使用 AEM On-Premise，请输入密钥的名称、客户端 ID、客户端密钥和私有密钥。如果您使用 AEM 即云服务，请输入密钥的名称、客户端 ID、客户端密钥、私有密钥、组织 ID、技术账户 ID 和 Adobe Identity Management System（IMS）主机。

保存并添加您的密钥。

e. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

f. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

g. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

h. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 同步范围 - 设置爬取某些内容类型、页面组件和根路径的限制，并使用正则表达式模式筛选内容。

i. 内容类型 - 选择是仅爬取页面或资产，还是同时爬取两者。

ii. （可选）其他配置 - 配置以下设置：

- 页面组件 - 页面组件的特定名称。页面组件是一个可扩展的页面组件，旨在与模板编辑器配合使用，允许使用 Adobe Experience Manager 模板编辑器组装页眉/页脚和结构组件。
- 内容片段变体 - 内容片段变体的具体名称。内容片段允许您在 Adobe Experience Manager 中设计、创建、策划和发布与页面无关的内容。它们允许您准备内容，以备在多个地点/通过多个渠道使用。

- 根路径 - 指向特定内容的根路径。
- 正则表达式模式 - 包含或排除某些文件的正则表达式模式。

b. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

c. 时区 ID - 如果您使用 AEM On-Premise，并且服务器的时区与 AEM 连接器或索引的时区不同，则可以指定与 Amazon Kendra AEM 连接器或索引相符的服务器时区。AEM On-Premise 的默认时区是 Amazon Kendra AEM 连接器或索引的时区。AEM 即云服务的默认时区是格林威治标准时间。

d. 频率同步运行计划-选择同步数据源内容和更新索引的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。要添加自定义数据来源字段，请创建要映射到的索引字段名称和字段数据类型。

b. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Adobe Experience Manager

您必须使用 TemplateConfiguration API 指定数据来源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构AEM时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。

- AEM 主机 URL - 指定 Adobe Experience Manager 主机 URL。例如，如果您使用 AEM On-Premise，则需要包含主机名和端口：https://hostname:port。或者，如果您使用 AEM 即云服务，则可以使用作者 URL：https://author-xxxxxx-xxxxxxx.adobeaemcloud.com。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅为新增、修改和删除的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 身份验证类型 - 指定要使用的身份验证类型，可以是 Basic 或 OAuth2。
- AEM 类型 - 指定您使用的 Adobe Experience Manager 类型，可以是 CLOUD 或 ON_PREMISE。
- 密钥 Amazon 资源名称（ARN）- 如果您想对 AEM 本地或云端使用基本身份验证，则需要提供一个用于存储您的用户名和密码的身份验证凭证的密钥。您提供密钥的亚马逊资源名称 (ARN)。AWS Secrets Manager 密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "aemUrl": "Adobe Experience Manager On-Premise host URL",
    "username": "user name with admin permissions",
    "password": "password with admin permissions"
}
```

如果您想对 AEM On-Premise 使用 OAuth 2.0 身份验证，则密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "aemUrl": "Adobe Experience Manager host URL",
    "clientId": "client ID",
    "clientSecret": "client secret",
    "privateKey": "private key"
}
```

如果您想对 AEM 即云服务使用 OAuth 2.0 身份验证，则密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "clientId": "client ID",
    "clientSecret": "client secret",
    "privateKey": "private key",
    "orgId": "organization ID",
    "technicalAccountId": "technical account ID",
    "imsHost": "Adobe Identity Management System (IMS) host"
}
```

- IAM 角色-指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Adobe Experience Manager 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Adobe Experience Manager 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 时区 ID-如果您使用 AEM 本地部署，并且服务器的时区与 AEM 连接器或索引的时区不同，则可以指定与 Amazon Kendra AEM 连接器或索引对齐的服务器时区。

  AEM 本地部署的默认时区是 AE Amazon Kendra M 连接器或索引的时区。AEM 即云服务的默认时区是格林威治标准时间。

  有关支持的时区 ID 的信息，请参阅 Adobe Experience Manager JSON 架构。
- 包含和排除筛选条件 - 指定是包含还是排除页面和资产。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访

问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

- 字段映射 - 选择将 Adobe Experience Manager 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅 Adobe Experience Manager 模板架构。

Alfresco

Alfresco 是一项内容管理服务，可帮助客户存储和管理其内容。您可以使用索引 Amazon Kendra 您的Alfresco文档库、Wiki 和博客。

Amazon Kendra 支持Alfresco本地部署和Alfresco云端（平台即服务）。

您可以使用Amazon Kendra 控制台或 TemplateConfigurationAPI Amazon Kendra 连接到您的Alfresco数据源。

要对 Amazon Kendra Alfresco 数据源连接器进行故障排除，请参阅。数据来源故障排除

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

Amazon Kendra Alfresco 数据来源连接器支持以下功能：

- 字段映射

- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- OAuth 2.0 和基本身份验证
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Alfresco 数据源之前，请在和中进行这些更改。Alfresco AWS 账户

在 Alfresco 中，请确保：

- 已复制您的 Alfresco 存储库 URL 和 Web 应用程序 URL。如果您只想为特定 Alfresco 站点编制索引，则还要复制该站点 ID。
- 已记下您的 Alfresco 身份验证凭证，其中包括至少具有读取权限的用户名和密码。如果要使用 OAuth 2.0 身份验证，则应将用户添加到 Alfresco 管理员组。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 可选：已在中配置 OAuth 2.0 凭据。Alfresco凭证包括客户端 ID、客户端密钥和令牌 URL。有关如何为 Alfresco On-Premises 配置客户端的更多信息，请参阅 Alfresco 文档。如果您使用 Alfresco Cloud（PaaS），则必须联系 Hyland 支持人员来进行 Alfresco OAuth 2.0 身份验证。
- 在 Alfresco 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Alfresco 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Alfresco 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。 Amazon Kendra如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 Alfresco 数据源，您必须提供 Alfresco 数据源的必要详细信息，Amazon Kendra 以便访问您的数据。如果您尚未为 Alfresco 配置 Amazon Kendra，请参阅。先决条件

Console

要连接 Amazon Kendra 到 Alfresco

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。

4. 在添加数据源页面上，选择 Alfresco 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Alfresco 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. Alfresco类型-选择使用Alfresco本地/服务器还是Alfresco云（平台即服务）。

   b. Alfresco 存储库 URL - 输入您的 Alfresco 存储库 URL。例如，如果您使用 Alfresco Cloud（PaaS），则存储库 URL 可能是 https://company.alfrescocloud.com。或者，如果您使用 Alfresco On-Premises，则存储库 URL 可能是 https://company-alfresco-instance.company-domain.suffix:port。

   c. Alfresco 用户应用程序。URL - 输入您的 Alfresco 用户界面 URL。您可以向 Alfresco 管理员获取存储库 URL。例如，用户界面 URL 可能是 https://example.com。

   d. SSL 证书位置-输入存储在存储 Amazon S3 桶中的 SSL 证书的路径。您可以通过安全 SSL 连接使用它连接到 Alfresco AEM On-Premise。

   e. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

   f. 身份验证 - 选择基本身份验证或 OAuth 2.0 身份验证。然后选择现有 Secrets Manager 密钥或创建新密钥来存储您的 Alfresco 凭证。如果您选择创建新密钥，则会打开一个 AWS Secrets Manager 秘密窗口。

      如果选择基本身份验证，请输入密钥的名称、Alfresco 用户名和密码。

      如果您选择 OAuth 2.0 身份验证，请输入密钥的名称、客户端 ID、客户端密钥和令牌 URL。

   g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

h. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

i. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

j. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 同步范围 - 设置爬取某些内容的限制，并使用正则表达式模式筛选内容。

b. i. 内容 - 选择是爬取 Alfresco 中标有“方面”的内容、特定 Alfresco 站点内的内容还是所有 Alfresco 站点内的内容。

ii. （可选）其他配置 - 设置以下设置：

- 包括评论 - 选择在 Alfresco 文档库和博客中包含评论。
- 正则表达式模式 - 包含或排除某些文件的正则表达式模式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在 “同步” 运行计划中，“频率”-选择同步数据源内容和更新索引的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 要添加自定义数据来源字段，请创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Alfresco

您必须使用 TemplateConfiguration API 指定数据来源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构ALFRESCO时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- Alfresco 站点 ID - 指定 Alfresco 站点 ID。
- Alfresco 存储库 URL - 指定 Alfresco 存储库 URL。您可以向 Alfresco 管理员获取存储库 URL。例如，如果您使用 Alfresco Cloud（PaaS），则存储库 URL 可能是 https://company.alfrescocloud.com。或者，如果您使用 Alfresco On-Premises，则存储库 URL 可能是 https://company-alfresco-instance.company-domain.suffix:port。
- Alfresco Web 应用程序 URL - 指定 Alfresco 用户界面 URL。您可以向 Alfresco 管理员获取存储库 URL。例如，用户界面 URL 可能是 https://example.com。
- 身份验证类型 - 指定要使用的身份验证类型，可以是 OAuth2 或 Basic。
- Alfresco 类型 - 指定您使用的 Alfresco 类型，可以是 PAAS（云/平台即服务）或 ON_PREM（本地）。
- 密钥 Amazon 资源名称（ARN）- 如果您想使用基本身份验证，则需要提供一个用于存储您的用户名和密码的身份验证凭证的密钥。您提供密钥的 Amazon 资源名称 (ARN)。 AWS Secrets Manager 密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "username": "user name",
    "password": "password"
}
```

如果您想使用 OAuth 2.0 身份验证，则密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "clientId": "client ID",
```

```
    "clientSecret": "client secret",
    "tokenUrl": "token URL"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Alfresco 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Alfresco 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 内容类型 - 您要爬取的内容类型，无论是 Alfresco 中标有“方面”的内容、特定 Alfresco 站点内的内容，还是所有 Alfresco 站点内的内容。您还可以列出特定的“方面”内容。
- 包含和排除筛选条件 - 指定是包含还是排除文件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

- 字段映射 - 选择将 Alfresco 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅 Alfresco 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与 Alfresco 数据源集成的更多信息，请参阅：

- 使用智能搜索Alfresco内容 Amazon Kendra

Aurora (MySQL)

Aurora 是专为云构建的关系数据库管理系统 (RDBMS)。如果您是 Aurora 用户，则可以使用索 Amazon Kendra 引您的Aurora (MySQL)数据源。 Amazon Kendra Aurora (MySQL)数据源连接器支持 Aurora MySQL 3 和 Aurora 无服务器 MySQL 8.0。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的Aurora (MySQL)数据源。

要对 Amazon Kendra Aurora (MySQL)数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 注意

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引Aurora (MySQL)数据源之前，请在Aurora (MySQL)和 AWS 帐户中进行这些更改。

在 Aurora (MySQL) 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 作为最佳实践，请提供只读 Amazon Kendra 的数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。你可以在 Amazon RDS 控制台上找到这些信息。
- 在 Aurora (MySQL) 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Aurora (MySQL) 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将Aurora (MySQL)数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的Aurora (MySQL)数据源，您必须提供Aurora (MySQL)凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置，Aurora (MySQL)请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 Aurora (MySQL)

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择Aurora (MySQL)连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的Aurora (MySQL)连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

a. 在源中，输入以下信息：

b. 主机 – 输入数据库主机 URL，例如：`http://instance URL.region.rds.amazonaws.com`。

c. 端口 –输入数据库端口，例如 5432。

d. 实例 - 输入数据库实例。

e. 在身份验证中 - 请输入以下信息：

- AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的Aurora (MySQL)身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

  A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

     I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Aurora (MySQL)-” 会自动添加到您的密钥名称中。

     II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

  B. 选择保存。

f. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

g. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

h. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB SQL 查询必须小于 32KB 且不包含任何分号 (;)。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。
- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Aurora (MySQL)

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 mySql。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。Aurora (MySQL)密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用Aurora (MySQL)连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Aurora (MySQL) S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 `VpcConfiguration`，以便调用 `CreateDataSource`。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 抓取文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Aurora (MySQL) 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称`_document_body`。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅Aurora (MySQL) 模板架构。

注意

- Amazon Kendra 检查已更新的内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。

- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 作为最佳实践，请提供只读 Amazon Kendra 的数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

Aurora (PostgreSQL)

Aurora 是专为云构建的关系数据库管理系统 (RDBMS)。如果您是 Aurora 用户，则可以使用索 Amazon Kendra 引您的Aurora (PostgreSQL)数据源。 Amazon Kendra Aurora (PostgreSQL)数据源连接器支持 Aurora PostgreSQL 1。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的Aurora (PostgreSQL)数据源。

要对 Amazon Kendra Aurora (PostgreSQL)数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引Aurora (PostgreSQL)数据源之前，请在Aurora (PostgreSQL)和 AWS 帐户中进行这些更改。

在 Aurora (PostgreSQL) 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 最佳做法是提供 Amazon Kendra 只读数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。
- 在 Aurora (PostgreSQL) 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Aurora (PostgreSQL) 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将Aurora (PostgreSQL)数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的Aurora (PostgreSQL)数据源，您必须提供Aurora (PostgreSQL)凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置，Aurora (PostgreSQL)请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 Aurora (PostgreSQL)

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择Aurora (PostgreSQL)连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的Aurora (PostgreSQL)连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 在源中，输入以下信息：
   b. 主机 – 输入数据库主机 URL，例如：`http://instance URL.region.rds.amazonaws.com`。
   c. 端口 –输入数据库端口，例如 `5432`。
   d. 实例 – 输入数据库实例，例如 `postgres`。
   e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

f. 在身份验证中 - 请输入以下信息：

- AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的Aurora (PostgreSQL)身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

  A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

     I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Aurora (PostgreSQL)-” 会自动添加到您的密钥名称中。

     II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

  B. 选择保存。

g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

h. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB SQL 查询必须小于 32KB 且不包含任何分号 (;)。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。

- 组列 - 输入包含允许访问内容的群组的列的名称。
- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Aurora (PostgreSQL)

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 postgresql。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。Aurora (PostgreSQL)密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用Aurora (PostgreSQL)连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Aurora (PostgreSQL) S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。

- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Aurora (PostgreSQL) 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅Aurora (PostgreSQL) 模板架构。

注意

- 在 Amazon Kendra 检查更新内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。
- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 最佳做法是提供 Amazon Kendra 只读数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

Amazon FSx（视窗）

Amazon FSx (Windows) 是一个完全托管的、基于云的文件服务器系统，提供共享存储功能。如果你是 Amazon FSx (Windows) 用户，则可以使用 Amazon Kendra 索引你的 Amazon FSx (Windows) 数据源。

> **Note**
>
> Amazon Kendra 现在支持升级版 Amazon FSx (Windows) 连接器。

控制台已自动为您升级。您在控制台上创建的任何新连接器都将使用升级后的架构。如果您使用 API，则现在必须使用TemplateConfiguration对象而不是FSxConfiguration对象来配置您的连接器。
使用较旧的控制台和 API 架构配置的连接器将继续按配置运行。但是，您将无法对其进行编辑或更新。如果要编辑或更新连接器配置，则必须创建新的连接器。
我们建议将您的连接器工作流程迁移到升级版本。对使用旧架构配置的连接器的支持计划于 2024 年 6 月结束。

您可以使用Amazon Kendra 控制台或 TemplateConfigurationAPI Amazon Kendra 连接到您的 Amazon FSx (Windows) 数据源。

要对您的 Amazon Kendra Amazon FSx (Windows) 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

Amazon Kendra Amazon FSx (Windows) 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 用户身份爬行
- 包含和排除过滤器
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Amazon FSx (Windows) 数据源之前，请先检查您的 Amazon FSx (Windows) 和的详细信息 AWS 账户。

对于 Amazon FSx (Windows)，请确保你有：

- 使用读取和装载权限进行设置 Amazon FSx (Windows)。
- 已记下您的文件系统 ID。您可以在 Amazon FSx (Windows) 控制台的 “文件系统” 仪表板上找到您的文件系统 ID。
- 使用您的 Amazon FSx (Windows) 文件系统所在 Amazon VPC 位置配置虚拟私有云。
- 记下了你的 Amazon FSx (Windows) Active Directory 用户帐户身份验证凭证。这包括你的 Active Directory 用户名以及你的 DNS 域名（例如，user@corp.example.com）和密码。

> **Note**
>
> 仅使用连接器运行所需的必要凭据。请勿使用诸如域管理员之类的特权凭据。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 已选中每个文档在 Amazon FSx (Windows) 中以及计划用于同一索引的其他数据源中都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将你的 Amazon FSx (Windows) 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，则记下该密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Amazon FSx (Windows) 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 Amazon FSx (Windows) 数据源，必须提供您的 Amazon FSx (Windows) 数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为 Amazon FSx (Windows) 进行配置 Amazon Kendra，请参阅先决条件。

Console

Amazon Kendra 连接到你的 Amazon FSx (Windows) 文件系统

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

   > **Note**
   >
   > 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Amazon FSx (Windows) 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Amazon FSx (Windows) 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

a. Amazon FSx (Windows) 文件系统 ID-从下拉列表中选择从 Amazon FSx (Windows) 获取的现有文件系统 ID。或者，创建一个 Amazon FSx (Windows) 文件系统。您可以在 Amazon FSx (Windows) 控制台的 “文件系统” 仪表板上找到您的文件系统 ID。

b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

c. 身份验证-选择现有 AWS Secrets Manager 密钥，或创建新密钥来存储您的文件系统凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

提供一个用于存储您的用户名和密码的身份验证凭据的密钥。用户名必须包含您的 DNS 域名。例如，user@corp.example.com。

保存并添加您的密钥。

d. 虚拟私有云 (VPC) — 你必须选择 Amazon VPC 你的 (Windows) Amazon FSx 所在的位置。您包括 VPC 子网和安全组。请参阅配置 Amazon VPC。

e. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

f. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 同步作用域、正则表达式模式-添加正则表达式模式以包含或排除某些文件。

b. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。

- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

c. 同步运行计划-对于频率，选择同步数据源内容和更新索引的频率。

d. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从 Amazon Kendra 生成的文件默认字段中选择要映射到索引的字段。要添加自定义数据来源字段，请创建要映射到的索引字段名称和字段数据类型。

b. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

Amazon Kendra 连接到你的 Amazon FSx (Windows) 文件系统

您必须使用 TemplateConfiguration API 指定数据来源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构FSX时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 文件系统 ID- Amazon FSx (Windows) 文件系统的标识符。您可以在 Amazon FSx (Windows) 控制台的 “文件系统” 仪表板上找到您的文件系统 ID。
- 文件系统类型 - 将文件系统的类型指定为 WINDOWS。
- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。

> **Note**
>
> 你必须选择你的 Amazon FSx (Windows) 所在的位置。 Amazon VPC 您包括 VPC 子网和安全组。

- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：

  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。
- 亚马逊秘密资源名称 (ARN)-提供包含您 Amazon FSx (Windows) 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "username": "user@corp.example.com",
    "password": "password"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Amazon FSx (Windows) 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Amazon FSx (Windows) 数据源的IAM 角色。

您还可以添加以下可选功能：

- 包含和排除筛选条件 - 指定是包含还是排除文件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 访问控制列表 (ACL)-如果您有 ACL 并希望将其用于访问控制，则指定是否要搜索文档的 ACL 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

> **Note**
>
> 要对用户测试用户上下文筛选，在发出查询时，必须将 DNS 域名作为用户名的一部分包含在内。您必须拥有 Active Directory 域的管理权限。您也可以根据组名称测试用户上下文筛选。

- 字段映射-选择将您的 Amazon FSx (Windows) 数据源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅 Amazon FSx (Windows) 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与你的 Amazon FSx (Windows) 数据源集成的更多信息，请参阅：

- 使用适用于 Amazon FSx (Windows) 的 Amazon Kendra 连接器，在 Windows 文件系统上安全地搜索非结构化数据。Windows File Server

Amazon FSx（NetApp ONTAP）

Amazon FSx (NetApp ONTAP) 是一个完全托管的、基于云的文件服务器系统，提供共享存储功能。如果您是 Amazon FSx (NetApp ONTAP) 用户，则可以使用 Amazon Kendra 索引您的 Amazon FSx (NetApp ONTAP) 数据源。

您可以使用Amazon Kendra 控制台或 TemplateConfigurationAPI Amazon Kendra 连接到您的 Amazon FSx (NetApp ONTAP) 数据源。

要对您的 Amazon Kendra Amazon FSx (NetApp ONTAP) 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

Amazon Kendra Amazon FSx (NetApp ONTAP) 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含和排除过滤器
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Amazon FSx (NetApp ONTAP) 数据源之前，请先检查您的 Amazon FSx (NetApp ONTAP) 和的详细信息。 AWS 账户

对于 Amazon FSx (NetApp ONTAP)，请确保您具有：

- 使用读取和挂载权限进行设置 Amazon FSx (NetApp ONTAP)。
- 已记下您的文件系统 ID。您可以在 Amazon FSx (NetApp ONTAP) 控制台的文件系统控制面板上找到您的文件系统 ID。
- 已记下用于文件系统的存储虚拟机 (SVM) ID。您可以前往 Amazon FSx (NetApp ONTAP) 控制台中的文件系统仪表板，选择您的文件系统 ID，然后选择存储虚拟机，找到您的 SVM ID。
- 使用您的 Amazon FSx (NetApp ONTAP) 文件系统所在 Amazon VPC 位置配置虚拟私有云。
- 记下您的 Amazon FSx (NetApp ONTAP) Active Directory 用户帐户身份验证凭据。这包括你的 Active Directory 用户名以及你的 DNS 域名（例如 user@corp.example.com）和密码。如果您的 (NetApp ONTAP) 文件系统使用网络文件系统 Amazon FSx (NFS) 协议，则身份验证凭据包括左 ID、右 ID 和预共享密钥。

> **Note**
>
> 仅使用连接器运行所需的必要凭据。请勿使用诸如域管理员之类的特权凭据。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 已选中每个文档在 Amazon FSx (NetApp ONTAP) 中以及计划用于同一索引的其他数据源中都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Amazon FSx (NetApp ONTAP) 身份验证凭据存储在 AWS Secrets Manager 密钥中，如果使用 API，则记下该密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Amazon FSx (NetApp ONTAP) 数据源连接到 Amazon Kendra时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 Amazon FSx （ NetApp ONTAP ）数据源，您必须提供您的 Amazon FSx （ NetApp ONTAP ）数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未配置 Amazon FSx (NetApp ONTAP) Amazon Kendra，请参阅先决条件。

Console

Amazon Kendra 连接到您的 Amazon FSx (NetApp ONTAP) 文件系统

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Amazon FSx (NetApp ONTAP) 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Amazon FSx (NetApp ONTAP) 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 来源-提供您的文件系统信息。
      - 文件系统协议-选择您的 Amazon FSx (NetApp ONTAP) 文件系统的协议。你可以选择通用互联网文件系统 (CIFS) 协议，也可以选择适用于 Linux 的网络文件系统 (NFS) 协议。

- Amazon FSx (NetApp ONTAP) 文件系统 ID-从下拉列表中选择从中获取的现有文件系统 ID Amazon FSx (NetApp ONTAP)。或者，创建 Amazon FSx (NetApp ONTAP) 文件系统。您可以在 Amazon FSx (NetApp ONTAP) 控制台的文件系统控制面板上找到您的文件系统 ID。
- SVM IDAmazon FSx （ NetApp ONTAP仅限 NetApp ONTAP ）-提供您 Amazon FSx 的（ ONTAP ）的存储虚拟机 (SVM) ID （ ONTAP ）。NetApp NetApp ONTAP您可以前往 Amazon FSx (NetApp ONTAP) 控制台中的文件系统仪表板，选择您的文件系统 ID，然后选择存储虚拟机，找到您的 SVM ID。

b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

c. 身份验证-选择现有 AWS Secrets Manager 密钥，或创建新密钥来存储您的文件系统凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

   提供一个用于存储您的用户名和密码的身份验证凭据的密钥。用户名必须包含您的 DNS 域名。例如，user@corp.example.com。

   如果您的 Amazon FSx (NetApp ONTAP) 文件系统使用 NFS 协议，请提供一个密钥来存储您的身份验证凭据，包括左 ID、右 ID 和预共享密钥。

   保存并添加您的密钥。

d. 虚拟私有云 (VPC)-您必须选择您的 NetApp (ONTA Amazon VPC P) Amazon FSx 所在的位置。您包括 VPC 子网和安全组。请参阅配置 Amazon VPC。

e. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

f. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 同步范围、正则表达式模式-添加正则表达式模式以包含或排除某些文件。

b. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

c. 同步运行计划-对于频率，选择同步数据源内容和更新索引的频率。

d. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从 Amazon Kendra 生成的文件默认字段中选择要映射到索引的字段。要添加自定义数据来源字段，请创建要映射到的索引字段名称和字段数据类型。

b. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

Amazon Kendra 连接到您的 Amazon FSx (NetApp ONTAP) 文件系统

您必须使用 TemplateConfiguration API 指定数据来源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构FSXONTAP时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 文件系统 ID- Amazon FSx (NetApp ONTAP) 文件系统的标识符。您可以在 Amazon FSx (NetApp ONTAP) 控制台的文件系统控制面板上找到您的文件系统 ID。
- SVM ID-用于文件系统的存储虚拟机 (SVM) ID。您可以前往 Amazon FSx (NetApp ONTAP) 控制台中的文件系统仪表板，选择您的文件系统 ID，然后选择存储虚拟机，找到您的 SVM ID。
- 协议类型-指定是使用通用互联网文件系统 (CIFS) 协议，还是用于 Linux 的网络文件系统 (NFS) 协议。
- 文件系统类型-将文件系统的类型指定为任一FSXONTAP类型。
- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。

> **Note**
>
> 您必须选择您的 Amazon VPC Amazon FSx (NetApp ONTAP) 所在的位置。您包括 VPC 子网和安全组。

- 亚马逊秘密资源名称 (ARN)-提供包含您 Amazon FSx (ONTAP) 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。NetApp 密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "username": "user@corp.example.com",
    "password": "password"
}
```

如果您的 Amazon FSx (NetApp ONTAP) 文件系统使用 NFS 协议，则密钥将存储在 JSON 结构中，其中包含以下密钥：

```
{
    "leftId": "left ID",
    "rightId": "right ID",
    "preSharedKey": "pre-shared key"
}
```

- IAM 角色-指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Amazon FSx (NetApp ONTAP) 连接器所需的公共 API 的权限，以及。Amazon Kendra有关更多信息，请参阅 Amazon FSx (NetApp ONTAP) 数据源的IAM 角色。

您还可以添加以下可选功能：

- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - `FORCED_FULL_CRAWL`对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - `FULL_CRAWL`每次数据源与索引同步时，仅为新增、修改和删除的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 包含和排除筛选条件 - 指定是包含还是排除文件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 访问控制列表 (ACL)-如果您有 ACL 并希望将其用于访问控制，则指定是否要搜索文档的 ACL 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

> **Note**
>
> 要对用户测试用户上下文筛选，在发出查询时，必须将 DNS 域名作为用户名的一部分包含在内。您必须拥有 Active Directory 域的管理权限。您也可以根据组名称测试用户上下文筛选。

- 字段映射-选择将您的 Amazon FSx (NetApp ONTAP) 数据源字段映射到索引字段。 Amazon Kendra 有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅 Amazon FSx (NetApp ONTAP) 模板架构。

Amazon RDS/Aurora

您可以使用数据库数据来源为存储在数据库中的文档编制索引。提供数据库的连接信息后， Amazon Kendra 连接和索引文档。

Amazon Kendra 支持以下数据库：

- Amazon Aurora MySQL

- Amazon Aurora PostgreSQL
- Amazon RDS 适用于 MySQL
- Amazon RDS 适用于 PostgreSQL

> **Note**
>
> 不支持无服务器 Aurora 数据库。

> **Important**
>
> 此 Amazon RDS/Aurora 连接器计划于 2023 年底之前弃用。
> Amazon Kendra 现在支持新的数据库数据源连接器。为了改善体验，我们建议您选择以下适用于您的用例的新连接器：
>
> - Aurora (MySQL)
> - Aurora (PostgreSQL)
> - Amazon RDS (MySQL)
> - Amazon RDS （微软 SQL Server）
> - Amazon RDS （甲骨文）
> - Amazon RDS (PostgreSQL)
> - IBM DB2
> - Microsoft SQL Server
> - MySQL
> - Oracle Database
> - PostgreSQL

您可以使用Amazon Kendra 控制台和 DatabaseConfigurationAPI Amazon Kendra 连接到数据库数据源。

要对 Amazon Kendra 数据库数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征

- 先决条件
- 连接说明

支持的特征

Amazon Kendra 数据库数据源连接器支持以下功能：

- 字段映射
- 用户上下文筛选
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引数据库数据源之前，请先在数据库和 AWS 帐户中进行这些更改。

在数据库中，请确保：

- 记下您的数据库用户名和密码基本身份验证凭证。
- 已复制主机名、端口号、主机地址、数据库名称以及包含文档数据的数据表的名称。对于 PostgreSQL，数据表必须是公共表或公共架构。

> **Note**
>
> 主机和端口告诉您在 Internet 上的 Amazon Kendra 何处可以找到数据库服务器。数据库名和表名告诉你 Amazon Kendra 在数据库服务器上哪里可以找到文档数据。

- 已复制数据表中包含文档数据的列的名称。必须包括文档 ID、文档正文、用于检测文档是否已更改的列（例如，上次更新的列）以及映射到自定义索引字段的可选数据表列。您也可以将任何 Amazon Kendra 保留字段名称映射到表列。
- 已复制数据库引擎类型信息，例如您使用的是 My Amazon RDS SQL 还是其他类型。
- 在数据库以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。

- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的数据库身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将数据库数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到数据库数据源，必须提供数据库数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为配置数据库 Amazon Kendra，请参见先决条件。

Console

Amazon Kendra 连接到数据库

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。

4. 在添加数据源页面上，选择数据库连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的数据库连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 端点 - DNS 主机名、IPv4 地址或 IPv6 地址。

   b. 端口 - 端口号。

   c. 数据库 - 数据库名称。

   d. 表名 - 表名。

   e. 对于身份验证类型，请选择现有或新建，以便存储您的数据库身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

      - 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

        A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-database-” 会自动添加到您的密钥名称中。

        B. 对于用户名和密码 - 输入您从数据库账户中复制的身份验证凭证值。

        C. 选择保存身份验证。

   f. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

> **Note**
>
> 您必须使用私有子网。如果您的 RDS 实例位于您的 VPC 的公有子网中，则可以创建一个私有子网，该子网可以出站访问公有子网中的 NAT 网关。VPC 配置中提供的子网必须位于美国西部（俄勒冈州）、美国东部（弗吉尼亚北部）和欧洲（爱尔兰）。

g. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

h. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 根据您的使用案例，选择 Aurora MySQL、MySQL、Aurora PostgreSQL 和 PostgreSQL。

   b. 用双引号将 SQL 标识符括起来 - 选择使用双引号将 SQL 标识符括起来。例如，“columnName”。

   c. ACL 列和变更检测列-配置 Amazon Kendra 用于变更检测的列（例如，上次更新的列）和您的访问控制列表。

   d. 在 “同步运行计划” 中，“频率”-选择与数据源同步的频率。 Amazon Kendra

   e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

   a. Amazon Kendra 默认字段映射-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。必须为 `document_id` 和 `document_body` 添加数据库列值

   b. 自定义字段映射 - 要添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

   c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

Amazon Kendra 连接到数据库

您必须指定以下 DatabaseConfigurationAPI：

- ColumnConfiguration— 有关索引应从数据库中从何处获取文档信息的信息。有关更多详细信息，请参阅ColumnConfiguration。必须指定 `DocumentDataColumnName`（文档正文或朱文本）、`DocumentIdColumnName` 和 `ChangeDetectingColumn`（例如，上次更新的列）字

段。映射到 DocumentIdColumnName 字段的列必须是整数列。以下示例显示数据库数据来源的简单列配置：

```
"ColumnConfiguration": {
    "ChangeDetectingColumns": [
        "LastUpdateDate",
        "LastUpdateTime"
    ],
    "DocumentDataColumnName": "TextColumn",
    "DocumentIdColumnName": "IdentifierColumn",
    "DocoumentTitleColumnName": "TitleColumn",
    "FieldMappings": [
        {
            "DataSourceFieldName": "AbstractColumn",
            "IndexFieldName": "Abstract"
        }
    ]
}
```

- ConnectionConfiguration— 连接数据库所需的配置信息。有关更多详细信息，请参阅ConnectionConfiguration。
- DatabaseEngineType— 运行数据库的数据库引擎的类型。的DatabaseHost字段ConnectionConfiguration必须是数据库的 Amazon Relational Database Service (Amazon RDS) 实例终端节点。不要使用集群端点。
- 亚马逊秘密资源名称 (ARN)-提供包含数据库账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "username": "user name",
    "password": "password"
}
```

以下示例显示数据库配置，包括密钥 ARN。

```
"DatabaseConfiguration": {
"ConnectionConfiguration": {
"DatabaseHost": "host.subdomain.domain.tld",
        "DatabaseName": "DocumentDatabase",
        "DatabasePort": 3306,
        "SecretArn": "arn:aws:secretmanager:region:account ID:secret/secret name",
```

```
        "TableName": "DocumentTable"
    }
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用数据库连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅数据库数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 在数据来源配置中指定 `VpcConfiguration`。请参阅配置 Amazon Kendra 以使用 VPC。

> **Note**
>
> 您只能使用私有子网。如果您的 RDS 实例位于您的 VPC 的公有子网中，则可以创建一个私有子网，该子网可以出站访问公有子网中的 NAT 网关。VPC 配置中提供的子网必须位于美国西部（俄勒冈州）、美国东部（弗吉尼亚北部）和欧洲（爱尔兰）。

- 字段映射 - 选择将数据库数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称`_document_body`。其他所有字段均为可选字段。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 抓取文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

Amazon RDS（微软 SQL Server）

SQL Server 是由微软开发的数据库管理系统。 Amazon RDS for SQL Server 可以轻松地在云中设置、操作和扩展 SQL Server 部署。如果你是 Amazon RDS（微软 SQL Server）用户，则可以使用 Amazon Kendra 索引你的 Amazon RDS（微软 SQL Server）数据源。 Amazon Kendra JDBC 数据源连接器支持微软 SQL Server 2019。

你可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到你的 Amazon RDS（微软 SQL Server）数据源。

要对你的 Amazon Kendra Amazon RDS（微软 SQL Server）数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引你的 Amazon RDS（微软 SQL Server）数据源之前，请在你的 Amazon RDS（微软 SQL Server）和 AWS 账户中进行这些更改。

在 Amazon RDS（微软 SQL Server）中，确保你有：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 最佳做法是提供 Amazon Kendra 只读数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。
- 已检查每个文档在 Amazon RDS （ Microsoft SQL Server ）中以及计划用于同一索引的其他数据源中都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将你的 Amazon RDS （ Microsoft SQL Server ）身份验证凭据存储在 AWS Secrets Manager 密钥中，如果使用 API，则记下该密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Amazon RDS （ Microsoft SQL Server ）数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到你的 Amazon RDS （微软 SQL Server ）数据源，你必须提供你的 Amazon RDS （微软 SQL Server ）凭据的详细信息，这样 Amazon Kendra 才能访问你的数据。如果你尚未配置 Amazon RDS （微软 SQL Server ）， Amazon Kendra 请参阅先决条件。

Console

要连接 Amazon Kendra 到 Amazon RDS （微软 SQL Server ）

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Amazon RDS (Microsoft SQL Server) 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Amazon RDS （微软 SQL Server ）连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 在源中，输入以下信息：
   b. 主机 - 输入数据库主机名。
   c. 端口 - 输入数据库端口。
   d. 实例 - 输入数据库实例。
   e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

f. 在身份验证中 - 请输入以下信息：

- AWS Secrets Manager s@@ ec ret — 选择现有密钥或创建新 Secrets Manager 密钥来存储你的 Amazon RDS （ Microsoft SQL Server ）身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

  A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

     I. 密钥名称 - 密钥的名称。前缀 'AmazonKendra-Amazon RDS （ Microsoft SQL Server ）-' 会自动添加到你的密钥名称中。

     II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

  B. 选择保存。

g. 虚拟私有云（ VPC ）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

h. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。

> **Note**
>
> 如果表名中包含特殊字符（非字母数字），则必须在表名周围使用方括号。例如，*# [my-database-table] ###**

- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。
- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Amazon RDS （ 微软 SQL Server ）

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 sqlserver。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。

  > **Note**
  >
  > 如果表名中包含特殊字符（非字母数字），则必须在表名周围使用方括号。例如，*#[my-database-table] ###**

- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供密钥的亚马逊资源名称 (ARN)，该 Secrets Manager 密钥包含您在您的（ Amazon RDS 微软 SQL Server ）账户中创建的身份验证证书。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM 角色 —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Amazon RDS （Microsoft SQL Server）连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Amazon RDS （微软 SQL Server）数据源的IAM角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 `VpcConfiguration`，以便调用 `CreateDataSource`。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射-选择将你的（ Amazon RDS Microsoft SQL Server）数据源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称`_document_body`。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅Amazon RDS （微软 SQL Server）模板架构。

注意

- 在 Amazon Kendra 检查更新内容时，不会跟踪已删除的数据库行。

- 在数据库的一行中，字段名和值的大小不能超过 400KB。
- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 最佳做法是提供 Amazon Kendra 只读数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

Amazon RDS (MySQL)

Amazon RDS （Amazon Relational Database Service）是一项网络服务，可以更轻松地在 AWS 云中设置、操作和扩展关系数据库。如果您是 Amazon RDS 用户，则可以使用索 Amazon Kendra 引您的 Amazon RDS (MySQL)数据源。 Amazon Kendra 数据源连接器支持 Amazon RDS MySql 5.6、5.7 和 8.0。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的Amazon RDS (MySQL)数据源。

要对 Amazon Kendra Amazon RDS (MySQL)数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引Amazon RDS (MySQL)数据源之前，请在Amazon RDS (MySQL)和 AWS 帐户中进行这些更改。

在 Amazon RDS (MySQL) 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 作为最佳实践，请提供只读 Amazon Kendra 的数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。你可以在 Amazon RDS 控制台上找到这些信息。
- 在 Amazon RDS (MySQL) 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Amazon RDS (MySQL) 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将Amazon RDS (MySQL)数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的Amazon RDS (MySQL)数据源，您必须提供Amazon RDS (MySQL)凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置，Amazon RDS (MySQL)请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 Amazon RDS (MySQL)

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择Amazon RDS (MySQL)连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的Amazon RDS (MySQL)连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 在源中，输入以下信息：
   b. 主机 – 输入数据库主机 URL，例如：`http://instance URL.region.rds.amazonaws.com`。

c. 端口 –输入数据库端口，例如 5432。

d. 实例 – 输入数据库实例，例如 postgres。

e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

f. 在身份验证中 - 请输入以下信息：

- AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的Amazon RDS (MySQL)身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

  A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

     I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Amazon RDS (MySQL)-” 会自动添加到您的密钥名称中。

     II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

  B. 选择保存。

g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

h. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB SQL 查询必须小于 32KB 且不包含任何分号 (;)。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。
- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Amazon RDS (MySQL)

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 mySql。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。Amazon RDS (MySQL)密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用Amazon RDS (MySQL)连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Amazon RDS (MySQL) S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（ VPC ）- 指定 `VpcConfiguration`，以便调用 `CreateDataSource`。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 字段映射 - 选择将 Amazon RDS (MySQL) 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称`_document_body`。其他所有字段均为可选字段。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 抓取文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

有关要配置的其他重要 JSON 键的列表，请参阅Amazon RDS (MySQL) 模板架构。

注意

- Amazon Kendra 检查已更新的内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。
- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 作为最佳实践，请提供只读 Amazon Kendra 的数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（ PII ）的表。

Amazon RDS (Oracle)

Amazon RDS （ Amazon Relational Database Service ）是一项网络服务，可以更轻松地在 AWS 云中设置、操作和扩展关系数据库。如果您是Amazon RDS (Oracle)用户，则可以使用索 Amazon Kendra 引您的Amazon RDS (Oracle)数据源。 Amazon Kendra Amazon RDS (Oracle)数据源连接器支持 Amazon RDS Oracle 数据库 21c、甲骨文数据库 19c、甲骨文数据库 12c。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的Amazon RDS (Oracle)数据源。

要对 Amazon Kendra Amazon RDS (Oracle)数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引Amazon RDS (Oracle)数据源之前，请在Amazon RDS (Oracle)和 AWS 帐户中进行这些更改。

在 Amazon RDS (Oracle) 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 最佳做法是提供 Amazon Kendra 只读数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。
- 在 Amazon RDS (Oracle) 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Amazon RDS (Oracle) 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将Amazon RDS (Oracle)数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的Amazon RDS (Oracle)数据源，您必须提供Amazon RDS (Oracle)凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置，Amazon RDS (Oracle)请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 Amazon RDS (Oracle)

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择Amazon RDS (Oracle)连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的Amazon RDS (Oracle)连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 在源中，输入以下信息：
   b. 主机 - 输入数据库主机名。
   c. 端口 - 输入数据库端口。
   d. 实例 - 输入数据库实例。
   e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

f. 在身份验证中 - 请输入以下信息：

- AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的Amazon RDS (Oracle)身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

  A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

     I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Amazon RDS (Oracle)-” 会自动添加到您的密钥名称中。

     II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

  B. 选择保存。

g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

h. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。

- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Amazon RDS (Oracle)

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 oracle。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。Amazon RDS (Oracle)密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用Amazon RDS (Oracle)连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Amazon RDS (Oracle) S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。

- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Amazon RDS (Oracle) 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅Amazon RDS (甲骨文) 模板架构。

注意

- 在 Amazon Kendra 检查更新内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。
- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 最佳做法是提供 Amazon Kendra 只读数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

Amazon RDS (PostgreSQL)

Amazon RDS 是一项 Web 服务，可以更轻松地在 AWS 云中设置、操作和扩展关系数据库。如果您是 Amazon RDS 用户，则可以使用索 Amazon Kendra 引您的Amazon RDS (PostgreSQL)数据源。Amazon Kendra Amazon RDS (PostgreSQL)数据源连接器支持 PostgreSQL 9.6。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的Amazon RDS (PostgreSQL)数据源。

要对 Amazon Kendra Amazon RDS (PostgreSQL)数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引Amazon RDS (PostgreSQL)数据源之前，请在Amazon RDS (PostgreSQL)和 AWS 帐户中进行这些更改。

在 Amazon RDS (PostgreSQL) 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 作为最佳实践，请提供只读 Amazon Kendra 的数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。你可以在 Amazon RDS 控制台上找到这些信息。
- 在 Amazon RDS (PostgreSQL) 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Amazon RDS (PostgreSQL) 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将Amazon RDS (PostgreSQL)数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的Amazon RDS (PostgreSQL)数据源，您必须提供Amazon RDS (PostgreSQL)凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置，Amazon RDS (PostgreSQL)请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 Amazon RDS (PostgreSQL)

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择Amazon RDS (PostgreSQL)连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的Amazon RDS (PostgreSQL)连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 在源中，输入以下信息：

   b. 主机 – 输入数据库主机 URL，例如：`http://instance URL.region.rds.amazonaws.com`。

   c. 端口 –输入数据库端口，例如 `5432`。

   d. 实例 – 输入数据库实例，例如 `postgres`。

   e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

   f. 在身份验证中 - 请输入以下信息：

      - AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的Amazon RDS (PostgreSQL)身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

        A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

           I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Amazon RDS (PostgreSQL)-” 会自动添加到您的密钥名称中。

           II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

        B. 选择保存。

   g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

   h. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB SQL 查询必须小于 32KB 且不包含任何分号 (;)。 Amazon Kendra 将抓取与您的查询相匹配的所有数据库内容。
- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。
- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。

- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Amazon RDS (PostgreSQL)

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 `postgresql`。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - `FORCED_FULL_CRAWL`对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - `FULL_CRAWL`每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。Amazon RDS (PostgreSQL)密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用Amazon RDS (PostgreSQL)连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Amazon RDS (PostgreSQL) S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Amazon RDS (PostgreSQL) 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅Amazon RDS (PostgreSQL) 模板架构。

注意

- 在 Amazon Kendra 检查更新内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。
- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 作为最佳实践，请提供只读 Amazon Kendra 的数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

Amazon S3

Amazon S3 是一种对象存储服务，可将数据作为对象存储在存储桶中。您可以使用 Amazon Kendra 为 Amazon S3 存储桶中的文档存储库编制索引。

> **Warning**
>
> Amazon Kendra 不使用向 Amazon Kendra 委托人授予与 S3 存储桶交互的权限的存储桶策略。相反，它使用 IAM 角色。请确保该成员 Amazon Kendra 未作为可信成员包含在存储桶策略中，以避免在意外向任意委托人授予权限时出现任何数据安全问题。但是，您可以添加存储桶策略，以便在不同的账户中使用 Amazon S3 存储桶。有关更多信息，请参阅跨账户使用 Amazon S3 的策略（在“S3 IAM 角色”选项卡的数据来源的IAM 角色下）。有关 S3 数据源的 IAM 角色的信息，请参阅IAM 角色。

> **Note**
>
> Amazon Kendra 现在支持升级后的 Amazon S3 连接器。
> 控制台已自动为您升级。您在控制台中创建的任何新连接器都将使用升级后的架构。如果您使用 API，则现在必须使用TemplateConfiguration对象而不是S3DataSourceConfiguration对象来配置您的连接器。
> 使用较旧的控制台和 API 架构配置的连接器将继续按配置运行。但是，您将无法对其进行编辑或更新。如果要编辑或更新连接器配置，则必须创建新的连接器。
> 我们建议将您的连接器工作流程迁移到升级版本。对使用旧架构配置的连接器的支持计划于 2024 年 6 月结束。

您可以使用Amazon Kendra 控制台或 TemplateConfigurationAPI 连接到您的 Amazon S3 数据源。

> **Note**
>
> 要为您的 Amazon S3 数据源生成同步状态报告，请参阅数据源疑难解答。

要 Amazon Kendra 对 S3 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户访问控制
- 包含/排除筛选条件

- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 S3 数据源之前，请在您的 S3 和 AWS 账户中进行这些更改。

在 S3 中，请确保：

- 已复制 Amazon S3 存储桶的名称。

> **Note**
>
> 您的存储桶必须与您的 Amazon Kendra 索引位于同一区域，并且您的索引必须有权访问包含您的文档的存储桶。

- 在 S3 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在您的 AWS 账户中，请确保您有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

如果您没有现有 IAM 角色，则可以在将 S3 数据源连接到时使用控制台创建新 IAM 角色 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色的 ARN 和索引 ID。

连接说明

要连接 Amazon Kendra 到 S3 数据源，您必须提供 S3 数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为 S3 配置 Amazon Kendra，请参阅先决条件。

Console

要连接 Amazon Kendra 到 Amazon S3

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。

2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 S3 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 S3 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，输入以下可选信息：

   a. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

   > **Note**
   >
   > IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

   b. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

   c. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 对于数据源位置-指定存储数据的 Amazon S3 存储桶的路径。选择浏览 S3 以选择您的 S3 存储桶。

   b. 对于最大文件大小-指定一个以 MB 为单位的限制，以仅搜索低于此限制的文件。允许的最大文件 Amazon Kendra 大小为 50 MB。

c. 对于（可选）元数据文件前缀文件夹位置-指定存储字段/属性和其他文档元数据的文件夹的路径。选择浏览 S3 以找到元数据文件夹。

d. 对于（可选）访问控制列表配置文件的位置-指定文件路径，该文件包含用户的 JSON 结构及其对文档的访问权限。选择浏览 S3 以找到 ACL 文件。

e. （可选）选择解密密钥 - 选择使用解密密钥。您可以选择使用现有 AWS KMS 密钥。

f. 对于（可选）其他配置-添加模式以包含或排除某些文件。所有路径都是相对于数据来源位置 S3 存储桶的路径。

g. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

h. 在 “同步” 运行计划中，“频率”-选择同步数据源内容和更新索引的频率。

i. 选择下一步。

8. 在设置字段映射页面上，输入以下可选信息：

a. 默认字段映射-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 选择添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Amazon S3

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构S3时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- BucketName— 包含文档的存储桶的名称。

- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 S3 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除筛选器-指定是包括还是排除某些文件名、文件类型和文件路径。你可以使用 glob 模式（可以将通配符模式扩展为与给定模式匹配的路径名列表的模式）。有关示例，请参阅《 AWS CLI 命令参考》中的 “排除和包含筛选器的使用”。
- 文档元数据和访问控制配置-添加包含源 URI、文档作者或自定义文档属性/字段等信息的文档元数据和访问控制文件，以及您的用户以及他们可以访问哪些文档。每个元数据文件都包含有关单个文档的元数据。
- 字段映射 - 选择将 S3 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅 S3 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与 S3 数据源集成的更多信息，请参阅：

- 使用支持 VPC Amazon Kendra 的 S3 连接器准确搜索答案

创建 Amazon S3 数据源

以下示例演示如何创建 Amazon S3 数据源。这些示例假设您已经创建了一个索引和一个有权从索引中读取数据的 IAM 角色。有关该 IAM 角色的更多信息，请参阅IAM 访问角色。有关创建索引的更多信息，请参阅创建索引。

CLI

```
aws kendra create-data-source \
 --index-id index ID \
 --name example-data-source \
 --type S3 \
 --configuration '{"S3Configuration":{"BucketName":"bucket name"}}'
 --role-arn 'arn:aws:iam::account id:role:/role name
```

Python

以下 Python 代码片段创建了一个 Amazon S3 数据源。有关完整示例，请参阅入门 (AWS SDK for Python (Boto3))。

```
print("Create an Amazon S3 data source.")

    # Provide a name for the data source
    name = "getting-started-data-source"
    # Provide an optional description for the data source
    description = "Getting started data source."
    # Provide the IAM role ARN required for data sources
    role_arn = "arn:aws:iam::${accountID}:role/${roleName}"
    # Provide the data soource connection information
    s3_bucket_name = "S3-bucket-name"
    type = "S3"
    # Configure the data source
    configuration = {"S3DataSourceConfiguration":
        {
            "BucketName": s3_bucket_name
        }
    }

    data_source_response = kendra.create_data_source(
        Configuration = configuration,
```

```
        Name = name,
        Description = description,
        RoleArn = role_arn,
        Type = type,
        IndexId = index_id
    )
```

创建数据来源可能需要一些时间。您可以使用 DescribeDataSourceAPI 监控进度。当数据来源状态为 ACTIVE 时，数据来源就已准备就绪。

以下示例演示如何获取数据来源的状态。

CLI

```
aws kendra describe-data-source \
 --index-id index ID \
 --id data source ID
```

Python

以下 Python 代码片段用于获取有关 S3 数据来源的信息。有关完整示例，请参阅入门 (AWS SDK for Python (Boto3))。

```
print("Wait for Amazon Kendra to create the data source.")

    while True:
        data_source_description = kendra.describe_data_source(
            Id = "data-source-id",
            IndexId = "index-id"
        )
        status = data_source_description["Status"]
        print(" Creating data source. Status: "+status)
        time.sleep(60)
        if status != "CREATING":
            break
```

此数据来源没有计划，因此不会自动运行。要为数据源编制索引，StartDataSourceSyncJob需要调用将索引与数据源同步。

以下示例演示如何同步数据来源。

CLI

```
aws kendra start-data-source-sync-job \
 --index-id index ID \
 --id data source ID
```

Python

以下 Python 代码片段同步了一个 Amazon S3 数据来源。有关完整示例，请参阅入门 (AWS SDK for Python (Boto3))。

```
print("Synchronize the data source.")

    sync_response = kendra.start_data_source_sync_job(
        Id = "data-source-id",
        IndexId = "index-id"
    )
```

Amazon S3 文档元数据

您可以使用元数据文件向 Amazon S3 存储桶中的文档添加元数据（有关文档的其他信息）。每个元数据文件都与一个已编入索引的文档相关联。

您的元数据文件必须与已编入索引的文件存储在同一个存储桶中。创建数据源时，您可以使用控制台或DocumentsMetadataConfiguration参数S3Prefix字段为元 Amazon S3 数据文件指定存储桶中的位置。如果未指定 Amazon S3 前缀，则元数据文件必须与已编入索引的文档存储在相同的位置。

如果您为元数据文件指定 Amazon S3 前缀，则这些文件位于与已编入索引的文档平行的目录结构中。Amazon Kendra 仅在指定目录中查找您的元数据。如果未读取元数据，请检查目录位置是否与元数据的位置相匹配。

以下示例展示了如何将已编入索引的文档位置映射到元数据文件位置。请注意，文档的 Amazon S3 密钥会附加到元数据的前 Amazon S3 缀后面，然后在后缀后面加上.metadata.json以形成元数据文件的路径。 Amazon S3 带有元数据 Amazon S3 前缀和.metadata.json后缀的组合 Amazon S3 密钥总长度不得超过 1024 个字符。建议将 Amazon S3 密钥保持在 1000 个字符以下，以便在将密钥与前缀和后缀组合时考虑其他字符。

```
Bucket name:
     s3://bucketName
```

```
Document path:
     documents
Metadata path:
     none
File mapping
     s3://bucketName/documents/file.txt ->
        s3://bucketName/documents/file.txt.metadata.json
```

```
Bucket name:
     s3://bucketName
Document path:
     documents/legal
Metadata path:
     metadata
File mapping
     s3://bucketName/documents/legal/file.txt ->
        s3://bucketName/metadata/documents/legal/file.txt.metadata.json
```

文档元数据在 JSON 文件中定义。该文件必须是没有 BOM 标记的 UTF-8 文本文件。JSON 文件的文件名必须是 <document>.<extension>.metadata.json。在此示例中，“document”是应用元数据的文档的名称，“extension”是该文档的文件扩展名。在 <document>.<extension>.metadata.json 中，文档 ID 必须是唯一的。

JSON 文件的内容遵循此模板。所有属性/字段都是可选的，因此不必包含所有属性。必须为要包含的每个属性提供一个值；该值不能为空。如果您未指定_source_uri，则搜索结果 Amazon Kendra 中返回的链接将指向包含该文档的 Amazon S3 存储桶。 DocumentId映射到字段，s3_document_id并且是 S3 中文档的绝对路径。

```
{
    "DocumentId": "S3 document ID, the S3 path to doc",
    "Attributes": {
        "_category": "document category",
        "_created_at": "ISO 8601 encoded string",
        "_last_updated_at": "ISO 8601 encoded string",
        "_source_uri": "document URI",
        "_version": "file version",
        "_view_count": number of times document has been viewed,
        "custom attribute key": "custom attribute value",
        additional custom attributes
    },
    "AccessControlList": [
```

```
        {
            "Name": "user name",
            "Type": "GROUP | USER",
            "Access": "ALLOW | DENY"
        }
    ],
    "Title": "document title",
    "ContentType": "For example HTML | PDF. For supported content types, see Types of
 documents."
}
```

_created_at 和 _last_updated_at 元数据字段是 ISO 8601 编码的日期。例如，2012-03-25T12:30:10+01:00 是中部欧洲时间 2012 年 3 月 25 日中午 12:30（10 秒）的 ISO 8601 日期-时间格式。

您可以向 Attributes 字段添加有关文档的其他信息，该文档用于筛选查询或对查询响应进行分组。有关更多信息，请参阅 创建自定义文档字段。

您可以使用 AccessControlList 字段来筛选查询响应。这样，只有特定的用户和组才能访问文档。有关更多信息，请参阅 根据用户上下文进行筛选。

Amazon S3 数据源的访问控制

您可以使用配置文件控制对 Amazon S3 数据源中文档的访问权限。您可以在控制台中指定文件，或者在调用CreateDataSource或 UpdateDataSourceAPI 时将其指定为AccessControlListConfiguration参数。

配置文件包含标识 S3 前缀并列出该前缀的访问设置的 JSON 结构。该前缀可以是路径，也可以是单个文件。如果该前缀是路径，则会将访问设置应用于该路径中的所有文件。JSON 配置文件中有 S3 前缀的最大数量和默认的最大文件大小。有关更多信息，请参阅 的配额 Amazon Kendra。

您可以在访问设置中同时指定用户和组。查询索引时，您可以指定用户和组信息。有关更多信息，请参阅 按属性筛选。

配置文件的 JSON 结构必须为以下格式：

```
[
    {
        "keyPrefix": "s3://BUCKETNAME/prefix1/",
        "aclEntries": [
            {
```

```
                "Name": "user1",
                "Type": "USER",
                "Access": "ALLOW"
            },
            {
                "Name": "group1",
                "Type": "GROUP",
                "Access": "DENY"
            }
        ]
    },
    {
        "keyPrefix": "s3://prefix2",
        "aclEntries": [
            {
                "Name": "user2",
                "Type": "USER",
                "Access": "ALLOW"
            },
            {
                "Name": "user1",
                "Type": "USER",
                "Access": "DENY"
            },
            {
                "Name": "group1",
                "Type": "GROUP",
                "Access": "DENY"
            }
        ]
    }
]
```

Amazon VPC 与 Amazon S3 数据源一起使用

本主题提供的 step-by-step 示例展示了如何使用亚马逊 S3 连接器通过 Amazon VPC 连接到 Amazon S3 存储桶。该示例假设您从现有的 S3 存储桶开始。我们建议您仅将几个文档上传到 S3 存储桶以测试示例。

您可以通过 Amazon Kendra 连接到您的 Amazon S3 存储桶 Amazon VPC。为此，您必须在创建 Amazon S3 数据源连接器时指定 Amazon VPC 子网和 Amazon VPC 安全组。

> **Important**
>
> 为了让 Amazon Kendra Amazon S3 连接器可以访问您的 Amazon S3 存储桶，请确保您已为虚拟私有云 (VPC) 分配了 Amazon S3 终端节点。

Amazon Kendra 要通过同步 Amazon S3 存储桶中的文档 Amazon VPC，您必须完成以下步骤：

- 为设置 Amazon S3 终端节点 Amazon VPC。有关如何设置 Amazon S3 终端节点的更多信息，请参阅AWS PrivateLink 指南 Amazon S3中的网关终端节点。
- （可选）已检查您的 Amazon S3 存储桶策略，确保可以从您分配到的虚拟私有云 (VPC) 访问 Amazon S3 存储桶 Amazon Kendra。有关更多信息，请参阅 Amazon S3 用户指南中的使用存储桶策略控制 VPC 终端节点的访问

步骤

- 步骤 1：配置 Amazon VPC
- （可选）步骤 2：配置 Amazon S3 存储桶策略
- 步骤 3：创建测试 Amazon S3 数据源连接器

步骤 1：配置 Amazon VPC

创建一个 VPC 网络，包括一个带有网 Amazon S3 关终端节点和安全组的私有子网 Amazon Kendra ，供以后使用。

为 VPC 配置私有子网、S3 终端节点和安全组

1. 登录 AWS Management Console 并打开 Amazon VPC 控制台，网址为https://console.aws.amazon.com/vpc/。
2. 创建具有私有子网和 S3 终端节点的 VPC Amazon Kendra 以供使用：

   在导航窗格中，选择您的 VPC，然后选择创建 VPC。

   a. 对于要创建的资源，选择 VPC 等。
   b. 对于 “名称标签”，启用 “自动生成”，然后输入**kendra-s3-example**。
   c. 对于 IPv4/IPv6 CIDR 块，请保留默认值。
   d. 对于可用区 (AZ) 的数量，请选择数字 1。

e. 选择 “自定义可用区”，然后从第一个可用区列表中选择一个可用区。

Amazon Kendra 仅支持一组特定的可用区。

f. 在 “公有子网数量” 中，选择数字 0。

g. 在私有子网数量中，选择数字 1。

h. 对于 NAT gateways（NAT 网关），选择 None（无）。

i. 对于 VPC 终端节点，请选择Amazon S3 网关。 。

j. 将其余值保留为默认设置。

k. 选择 Create VPC。

等到创建 VPC 工作流程完成。然后，选择查看 VPC 以检查您刚刚创建的 VPC。

现在，您已经创建了一个带有私有子网的 VPC 网络，该子网无法访问公共互联网。

3. 复制您的 Amazon S3 终端节点的 VPC 终端节点 ID：

a. 在导航窗格中，选择端点。

b. 在终端节点列表中，找到您刚刚与您的 VPC 一起创建的 Amazon S3 终端节点`kendra-s3-example-vpce-s3`。

c. 记下 VPC 终端节点 ID。

现在，您已经创建了一个 Amazon S3 网关终端节点，用于通过子网访问您的 Amazon S3 存储桶。

4. 创建安全组 Amazon Kendra 以供使用：

a. 在导航窗格中，选择安全组，然后选择创建安全组。

b. 对于安全组名称，输入 **`s3-data-source-security-group`**。

c. 从Amazon VPC列表中选择您的 VPC。

d. 将入站规则和出站规则保留为默认值。

e. 选择创建安全组。

现在，您已经创建了一个 VPC 安全组。

在连接器配置过程中，您将创建的子网和安全组分配给 Amazon Kendra Amazon S3 数据源连接器。

（可选）步骤 2：配置 Amazon S3 存储桶策略

在此可选步骤中，学习如何配置 Amazon S3 存储桶策略，以便只能从您分配给的 VPC 访问您的 Amazon S3 存储桶 Amazon Kendra。

Amazon Kendra 使用 IAM 角色访问您的 Amazon S3 存储桶，并且不需要您配置 Amazon S3 存储桶策略。但是，如果您想使用 Amazon S3 存储桶配置 Amazon S3 连接器，而该存储桶具有限制从公共 Internet 访问的现有策略，那么创建存储桶策略可能会很有用。

配置您的 Amazon S3 存储桶策略

1. 打开 Amazon S3 控制台，网址为 https://console.aws.amazon.com/s3/。
2. 从导航窗格中选择 Buckets。
3. 选择您要与之同步的 Amazon S3 存储桶的名称 Amazon Kendra。
4. 选择 “权限” 选项卡，向下滚动到 “存储桶策略”，然后单击 “编辑”。
5. 添加或修改您的存储桶策略，使其仅允许从您创建的 VPC 终端节点进行访问。

   下面是一个示例存储桶策略。将*bucket-name*和*vpce-id*替换为您的 Amazon S3 存储桶名称和您之前记下的 Amazon S3 终端节点 ID。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Deny",
      "Principal": "*",
      "Action": "s3:*",
      "Resource": "arn:aws:s3:::bucket-name/*",
      "Condition": {
        "StringNotEquals": {
          "aws:SourceVpce": "vpce-id"
        }
      }
    }
  ]
}
```

6. 选择保存更改。

现在，只能从您创建的特定 VPC 访问您的 S3 存储桶。

步骤 3：创建测试 Amazon S3 数据源连接器

要测试您的 Amazon VPC 配置，请创建一个 Amazon S3 连接器。然后，按照中概述的步骤，使用您创建的 VPC 对其进行配置Amazon S3。

对于 Amazon VPC 配置值，请选择您在本示例中创建的值：

- Amazon VPC(VPC) — `kendra-s3-example-vpc`
- 子网 — `kendra-s3-example-subnet-private1-[availability zone]`
- 安全组 — `s3-data-source-security-group`

等待连接器完成创建。创建 Amazon S3 连接器后，选择 “立即同步” 以启动同步。

完成同步可能需要几分钟到几小时，具体取决于 Amazon S3 存储桶中有多少文档。为了测试该示例，我们建议您只将几个文档上传到 S3 存储桶。如果您的配置正确，您最终应该会看到同步状态为 “已完成”。

如果遇到任何错误，请参阅Amazon VPC 连接疑难解答。

Amazon Kendra 网络爬虫

您可以使用 Amazon Kendra Web Crawler 来抓取和索引网页。

您只能爬取公共网站和使用安全通信协议（安全超文本传输协议（HTTPS））的公司内部网站。如果您在爬取网站时收到错误，则可能是该网站被阻止爬网。要爬取内部网站，可以设置 Web 代理。Web 代理必须面向公众。您还可以使用身份验证来访问和爬取网站。

当选择要编制索引的网站时，您必须遵守 Amazon 可接受使用政策以及所有其他 Amazon 条款。请记住，您只能使用 Amazon Kendra Web Crawler 来索引自己的网页或您有权编制索引的网页。要了解如何阻止 Amazon Kendra Web Crawler 将您的网站编入索引，请参阅为 Amazon Kendra Web 爬网程序配置 *robots.txt* 文件。

> **Note**
>
> 滥用 Amazon Kendra Web Crawler 来积极抓取你不拥有的网站或网页是不被视为不可接受的用法。

Amazon Kendra 有两个版本的web crawler连接器。每个版本支持的功能包括：

Amazon Kendra Web Crawler 连接器 v1.0/API WebCrawlerConfiguration

- Web 代理
- 包含/排除筛选条件

Amazon Kendra Web Crawler 连接器 v2.0/API TemplateConfiguration

- 字段映射
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Web 代理
- 网站的基本、NTLM/Kerberos、SAML 和表单身份验证
- Virtual Private Cloud (VPC)

> **Important**
>
> 不支持 Web Crawler v2.0 连接器的创建。 AWS CloudFormation如果需要 AWS CloudFormation 支持，请使用 Web Crawler v1.0 连接器。

要对 Amazon Kendra 网络爬虫数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- Amazon Kendra 网络爬虫连接器 v1.0
- Amazon Kendra 网络爬虫连接器 v2.0
- 为 Amazon Kendra Web 爬网程序配置 robots.txt 文件

Amazon Kendra 网络爬虫连接器 v1.0

您可以使用 Amazon Kendra Web Crawler 来抓取和索引网页。

您只能爬取公共网站和使用安全通信协议（安全超文本传输协议（HTTPS））的网站。如果您在爬取网站时收到错误，则可能是该网站被阻止爬网。要爬取内部网站，可以设置 Web 代理。Web 代理必须面向公众。

当选择要编制索引的网站时，您必须遵守 Amazon 可接受使用政策以及所有其他 Amazon 条款。请记住，您只能使用 Amazon Kendra Web Crawler 来索引自己的网页或您有权编制索引的网页。要了解如

何阻止 Amazon Kendra Web Crawler 将您的网站编入索引，请参阅为 Amazon Kendra Web 爬网程序配置 *robots.txt* 文件。

> **Note**
>
> 滥用 Amazon Kendra Web Crawler 来积极抓取你不拥有的网站或网页是不被视为不可接受的用法。

要对 Amazon Kendra 网络爬虫数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

- Web 代理
- 包含/排除筛选条件

先决条件

在使用 Amazon Kendra 索引您的网站之前，请先检查您的网站和 AWS 帐户的详细信息。

对于您的网站，请确保：

- 已复制要编制索引的网站的种子或站点地图 URL。
- 对于需要基本身份验证的网站：记下用户名和密码，并复制网站的主机名和端口号。
- 可选：如果您想使用 Web 代理连接到要爬取的内部网站，请复制网站的主机名和端口号。该 Web 代理必须面向公众。 Amazon Kendra 支持连接到通过基本身份验证提供支持的 Web 代理服务器，或者无需身份验证即可连接的服务器。
- 在网页以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在您的 AWS 账户中，请确保您有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 对于需要身份验证的网站，或者如果使用带身份验证的 Web 代理，请将您的身份验证凭据存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将web crawler数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的web crawler数据源，您必须提供web crawler数据源的必要详细信息，Amazon Kendra 以便访问您的数据。如果您尚未进行配置，web crawler请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 web crawler

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在 “添加数据源” 页面上，选择 Web Crawler 连接器，然后选择 “添加连接器”。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Web 爬网程序连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 对于来源，请根据您的使用案例选择来源 URL 和来源站点地图，然后分别输入值。

   您最多可以添加 10 个来源 URL 和三个站点地图。

   > **Note**
   >
   > 如果要爬取站点地图，请检查基本 URL 或根 URL 是否与站点地图页面上列出的 URL 相同。例如，如果您的站点地图 URL 是 https://example.com/sitemap-page.html，则此站点地图页面上列出的 URL 也应使用基本 URL“https://example.com/”。

   b. （可选）对于 Web 代理 - 请输入以下信息：

      i. 主机名 - 需要 Web 代理的主机名。

      ii. 端口号 - 主机 URL 传输协议使用的端口。端口号应为介于 0 到 65535 之间的数值。

      iii. 对于 Web 代理凭证 - 如果您的 Web 代理连接需要身份验证，请选择现有密钥或创建新密钥来存储您的身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

iv. 在创建 AWS Secrets Manager Secrets Manager 密钥窗口中输入以下信息：

A. 密钥名称 - 密钥的名称。前缀“AmazonKendra-WebCrawler-”会自动添加到您的密钥名称中。

B. 对于用户名和密码 - 输入网站的这些基本身份验证凭证。

C. 选择保存。

c. （可选）需要身份验证的主机 - 选择添加其他需要身份验证的主机。

d. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

e. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 爬取范围 - 选择要爬取的网页类型。

b. 爬行深度-从种子 URL 中选择 Amazon Kendra 应该爬行的关卡数。

c. 高级爬取设置和其他配置，输入以下信息：

i. 最大文件大小 - 要爬取的最大网页或附件大小。最小 0.000001 MB（1 字节）。最大 50 MB。

ii. 每页最多链接数 - 在每个页面上爬取的最大链接数量。按显示顺序爬取链接。最少 1 个链接/页面。最多 1000 个链接/页面。

iii. 最大限制 - 每分钟爬取的每个主机名的最大 URL 数量。每分钟每个主机名最少 1 个 URL。每分钟每个主机名最多 300 个 URL。

iv. 正则表达式模式 - 添加包含或排除某些 URL 的正则表达式模式。最多可以添加 100 个模式。

d. 在 “同步运行计划” 中，“频率”-选择与数据源同步的频率。 Amazon Kendra

e. 选择下一步。

8. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 web crawler

您必须使用 WebCrawlerConfigurationAPI 指定以下内容：

- URL - 指定您想要使用 SeedUrlConfiguration 和 SiteMapsConfiguration 爬取的网站或站点地图 URL 的种子或起点 URL。

  > **Note**
  >
  > 如果要爬取站点地图，请检查基本 URL 或根 URL 是否与站点地图页面上列出的 URL 相同。例如，如果您的站点地图 URL 是 https://example.com/sitemap-page.html，则此站点地图页面上列出的 URL 也应使用基本 URL“https://example.com/”。

- 密钥 Amazon资源名称（ARN）- 如果网站需要基本身份验证，则需要提供主机名、端口号和用于存储您的用户名和密码的基本身份验证凭证的密钥。您可以使用 AuthenticationConfiguration API 提供密钥 ARN。密钥必须使用具有以下键的 JSON 结构存储：

  ```
  {
      "username": "user name",
      "password": "password"
  }
  ```

  您也可以使用 AWS Secrets Manager 密钥提供 Web 代理凭证。您可以使用 ProxyConfiguration API 提供网站主机名称和端口号，还可以选择提供存储您的 Web 代理凭证的密钥。

- IAM ro le —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Web 爬网程序连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Web 爬网程序数据来源的IAM 角色。

您还可以添加以下可选功能：

- 爬取模式 - 选择是仅爬取网站主机名，还是爬取带有子域的主机名，还是同时爬取网页链接到的其他域名。
- 从种子层爬取的“深度”或层数。例如，种子 URL 页面的深度为 1，在该页面上同时爬取的所有超链接的深度都是 2。
- 爬取单个网页时要包含的 URL 的最大数量。
- 要爬取的网页的最大大小（以 MB 为单位）。

- 每分钟爬取的每个网站主机的最大 URL 数量。
- 用于连接和爬取内部网站的 Web 代理主机和端口号。例如，https://a.example.com/page1.html 的主机名是“a.example.com”，端口号是 443，这是 HTTPS 的标准端口。如果连接至网站主机需要 Web 代理凭证，您可以创建存储凭证的 AWS Secrets Manager 。
- 访问和爬取需要用户身份验证的网站的身份验证信息。
- 您可以使用自定义文档富集工具将 HTML 元标签提取为字段。有关更多信息，请参阅在提取过程中自定义文档元数据。有关提取 HTML 元标签的示例，请参阅 CDE 示例。
- 包含和排除筛选条件 - 指定是包含还是排除 URL：

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

了解更多信息

要了解有关 Amazon Kendra 与web crawler数据源集成的更多信息，请参阅：

- 使用 Amazon Kendra Web Crawler 重新构想知识发现

Amazon Kendra 网络爬虫连接器 v2.0

您可以使用 Amazon Kendra Web Crawler 来抓取和索引网页。

您只能爬取公共网站和使用安全通信协议（安全超文本传输协议（HTTPS））的公司内部网站。如果您在爬取网站时收到错误，则可能是该网站被阻止爬网。要爬取内部网站，可以设置 Web 代理。Web 代理必须面向公众。您还可以使用身份验证来访问和爬取网站。

Amazon Kendra Web Crawler v2.0 使用 Selenium 网络爬虫软件包和 Chromium 驱动程序。 Amazon Kendra 使用持续集成 (CI) 自动更新 Selenium 和 Chromium 驱动程序的版本。

当选择要编制索引的网站时，您必须遵守 Amazon 可接受使用政策以及所有其他 Amazon 条款。请记住，您只能使用 Amazon Kendra Web Crawler 来索引自己的网页或您有权编制索引的网页。要了解如何阻止 Amazon Kendra Web Crawler 将您的网站编入索引，请参阅为 Amazon Kendra Web 爬网程序

配置 *robots.txt* 文件。 。滥用 Amazon Kendra Web Crawler 来积极抓取你不拥有的网站或网页是不被视为不可接受的用法。

要对 Amazon Kendra 网络爬虫数据源连接器进行故障排除，请参阅数据来源故障排除。

> **Note**
>
> Web Crawler 连接器 v2.0 不支持从 AWS KMS 加密存储桶中抓取网站列表。 Amazon S3 它仅支持使用 Amazon S3 托管密钥进行服务器端加密。

> **Important**
>
> 不支持 Web Crawler v2.0 连接器的创建。 AWS CloudFormation如果需要 AWS CloudFormation 支持，请使用 Web Crawler v1.0 连接器。

主题

- 支持的特征
- 先决条件
- 连接说明

支持的特征

- 字段映射
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Web 代理
- 网站的基本、NTLM/Kerberos、SAML 和表单身份验证
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引您的网站之前，请先检查您的网站和 AWS 帐户的详细信息。

对于您的网站，请确保：

- 已复制要编制索引的网站的种子或站点地图 URL。您可以将 URL 存储在文本文件中，然后将其上传到 Amazon S3 存储桶。文本文件中的每个 URL 都必须对单行进行格式化。如果您想将站点地图存储在 Amazon S3 存储桶中，请确保已复制站点地图 XML 并将其保存在 XML 文件中。您也可以将多个站点地图 XML 文件压缩成一个 ZIP 文件。

  > **Note**
  >
  > （本地/服务器）Amazon Kendra 会检查中 AWS Secrets Manager 包含的端点信息是否与数据源配置详细信息中指定的端点信息相同。这有助于防止出现混淆代理人问题，这是一个安全问题，即用户无权执行操作，但可以将 Amazon Kendra 作为代理来访问配置的密钥和执行操作。如果以后更改端点信息，则必须创建一个新密钥来同步此信息。

- 对于需要基本、NTLM 或 Kerberos 身份验证的网站：
  - 已记下您的网站身份验证凭证，其中包括用户名和密码。

  > **Note**
  >
  > Amazon Kendra Web Crawler v2.0 支持包括密码哈希的 NTLM 身份验证协议和包括密码加密的 Kerberos 身份验证协议。

- 对于需要 SAML 或登录表单身份验证的网站：
  - 已记下您的网站身份验证凭证，其中包括用户名和密码。
  - 已复制用户名字段（如果使用 SAML，则还包括用户名按钮）、密码字段和按钮的 XPaths（XML 路径语言），并复制了登录页面 URL。您可以使用 Web 浏览器的开发者工具找到元素的 XPaths。XPaths 通常遵循以下格式：`//tagname[@Attribute='Value']`.

  > **Note**
  >
  > Amazon Kendra Web Crawler v2.0 使用无头 Chrome 浏览器和表单中的信息，通过受 OAuth 2.0 保护的网址对访问进行身份验证和授权。

- 可选：如果您想使用 Web 代理连接到要爬取的内部网站，请复制 Web 代理服务器的主机名和端口号。Web 代理必须面向公众。 Amazon Kendra 支持连接到由基本身份验证支持的 Web 代理服务器，或者无需身份验证即可连接。
- 可选：如果您想使用 VPC 连接到要爬取的内部网站，请复制虚拟私有云（VPC）子网 ID。有关更多信息，请参阅配置 Amazon VPC。

- 在网页以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在您的 AWS 账户中，请确保您有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该 IAM 角色的 Amazon 资源名称。

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 对于需要身份验证的网站，或者如果使用带身份验证的 Web 代理，请将您的身份验证凭据存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将web crawler数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的web crawler数据源，您必须提供web crawler数据源的必要详细信息，Amazon Kendra 以便访问您的数据。如果您尚未进行配置，web crawler请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 web crawler

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在 “添加数据源” 页面上，选择 Web Crawler 连接器，然后选择 “添加连接器”。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Web 爬网程序连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 来源 - 选择来源 URL、来源站点地图、来源 URL 文件、源站点地图文件。如果您选择使用包含最多 100 个种子 URL 列表的文本文件，则需要指定 Amazon S3 存储文件的存储桶的路径。如果您选择使用站点地图 XML 文件，则需要指定存储文件的 Amazon S3 存储桶的路径。您也可以将多个站点地图 XML 文件压缩成一个 ZIP 文件。否则，您最多可以手动输入 10 个种子或起点 URL，以及最多三个站点地图 URL。

   > **Note**
   >
   > 如果要爬取站点地图，请检查基本 URL 或根 URL 是否与站点地图页面上列出的 URL 相同。例如，如果您的站点地图 URL 是 https://example.com/sitemap-page.html，则此站点地图页面上列出的 URL 也应使用基本 URL“https://example.com/”。

如果您的网站需要身份验证才能访问，则可以选择基本身份验证、NTLM/Kerberos、SAML 或表单身份验证。否则，请选择不进行身份验证的选项。

> **Note**
>
> 如果您想稍后编辑数据来源以更改带有站点地图身份验证的种子 URL，则必须创建一个新的数据来源。 Amazon Kendra 使用 Secrets Manager 密钥中的种子 URL 端点信息为身份验证配置数据来源，因此在更改为站点地图时无法重新配置数据来源。

- AWS Secrets Manager s@@ ecret —如果您的网站需要相同的身份验证才能访问网站，请选择现有密钥或创建新 Secrets Manager 密钥来存储您的网站凭据。如果您选择创建新密钥，则会打开一个 AWS Secrets Manager 秘密窗口。

  如果您选择基本或 NTML/Kerberos 身份验证，请输入密钥的名称以及用户名和密码。NTLM 身份验证协议包括密码哈希，Kerberos 身份验证协议包括密码加密。

  如果您选择 SAML 或表单身份验证，请输入密钥的名称以及用户名和密码。使用用户名字段的 XPath（如果使用 SAML，则使用用户名按钮的 XPath）。使用密码字段和按钮的 XPaths，以及登录页面 URL。您可以使用 Web 浏览器的开发者工具找到元素的 XPaths（XML 路径语言）。XPaths 通常遵循以下格式：`//tagname[@Attribute='Value']`.

b. （可选）Web 代理 - 输入要用于连接内部网站的代理服务器的主机名和端口号。例如，https://a.example.com/page1.html 的主机名是“a.example.com”，端口号是 443，这是 HTTPS 的标准端口。如果需要网络代理凭据才能连接到网站主机，则 AWS Secrets Manager 可以创建存储凭据的。

c. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

d. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

e. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 同步范围 - 设置爬取网页的限制，包括其域名、文件大小和链接；并使用正则表达式模式筛选 URL。

      i. （可选）爬取域名范围 - 选择是仅爬取网站域名、包含子域名的域名，还是同时爬取网页链接到的其他域名。默认情况下， Amazon Kendra 仅抓取您要抓取的网站的域名。

      ii. （可选）其他配置 - 设置以下设置：

      - 爬取深度 -“深度”或从种子层开始的爬取层数。例如，种子 URL 页面的深度为 1，在该页面上同时爬取的所有超链接的深度都是 2。
      - 最大文件大小 - 要爬取的网页或附件的最大大小（以 MB 为单位）。
      - 每页的最大链接数量 - 要爬取的单个网页的最大 URL 数量。
      - 爬取速度的最大限制 - 每分钟爬取的每个网站主机名的最大 URL 数量。
      - 文件 - 选择爬取网页链接到的文件。
      - 爬取和索引 URL - 添加正则表达式模式以包含或排除爬取某些 URL，以及将这些 URL 网页上的任何超链接编入索引。

   b. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

      - 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
      - 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

   c. 同步运行计划 - 对于频率，选择 Amazon Kendra 与数据来源同步的频率。

   d. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

   a. 从 Amazon Kendra 生成的网页和文件默认字段中选择要映射到索引的网页。

   b. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 web crawler

您必须使用 TemplateConfiguration API 指定数据来源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构WEBCRAWLERV2时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- URL - 指定您想要使用爬取的网站或站点地图 URL 的种子或起点 URL。您可以指定存储种子 URL 列表的 Amazon S3 存储桶的路径。种子 URL 的文本文件中的每个 URL 都必须对单行进行格式化。您还可以指定存储站点地图 XML 文件的 Amazon S3 存储桶的路径。您可以将多个站点地图文件压缩成一个 ZIP 文件，然后将 ZIP 文件存储在 Amazon S3 存储桶中。

  > **Note**
  >
  > 如果要爬取站点地图，请检查基本 URL 或根 URL 是否与站点地图页面上列出的 URL 相同。例如，如果您的站点地图 URL 是 https://example.com/sitemap-page.html，则此站点地图页面上列出的 URL 也应使用基本 URL“https://example.com/”。

- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅为新增、修改和删除的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 身份验证 - 如果您的网站需要相同的身份验证，请指定 BasicAuth、NTLM_Kerberos、SAML 或 Form 身份验证。如果您的网站不需要身份验证，请指定NoAuthentication。
- 密钥 Amazon 资源名称（ARN）- 如果您的网站需要基本、NTLM 或 Kerberos 身份验证，则需要提供存储您的用户名和密码的身份验证凭证的密钥。您要提供 AWS Secrets Manager 密钥的 Amazon 资源名称（ARN）。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "seedUrlsHash": "Hash representation of all seed URLs",
    "userName": "user name",
    "password": "password"
}
```

如果您的网站需要 SAML 身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "seedUrlsHash": "Hash representation of all seed URLs",

    "userName": "user name",
    "password": "password",
    "userNameFieldXpath": "XPath for user name field",
    "userNameButtonXpath": "XPath for user name button",
    "passwordFieldXpath": "XPath for password field",
    "passwordButtonXpath": "XPath for password button",
    "loginPageUrl": "Full URL for website login page"
}
```

如果您的网站需要表单身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "seedUrlsHash": "Hash representation of all seed URLs",
    "userName": "user name",
    "password": "password",
    "userNameFieldXpath": "XPath for user name field",
    "passwordFieldXpath": "XPath for password field",
    "passwordButtonXpath": "XPath for password button",
    "loginPageUrl": "Full URL for website login page"
}
```

您可以使用 Web 浏览器的开发者工具找到元素的 XPaths（XML 路径语言）。XPaths 通常遵循以下格式：`//tagname[@Attribute='Value']`.

您也可以使用 AWS Secrets Manager 密钥提供 Web 代理凭证。

- IAM ro le —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Web 爬网程序连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Web 爬网程序数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 `VpcConfiguration`，以便调用 `CreateDataSource`。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。

- 域范围 - 选择是仅爬取带有子域名的网站域名，还是同时爬取网页链接到的其他域名。默认情况下， Amazon Kendra 仅抓取您要抓取的网站的域名。
- 从种子层爬取的“深度”或层数。例如，种子 URL 页面的深度为 1，在该页面上同时爬取的所有超链接的深度都是 2。
- 爬取单个网页时要包含的 URL 的最大数量。
- 要爬取的网页或附件的最大大小（以 MB 为单位）。
- 每分钟爬取的每个网站主机的最大 URL 数量。
- 用于连接和爬取内部网站的 Web 代理主机和端口号。例如，https://a.example.com/page1.html 的主机名是“a.example.com”，端口号是 443，这是 HTTPS 的标准端口。如果连接至网站主机需要 Web 代理凭证，您可以创建存储凭证的 AWS Secrets Manager 。
- 包含和排除筛选条件 - 指定是包括还是排除爬取某些 URL 并将这些 URL 网页上的任何超链接编入索引。

  > **Note**
  >
  > 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 字段映射-选择将网页和网页文件的字段映射到您的 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

有关要配置的其他重要 JSON 键的列表，请参阅 Amazon Kendra Web 爬网程序模板架构。

为 Amazon Kendra Web 爬网程序配置 `robots.txt` 文件

Amazon Kendra 是一种智能搜索服务， AWS 客户使用它来索引和搜索自己选择的文档。为了索引网络上的文档，客户可以使用 Amazon Kendra Web Crawler，指明应为哪些 URL 编制索引以及其他操作参数。 Amazon Kendra 在为任何特定网站编制索引之前，客户必须获得授权。

Amazon Kendra Web Crawler 尊重标准 robots.txt 指令，例如`Allow`和。`Disallow`您可以修改网站`robots.txt`文件以控制 Amazon Kendra Web Crawler 如何抓取您的网站。

配置 Amazon Kendra Web Crawler 如何访问您的网站

您可以使用Allow和指Disallow令控制 Amazon Kendra Web Crawler 如何为您的网站编制索引。您还可以控制为哪些网页编制索引，以及不爬取哪些网页。

要允许 Amazon Kendra Web Crawler 抓取除不允许的网页之外的所有网页，请使用以下指令：

```
User-agent: amazon-kendra    # Amazon Kendra Web Crawler
Disallow: /credential-pages/ # disallow access to specific pages
```

要允许 Amazon Kendra Web Crawler 仅抓取特定的网页，请使用以下指令：

```
User-agent: amazon-kendra    # Amazon Kendra Web Crawler
Allow: /pages/ # allow access to specific pages
```

要允许 Amazon Kendra Web Crawler 抓取所有网站内容并禁止任何其他机器人抓取，请使用以下指令：

```
User-agent: amazon-kendra # Amazon Kendra Web Crawler
Allow: / # allow access to all pages
User-agent: * # any (other) robot
Disallow: / # disallow access to any pages
```

阻止 Amazon Kendra Web Crawler 抓取您的网站

您可以使用该Disallow指令阻止 Amazon Kendra Web Crawler 将您的网站编入索引。您还可以控制爬取哪些网页以及不爬取哪些网页。

要阻止 Amazon Kendra Web Crawler 抓取网站，请使用以下指令：

```
User-agent: amazon-kendra # Amazon Kendra Web Crawler
Disallow: / # disallow access to any pages
```

Amazon Kendra Web Crawler 还支持 HTML 页面中元标记中的机器noindex人和nofollow指令。这些指令会阻止 Web 爬网程序将网页编入索引，并停止跟踪网页上的任何链接。在文档的部分中添加元标签可指定机器人规则。

例如，以下网页包含机器人指令 noindex 和 nofollow：

```
<html>
<head>
    <meta name="robots" content="noindex, nofollow"/>
    ...
</head>
<body>...</body>
</html>
```

如果您对 Amazon Kendra Web Crawler 有任何疑问或疑虑，可以联系AWS 支持团队。

Amazon WorkDocs

Amazon WorkDocs 是一项用于创建、编辑、存储和共享内容的安全内容协作服务。您可以使用索 Amazon Kendra 引您的 Amazon WorkDocs 数据源。

您可以使用Amazon Kendra 控制台和 WorkDocsConfigurationAPI Amazon Kendra 连接到您的 Amazon WorkDocs 数据源。

Amazon WorkDocs 已在俄勒冈州、北弗吉尼亚州、悉尼、新加坡和爱尔兰地区推出。

要对 Amazon Kendra WorkDocs 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

Amazon Kendra WorkDocs 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 更改日志

先决条件

在使用索 Amazon Kendra 引 WorkDocs 数据源之前，请在 WorkDocs 和 AWS 帐户中进行这些更改。

在中 WorkDocs，请确保你有：

- 已记下 Amazon WorkDocs 仓库的 Amazon WorkDocs 目录 ID（组织 ID）。
- 已选中每个文档在您计划用于同一索引的其他数据源中 WorkDocs 以及其他数据源中都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在您的 AWS 账户中，请确保您有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

如果您没有现有 IAM 角色，则可以在将 WorkDocs 数据源连接到时使用控制台创建新 IAM 角色 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色的 ARN 和索引 ID。

连接说明

要 Amazon Kendra 连接到您的 WorkDocs 数据源，您必须提供 WorkDocs 数据源的必要详细信息，以便 Amazon Kendra 能够访问您的数据。如果您尚未进行配置 Amazon Kendra， WorkDocs 请参阅先决条件。

Console

要连接 Amazon Kendra 到 Amazon WorkDocs

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择WorkDocs 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的WorkDocs 连接器。

5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 特定于您的 Amazon WorkDocs 站点的组织 ID-选择要编制索引的 Amazon WorkDocs 网站的 ID。您必须已创建一个站点。

   b. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

   > **Note**
   >
   > IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

   c. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 爬取文档备注 - 要爬取的 Amazon WorkDocs 实体或内容类型。

   b. 使用更改日志-选择仅使用新的或修改过的内容更新索引，而不是同步所有文件。

   c. 正则表达式模式 - 包含或排除某些文件的正则表达式模式。

   d. 在 F requency 的同步运行计划中-选择同步数据源内容和更新索引的频率。

   e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

   a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

   b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

   c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Amazon WorkDocs

您必须使用 WorkDocsConfigurationAPI 指定以下内容：

- Amazon WorkDocs 目录 ID-指定 Amazon WorkDocs 目录的组织 ID。您可以依次转到 Active Directory、目录，来查找 AWS Directory Service 中的组织 ID。
- IAM 角色-指定`RoleArn`何时调用`CreateDataSource`为 IAM 角色提供访问 WorkDocs目录和调用连接器所需的公共 API 的 WorkDocs 权限。 Amazon Kendra有关更多信息，请参阅 WorkDocs 数据源的 IAM 角色。

您还可以添加以下可选功能：

- 更改日志-是否 Amazon Kendra 应使用 WorkDocs 数据源更改日志机制来确定是否必须更新索引中的文档。

  > **Note**
  >
  > 如果您不想让 Amazon Kendra 扫描所有文档，请使用更改日志。如果您的更改日志很大，则扫描 WorkDocs 数据源中的文档所花费的时间可能比处理更改日志所需的时间 Amazon Kendra 少。如果您是首次将 WorkDocs数据源与索引同步，则会扫描所有文档。

- 包含和排除筛选条件 - 指定是包括还是排除某些文档和文档注释。每条注释都会作为单独的文档来编制索引。

  > **Note**
  >
  > 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

- 字段映射-选择将 WorkDocs 数据源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

了解更多信息

要了解有关 Amazon Kendra 与 WorkDocs 数据源集成的更多信息，请参阅：

- 开始使用 A Amazon Kendra mazon WorkDocs 连接器

Box

Box 是一项提供文件托管功能的云存储服务。您可以使用 Amazon Kendra 索引 Box 内容中的内容，包括评论、任务和网络链接。

您可以使用Amazon Kendra 控制台和 BoxConfigurationAPI 连接 Amazon Kendra 到 Box 数据源。

要对 Amazon Kendra Box 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

Amazon Kendra Box 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件

- 更改日志、完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Box 数据源之前，请在您的 Box 和 AWS 账户中进行这些更改。

在 Box 中，请确保：

- 一个 Box Enterprise 或 Box Enterprise Plus 账户
- 在 Box 开发者控制台中配置了 Box 自定义应用程序，并使用 JSON Web 令牌 (JWT) 进行服务器端身份验证。有关更多详细信息，请参阅有关创建自定义应用程序的 Box 文档和 配置 JWT 身份验证的 Box 文档。
- 将您的应用程序访问权限级别设置为 应用程序 + 企业版访问程序，并允许它使用 as-user 标头进行 API 调用。
- 使用管理员用户在您的 Box 应用程序中添加以下应用程序范围：
  - 写入存储在 Box 中的所有文件和文件夹
  - 管理用户
  - 管理组
  - 管理企业版属性
- 已配置的公共/私有密钥对，包括客户端 ID、客户机密、公钥 ID、私钥 ID、密码和企业 ID，用作身份验证凭证。有关更多详细信息，请参阅公钥和私钥 pair。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 已从 Box 开发者控制台设置或 Box 应用程序中复制您的 Box 企业 ID。例如，*801234567*。
- 在 Box 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Box 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Box 数据源连接至时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要连接 Amazon Kendra 到 Box 数据源，您必须提供 Box 数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为 Box 配置 Amazon Kendra，请参阅先决条件。

Console

连接 Amazon Kendra 到 Box

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。

4. 在添加数据源页面上，选择 Box 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Box 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. Box 企业版 ID - 输入您的 Box 企业版 ID。例如，*801234567*。
   b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
   c. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Box 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。
      i. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Box-” 会自动添加到您的密钥名称中。
      ii. 对于 “客户端 ID”、“客户机密”、“公钥 ID”、“私钥 ID” 和 “密码短语”，请输入您在框中配置的公钥/私钥中的值。
      iii. 添加并保存您的密钥。
   d. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。
   e. 身份搜寻器-指定是否开启 Amazon Kendra身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。
   f. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

g. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. Box 文件夹 ID-输入要抓取的某些 Box 文件夹 ID，否则会抓取所有文件夹中的内容。

   b. Box files-选择是否抓取 Web 链接、评论和任务。

   c. 对于其他配置-添加正则表达式模式以包含或排除某些内容。

   d. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

      - 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
      - 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
      - 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

   e. 在 F requency 的同步运行计划中-选择同步数据源内容和更新索引的频率。

   f. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

   a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

   b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

   c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接 Amazon Kendra 到 Box

您必须使用 BoxConfigurationAPI 指定以下内容：

Box 企业版 ID - 输入您的 Box 企业版 ID。您可以在 Box 开发者控制台设置中或在 Box 中配置应用程序时找到企业 ID。

- 亚马逊秘密资源名称 (ARN)-提供包含您的 Box 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "clientID": "client-id",
    "clientSecret": "client-secret",
    "publicKeyID": "public-key-id",
    "privateKey": "private-key",
    "passphrase": "pass-phrase"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Box 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Box 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 在数据来源配置中指定 VpcConfiguration。请参阅配置 Amazon Kendra 以使用 VPC。
- 更改日志-是否 Amazon Kendra 应使用 Box 数据源更改日志机制来确定是否必须更新索引中的文档。

> **Note**
>
> 如果您不想让 Amazon Kendra 扫描所有文档，请使用更改日志。如果您的更改日志很大，则扫描 Box 数据源中的文档所花费的时间可能比处理更改日志所需的时间 Amazon Kendra 少。如果您是首次将 Box 数据来源与索引同步，则会扫描所有文档。

- 评论、任务、Web 链接-指定是否抓取这些类型的内容。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件

不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 包含和排除筛选器-指定是包含还是排除某些 Box 文件和文件夹。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Box 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

了解更多信息

要了解有关 Amazon Kendra 与 Box 数据源集成的更多信息，请参阅：

- Amazon Kendra Box 连接器入门

Confluence

Confluence 是一款协作式工作管理工具，专为共享、存储和处理项目规划、软件开发和产品管理而设计。您可以使用 Amazon Kendra 索引 Confluence 空间、页面（包括嵌套页面）、博客以及已编入索引的页面和博客的评论和附件。

Amazon Kendra 同时支持 Confluence 服务器/数据中心和 Confluence 云。

> **Note**
>
> 默认情况下，Amazon Kendra 不索引 Confluence 档案和个人空间。在创建数据来源时，您可以选择为它们编制索引。如果您不想 Amazon Kendra 为空间编制索引，请在 Confluence 中将其标记为私有。

您可以使用Amazon Kendra 控制台、API 或 TemplateConfigurationAPI Amazon Kendra 连接到您的 Confluence 数据源。ConfluenceConfiguration

Amazon Kendra 有两个版本的 Confluence 连接器。每个版本支持的功能包括：

Confluence 连接器 V1.0/API ConfluenceConfiguration

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- （仅适用于 Confluence Server）虚拟私有云（VPC）

Confluence 连接器 V2.0/API TemplateConfiguration

- 字段映射
- 用户访问控制
- 包含/排除模式
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

> **Note**
>
> 对 Confluence 连接器 V1.0/ ConfluenceConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 Confluence 连接器 V2.0/API。 TemplateConfiguration

要对 Amazon Kendra Confluence 数据源连接器进行故障排除，请参阅。数据来源故障排除

主题

- Confluence 连接器 V1.0
- Confluence 连接器 V2.0

Confluence 连接器 V1.0

Confluence 是一款协作式工作管理工具，专为共享、存储和处理项目规划、软件开发和产品管理而设计。您可以使用 Amazon Kendra 为 Confluence 空间、页面（包括嵌套页面）、博客以及已编入索引的页面和博客的评论和附件编制索引。

> **Note**
>
> 对 Confluence 连接器 V1.0/ ConfluenceConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 Confluence 连接器 V2.0/API。 TemplateConfiguration

要对 Amazon Kendra Confluence 数据源连接器进行故障排除，请参阅。数据来源故障排除

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

Amazon Kendra Confluence 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- （仅适用于 Confluence Server）虚拟私有云（VPC）

先决条件

在使用 Amazon Kendra 索引您的 Confluence 数据源之前，请先在您的 Confluence 和账户中进行这些更改。 AWS

在 Confluence 中，请确保：

- 通过以下 Amazon Kendra 方式授予查看您的 Confluence 实例中所有内容的权限：
  - Amazon Kendra 成为`confluence-administrators`群组成员。
  - 授予所有现有空间、博客和页面的站点管理员权限。
- 复制 Confluence 实例的 URL。
- 对于 SSO（单点登录）用户：在 Confluence 数据中心配置 Confluence 身份验证方法时，激活了显示登录页面以输入用户名和密码。
- 对于 Confluence Server
  - 记下了您的基本身份验证凭证，其中包含要连接到 Amazon Kendra的 Confluence 管理账户用户名和密码。

    > **Note**
    >
    > 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

  - 可选：在您的 Confluence 账户中生成了要连接的个人访问令牌。 Amazon Kendra有关更多信息，请参阅有关生成个人访问令牌的 Confluence 文档。
- 对于 Confluence Cloud
  - 记下了您的基本身份验证凭证，其中包含要连接到 Amazon Kendra的 Confluence 管理账户用户名和密码。
- 在 Confluence 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Confluence 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Confluence 数据源连接至时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。 Amazon Kendra如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到您的 Confluence 数据源，您必须提供您的 Confluence 凭据的详细信息，Amazon Kendra 以便访问您的数据。如果您尚未配置 Confluence，请参阅。 Amazon Kendra 先决条件

Console

连接到 Con Amazon Kendra fluence

1. 登录 AWS 管理控制台并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

   > **Note**
   >
   > 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据来源页面上，选择 Confluence 连接器 V1.0，然后选择添加数据来源。

5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 在 Confluence 云和 Confluenc e 服务器之间进行选择。

   b. 如果您选择 Confluence Cloud，请输入以下信息：

      i. Confluence URL - 您的 Confluence URL。

      ii. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Confluence 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

         • 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

            I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Confluence-” 会自动添加到您的密钥名称中。

            II. 对于用户名和密码-输入您的 Confluence 用户名和密码。

            III. 选择保存身份验证。

   c. 如果您选择 Confluence 服务器，请输入以下信息：

      i. Confluence URL - 您的 Confluence 用户名和密码。

      ii. （可选）对于 Web 代理 - 请输入以下信息：

         A. 主机名 - 您的 Confluence 账户的主机名。

         B. 端口号 - 主机 URL 传输协议使用的端口。

      iii. 对于身份验证，请选择基本身份验证或（仅限 Confluence 服务器）个人访问令牌。

      iv. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Confluence 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

- 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

  I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Confluence-” 会自动添加到您的密钥名称中。

  II. 对于用户名和密码-输入您在 Confluence 中配置的身份验证凭据值。如果使用基本身份验证，请使用您的 Confluence 用户名（电子邮件 ID）和密码（API 令牌）。如果使用个人访问令牌，请输入您在 Confluence 账户中配置的个人访问令牌的详细信息。

  III. 保存并添加您的密钥。

d. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

e. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 对于包括个人空间和包括存档空间 - 选择要包含在此数据来源中的可选空间类型。

   b. 对于其他配置 - 指定正则表达式模式以包含或排除某些文件。最多可以添加 100 个模式。

   c. 您还可以选择在所选空间内爬取附件。

   d. 在 “同步运行计划” 中，“频率”-选择与数据源同步的频率。 Amazon Kendra

   e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

   a. 对于 Spac e、Page、Blog-从 Amazon Kendra 生成的默认数据源字段或其他建议的字段映射中选择以添加索引字段。

   b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

   c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接到 Con Amazon Kendra fluence

您必须使用 ConfluenceConfigurationAPI 指定以下内容：

- Confluence 版本 - 指定您用作 CLOUD 或 SERVER 的 Confluence 实例的版本。
- 亚马逊秘密资源名称 (ARN)-提供包含您的 Confluence 身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。

  如果您使用 Confluence 服务器，则可以使用您的 Confluence 用户名和密码或您的个人访问令牌作为身份验证凭据。

  如果您使用 Confluence 用户名和密码作为身份验证凭据，则可以将以下凭据作为 JSON 结构存储在密钥 Secrets Manager 中：

  ```
  {
      "username": "user name",
      "password": "password"
  }
  ```

  如果您使用个人访问令牌连接 Confluence Server Amazon Kendra，则可以将以下凭据作为 JSON 结构存储在您的 Secrets Manager 密钥中：

  ```
  {
      "patToken": "personal access token"
  }
  ```

  如果你使用 Confluence Cloud，则使用你的 Confluence 用户名和在 Confluence 中配置的 API 令牌作为密码。您将以下凭证作为 JSON 结构存储在您的 Secrets Manager 密钥中：

  ```
  {
      "username": "user name",
      "password": "API token"
  }
  ```

- IAM 角色 —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Confluence 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Confluence 数据来源的IAM 角色。

您还可以添加以下可选功能：

- Web 代理 - 是否通过 Web 代理连接到您的 Confluence URL 实例。您可以将此选项用于 Confluence Server。
- （仅适用于 Confluence Server）虚拟私有云（VPC）- 在数据来源配置中指定 `VpcConfiguration`。请参阅配置 Amazon Kendra 以使用 VPC。
- 包含和排除筛选条件 - 指定用于包含或排除特定空间、博客文章、页面、空间和附件的正则表达式。如果您选择为附件编制索引，则仅对已编入索引的页面和博客的附件编制索引。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 字段映射 - 选择将 Confluence 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称`_document_body`。其他所有字段均为可选字段。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

了解更多信息

要了解有关 Amazon Kendra 与 Confluence 数据源集成的更多信息，请参阅：

- 配置您的 Amazon Kendra Confluence 服务器连接器

Confluence 连接器 V2.0

Confluence 是一款协作式工作管理工具，专为共享、存储和处理项目规划、软件开发和产品管理而设计。您可以使用 Amazon Kendra 为 Confluence 空间、页面（包括嵌套页面）、博客以及已编入索引的页面和博客的评论和附件编制索引。

要对 Amazon Kendra Confluence 数据源连接器进行故障排除，请参阅。数据来源故障排除

主题

支持的特征

Amazon Kendra Confluence 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除模式
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引您的 Confluence 数据源之前，请先在您的 Confluence 和账户中进行这些更改。 AWS

在 Confluence 中，请确保：

- 已复制 Confluence 实例的 URL。例如：*https://example.confluence.com*、*https://www.example.confluence.com/* 或 *https://atlassian.net/*。您需要 Confluence 实例 URL 才能连接到 Amazon Kendra。

  ```
  ####### Confluence Cloud########## atlassian.net/ ###
  ```

> **Note**
>
> 不支持以下 URL 格式：
>
> - *https://example.confluence.com/xyz*
> - *https://www.example.confluence.com//wiki/spacekey/xxx*
> - *https://atlassian.net/xyz*

> **Note**
>
> （本地/服务器）Amazon Kendra 会检查中 AWS Secrets Manager 包含的端点信息是否与数据源配置详细信息中指定的端点信息相同。这有助于防止出现混淆代理人问题，这是一个安全问题，即用户无权执行操作，但可以将 Amazon Kendra 作为代理来访问配置的密钥和执行操作。如果以后更改端点信息，则必须创建一个新密钥来同步此信息。

- 配置了包含用户名（用于登录 Confluence 的电子邮件 ID）和密码（Confluence API 令牌作为密码）的基本身份验证凭证。请参阅管理您的 Atlassian 账户的 API 令牌。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 可选：已配置包含 Confluence 应用程序密钥、Confluence 应用程序密钥、Confluence 访问令牌和 Confluence 刷新令牌的 OAuth 2.0 凭据，允许连接到你的 Confluence 实例。 Amazon Kendra 如果您的访问令牌过期，则可以使用刷新令牌重新生成访问令牌和刷新令牌对。或者，您可以重复授权过程。有关访问令牌的更多信息，请参阅管理 OAuth 访问令牌。
- （仅适用于 Confluence 服务器/数据中心）可选：在 Confluence 中配置了个人访问令牌 (PAT)。请参阅使用个人访问令牌。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Confluence 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Confluence 数据源连接至时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。 Amazon Kendra如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到您的 Confluence 数据源，您必须提供 Confluence 数据源的必要详细信息，Amazon Kendra 以便访问您的数据。如果您还没有为 Amazon Kendra 配置 Confluence，请参阅先决条件。

Console

连接到 Con Amazon Kendra fluence

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。

4. 在添加数据源页面上，选择 Confluence 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Confluence 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 在源中，选择 Confluence 云或 Confluen ce 服务器/数据中心。
   b. Confluence URL - 输入 Confluence 主机 URL。例如，*https://example.confluence.com*。
   c. （仅适用于 Confluence 服务器/数据中心）SSL 证书位置-可选- 输入 Confluence 服务器的 SSL 证书文件的 Amazon S3 路径。
   d. （仅适用于 Confluence 服务器/数据中心）Web 代理-可选- 输入 Web 代理主机名（不带`http://`或`https://`协议）和端口号（主机 URL 传输协议使用的端口）。端口号应为介于 0 到 65535 之间的数值。
   e. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
   f. 身份验证 -选择基本身份验证、Oauth 2.0 身份验证或（仅适用于 Confluence 服务器/数据中心）个人访问令牌身份验证。
   g. AWS Secrets Manager 密钥 - 选择现有密钥或创建新的 密钥来存储您的 Confluence 身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。在窗口中输入以下信息：
      i. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Confluence-” 会自动添加到您的密钥名称中。
      ii. 如果使用基本身份验证，请输入您在 Confluence 中配置的机密名称、用户名和密码（Confluence API 令牌作为密码）。

如果使用 OAuth2.0 身份验证，请输入您在 Confluence 中配置的密钥名称、应用程序密钥、应用程序密钥、访问令牌和刷新令牌。

（仅限 Confluence 服务器/数据中心）如果使用个人访问令牌身份验证，请输入您在 Confluence 中配置的密钥名称和 Confluence 令牌。

iii. 保存并添加您的密钥。

h. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

i. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

j. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

k. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围内，用于同步内容-选择同步以下内容类型：页面、页面评论、页面附件、博客、博客评论、博客附件、个人空间和存档空间。

> **Note**
>
> 只有当您选择同步页面时，才能选择页面评论和页面附件。只有选择同步博客时，才能选择博客评论和博客附件。

> **Important**
>
> 如果您未在 “其他配置” 中指定空格键正则表达式模式，则默认情况下将抓取所有页面和博客。

b. 在其他配置中，对于最大文件大小-指定要抓取的文件大小限制（以 MB Amazon Kendra 为单位）。 Amazon Kendra 将仅抓取您定义的大小限制内的文件。默认文件大小为 50 MB。最大文件大小应大于 0 MB 且小于或等于 50 MB。

对于 Spaces 正则表达式模式-使用以下方法指定是在索引中包含还是排除特定空格：

- 空格键（例如，*my-space-123*）

> **Note**
>
> 如果您未指定空格键正则表达式模式，则默认情况下将抓取所有页面和博客。

- 网址（例如，*# */ MySiteMyDocuments/)*
- 文件类型（例如，*.*\ .pdf#.*\ .t* xt）

对于实体标题正则表达式模式-指定正则表达式模式以按标题包含或排除某些博客、页面、评论和附件。

> **Note**
>
> 如果要包含或排除对特定页面或子页面的抓取，则可以使用页面标题正则表达式模式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率-选择同步数据源内容和更新索引的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。要添加自定义数据来源字段，请创建要映射到的索引字段名称和字段数据类型。

b. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接到 Con Amazon Kendra fluence

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构CONFLUENCEV2时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 主机 URL-指定 Confluence 主机 URL 实例。例如，*https://example.confluence.com*。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 身份验证类型-指定身份验证的类型，是否为、BasicOAuth2、（仅限 Confluence 服务器）。Personal-token
- （可选，仅限 Confluence Server）SSL 证书位置 - 指定用于存储 SSL 证书的 S3bucketName 和 s3certificateName。
- 亚马逊秘密资源名称 (ARN)-提供包含您在 Confluence 中配置的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。如果您使用基本身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
```

```
    "username": "email ID or user name",
    "password": "Confluence API token"
}
```

如果您使用 OAuth 2.0 身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "confluenceAppKey": "app key",
    "confluenceAppSecret": "app secret",
    "confluenceAccessToken": "access token",
    "confluenceRefreshToken": "refresh token"
}
```

（仅限 Confluence Server）如果您使用基本身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "hostUrl": "Confluence Server host URL",
    "username": "Confluence Server user name",
    "password": "Confluence Server password"
}
```

（仅限 Confluence Server）如果您使用个人访问令牌身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "hostUrl": "Confluence Server host URL",
    "patToken": "personal access token"
}
```

- IAM 角色 —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Confluence 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Confluence 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 `VpcConfiguration`，以便调用 `CreateDataSource`。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 文件大小-指定要抓取的最大文件大小。

- 文档/内容类型-指定是否抓取页面、页面评论、页面附件、博客、博客评论、博客附件、空间和存档空间。
- 包含和排除过滤器-指定是包含还是排除某些空间、页面、博客及其评论和附件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 网络代理-如果您想通过网络代理连接到 Confluence URL 实例，请指定您的网络代理信息。您可以将此选项用于 Confluence Server。
- 访问控制列表 (ACL)-如果您有 ACL 并希望将其用于访问控制，则指定是否要搜索文档的 ACL 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。
- 字段映射 - 选择将 Confluence 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅 Confluence 模板架构。

注意

- 个人访问令牌（PAT）不适用于 Confluence Cloud。

自定义数据来源连接器

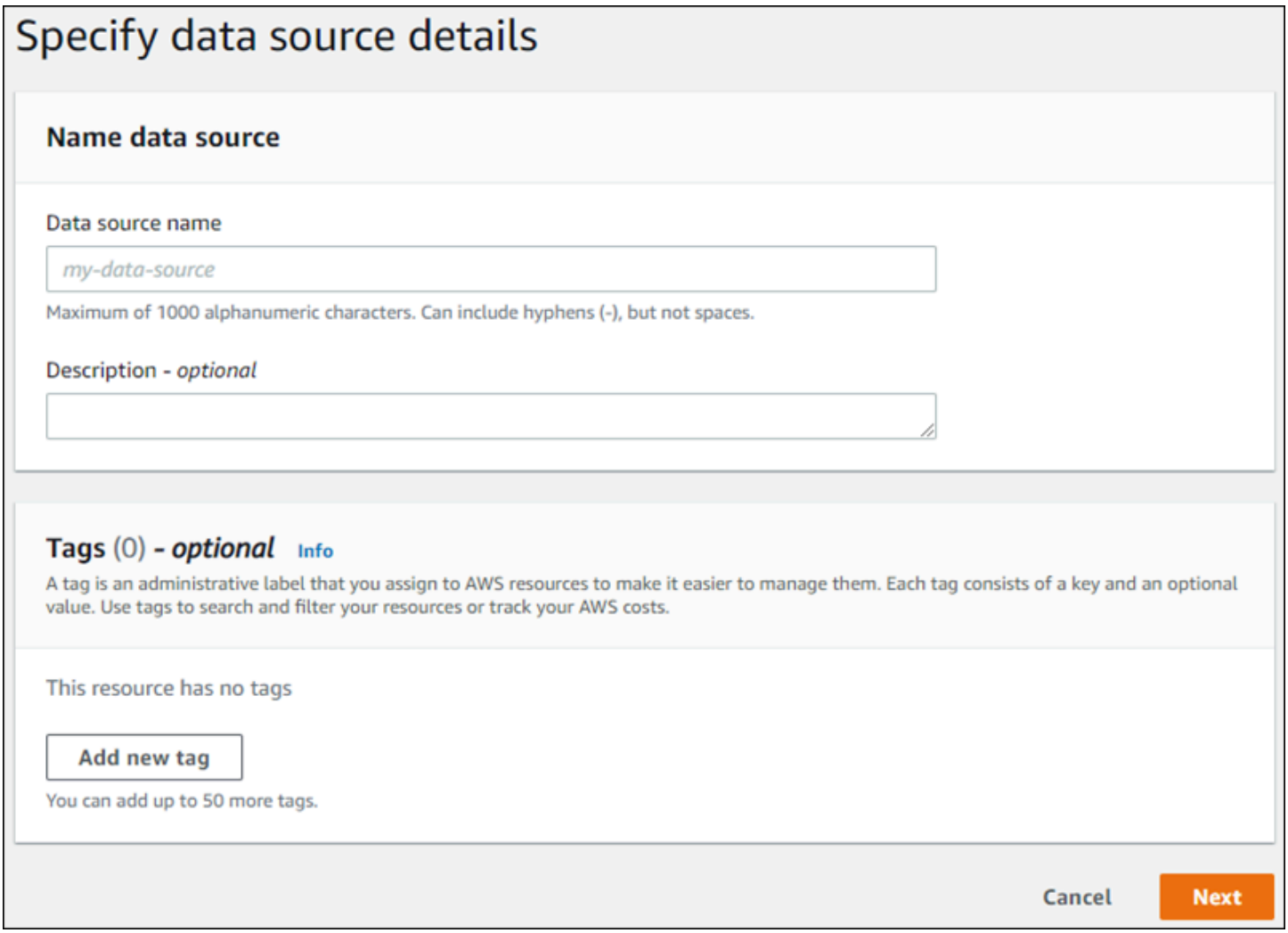
如果您的存储库尚未为其提供数据源连接器， Amazon Kendra 请使用自定义数据源。即使您无法使用 Amazon Kendra 数据源同步存储库，也可以使用 Amazon Kendra它来查看数据源提供的相同运行历史指标。使用它可以在 Amazon Kendra 数据源和自定义数据源之间创建一致的同步监控体验。具体而言，使用自定义数据源来查看您使用 Document and D BatchPutoc ument API 创建的数据源连接器的同步指标。BatchDelete

要对 Amazon Kendra 数据来源连接器进行故障排除，请参阅数据来源故障排除。

创建自定义数据源时，您可以完全控制要编制索引的文档的选择方式。 Amazon Kendra 仅提供可用于监控数据源同步作业的指标信息。您必须创建并运行用于确定数据来源索引文档的爬网程序。

您必须使用 Document 对象和`_source_uri`来指定文档的主标题，才能`DocumentTitle``DocumentURI`包含在`Query`结果的响应中。DocumentAttribute

您可以使用控制台或源 API 为自定义数据CreateData源创建标识符。要使用控制台，请为您的数据来源命名，并可选择提供描述和资源标签。创建数据来源后，将显示数据来源 ID。复制此 ID，以便在将数据来源与索引同步时使用。



您还可以使用 `CreateDataSource` API 创建自定义数据来源。API 会返回一个 ID，供您在同步数据来源时使用。使用 `CreateDataSource` API 创建自定义数据来源时，无法设置 `Configuration`、`RoleArn` 或 `Schedule` 参数。如果您设置了这些参数，则 Amazon Kendra 会返回`ValidationException`异常。

要使用自定义数据来源，请创建一个负责更新 Amazon Kendra 索引的应用程序。该应用程序依赖于您创建的爬网程序。爬网程序会读取存储库中的文档并确定应将哪些文档发送到 Amazon Kendra。您的应用程序应执行以下步骤：

1. 爬取您的存储库，列出存储库中添加、更新或删除的文档。
2. 调用 StartDataSourceSyncJob API 以表示同步作业正在启动。您需要提供一个数据源 ID 来标识正在同步的数据源。 Amazon Kendra 返回用于标识特定同步任务的执行 ID。
3. 调用BatchDelete文档 API 从索引中移除文档。您可以提供数据来源 ID 和执行 ID 来标识正在同步的数据来源以及与此更新关联的作业。

4. 调用 StopDataSourceSyncJob API 以发出同步任务结束的信号。调用 StopDataSourceSyncJob API 后，关联的执行 ID 不再有效。
5. 使用索引和数据源标识符调用 J ListDataSourceSyncob s API，列出数据源的同步任务并查看同步作业的指标。

结束同步作业后，您可以开始新的同步作业。可能需要一段时间才能将所有提交的文档添加到索引中。使用 ListDataSourceSyncJobs API 查看同步任务的状态。如果同步作业返回的 Status 为 SYNCING_INDEXING，则某些文档仍在编制索引。当上一个任务的状态为FAILED或时，您可以开始新的同步作业SUCCEEDED。

调用 StopDataSourceSyncJob API 后，您不能在调用 BatchPutDocument 或 BatchDeleteDocument API 时使用同步任务标识符。如果您这样做，则提交的所有文档都将在 API 的 FailedDocuments 响应消息中返回。

必需的属性

当您 Amazon Kendra 使用 BatchPutDocument API 向提交文档时，每个文档都需要两个属性来标识其所属的数据源和同步运行。要将自定义数据来源的文档正确映射到 Amazon Kendra 索引，您必须提供以下两个属性：

- _data_source_id - 数据来源的标识符。当您使用控制台或 CreateDataSource API 创建数据来源时，会返回此值。
- _data_source_sync_job_execution_id - 同步运行的标识符。当您开始与 StartDataSourceSyncJob API 进行索引同步时，会返回该值。

以下是使用自定义数据来源为文档编制索引所需的 JSON。

```
{
    "Documents": [
        {
            "Attributes": [
                {
                    "Key": "_data_source_id",
                    "Value": {
                        "StringValue": "data source identifier"
                    }
                },
                {
                    "Key": "_data_source_sync_job_execution_id",
```

```
                    "Value": {
                        "StringValue": "sync job identifier"
                    }
                }
            ],
            "Blob": "document content",
            "ContentType": "content type",
            "Id": "document identifier",
            "Title": "document title"
        }
    ],
    "IndexId": "index identifier",
    "RoleArn": "IAM role ARN"
}
```

使用 BatchDeleteDocument API 从索引中移除文档时，需要在 DataSourceSyncJobMetricTarget 参数中指定以下两个字段：

- DataSourceId - 数据来源的标识符。当您使用控制台或 CreateDataSource API 创建数据来源时，会返回此值。
- DataSourceSyncJobId - 同步运行的标识符。当您开始与 StartDataSourceSyncJob API 进行索引同步时，会返回该值。

以下是使用 BatchDeleteDocument API 从索引中删除文档所需的 JSON。

```
{
    "DataSourceSyncJobMetricTarget": {
        "DataSourceId": "data source identifier",
        "DataSourceSyncJobId": "sync job identifier"
    },
    "DocumentIdList": [
        "document identifier"
    ],
    "IndexId": "index identifier"
}
```

查看 指标

同步任务完成后，你可以使用 M DataSourceSyncJobetrics API 来获取与同步任务相关的指标。使用它来监控您的自定义数据来源同步。

如果您多次提交同一个文档，无论是作为 BatchPutDocument API、BatchDeleteDocument API 的一部分，还是为添加和删除提交该文档，该文档在指标中仅计入一次。

- DocumentsAdded - 使用与首次添加到索引中的此同步作业关联的 BatchPutDocument API 提交的文档数量。如果在同步中为添加多次提交文档，则该文档在指标中仅计入一次。
- DocumentsDeleted - 使用与从索引中删除的此同步作业关联的 BatchDeleteDocument API 提交的文档数量。如果在同步中为删除多次提交文档，则该文档在指标中仅计入一次。
- DocumentsFailed - 与该同步作业关联但编制索引失败的文档数量。 Amazon Kendra 已接受为这些文档编制索引，但无法编制索引或已删除。如果文档未被接受 Amazon Kendra，则该文档的标识符将在BatchPutDocument和 BatchDeleteDocument API 的FailedDocuments响应属性中返回。
- DocumentsModified— 使用与此同步作业关联的 BatchPutDocument API 提交的已修改文档的数量，这些文档已在 Amazon Kendra 索引中进行了修改。

Amazon Kendra 在为文档编制索引时也会发出 Amazon CloudWatch 指标。有关更多信息，请参阅Amazon Kendra 使用进行监控 Amazon CloudWatch。

Amazon Kendra 不返回自定义数据源的DocumentsScanned指标。它还会 CloudWatch 发出文档中列出的Amazon Kendra 数据源指标。

了解更多信息

要了解有关 Amazon Kendra 与自定义数据源集成的更多信息，请参阅：

- 将自定义数据源添加到 Amazon Kendra

自定义数据来源（Java）

以下代码提供了使用 Java 实现自定义数据来源的示例。该程序首先创建自定义数据来源，然后将新添加至索引的文档与自定义数据来源同步。

以下代码演示了如何创建和使用自定义数据来源。当您在应用程序中使用自定义数据来源时，无需在每次将索引与数据来源同步时都创建新的数据来源（一次性流程）。您可以使用索引 ID 和数据来源 ID 来同步数据。

```
package com.amazonaws.kendra;

import java.util.concurrent.TimeUnit;
import software.amazon.awssdk.services.kendra.KendraClient;
```

```
import csoftware.amazon.awssdk.services.kendra.model.BatchPutDocumentRequest;
import csoftware.amazon.awssdk.services.kendra.model.BatchPutDocumentResponse;
import software.amazon.awssdk.services.kendra.model.CreateDataSourceRequest;
import software.amazon.awssdk.services.kendra.model.CreateDataSourceResponse;
import software.amazon.awssdk.services.kendra.model.DataSourceType;
import software.amazon.awssdk.services.kendra.model.Document;
import software.amazon.awssdk.services.kendra.model.ListDataSourceSyncJobsRequest;
import software.amazon.awssdk.services.kendra.model.ListDataSourceSyncJobsResponse;
import software.amazon.awssdk.services.kendra.model.StartDataSourceSyncJobRequest;
import software.amazon.awssdk.services.kendra.model.StartDataSourceSyncJobResponse;
import software.amazon.awssdk.services.kendra.model.StopDataSourceSyncJobRequest;
import software.amazon.awssdk.services.kendra.model.StopDataSourceSyncJobResponse;

public class SampleSyncForCustomDataSource {
  public static void main(String[] args) {
    KendraClient kendra = KendraClient.builder().build();

    String myIndexId = "yourIndexId";
    String dataSourceName = "custom data source";
    String dataSourceDescription = "Amazon Kendra custom data source connector"

    // Create custom data source
    CreateDataSourceRequest createDataSourceRequest = CreateDataSourceRequest
        .builder()
        .indexId(myIndexId)
        .name(dataSourceName)
        .description(dataSourceDescription)
        .type(DataSourceType.CUSTOM)
        .build();

    CreateDataSourceResponse createDataSourceResponse =
 kendra.createDataSource(createDataSourceRequest);
    System.out.println(String.format("Response of creating data source: %s",
 createDataSourceResponse));

    // Get the data source ID from createDataSourceResponse
    String dataSourceId = createDataSourceResponse.Id();

    // Wait for the custom data source to become active
    System.out.println(String.format("Waiting for Amazon Kendra to create the data
 source %s", dataSourceId));
    // You can use the DescribeDataSource API to check the status
    DescribeDataSourceRequest describeDataSourceRequest = DescribeDataSourceRequest
        .builder()
```

```
        .indexId(myIndexId)
        .id(dataSourceId)
        .build();

   while (true) {
        DescribeDataSourceResponse describeDataSourceResponse =
kendra.describeDataSource(describeDataSourceRequest);

        DataSourceStatus status = describeDataSourceResponse.status();
        System.out.println(String.format("Creating data source. Status: %s", status));
        if (status != DataSourceStatus.CREATING) {
            break;
        }

        TimeUnit.SECONDS.sleep(60);
   }

   // Start syncing yor data source by calling StartDataSourceSyncJob and providing
your index ID
   // and your custom data source ID
   System.out.println(String.format("Synchronize the data source %s", dataSourceId));
   StartDataSourceSyncJobRequest startDataSourceSyncJobRequest =
StartDataSourceSyncJobRequest
        .builder()
        .indexId(myIndexId)
        .id(dataSourceId)
        .build();
   StartDataSourceSyncJobResponse startDataSourceSyncJobResponse =
kendra.startDataSourceSyncJob(startDataSourceSyncJobRequest);

   // Get the  sync job execution ID from startDataSourceSyncJobResponse
   String executionId = startDataSourceSyncJobResponse.ExecutionId();
System.out.println(String.format("Waiting for the data source to sync with the index
%s for execution ID %s", indexId, startDataSourceSyncJobResponse.executionId()));

   // Add 2 documents uploaded to S3 bucket to your index using the BatchPutDocument
API
   // The added documents should sync with your custom data source
   Document pollyDoc = Document
        .builder()
        .s3Path(
            S3Path.builder()
            .bucket("s3-test-bucket")
            .key("what_is_Amazon_Polly.docx")
```

```
            .build())
        .title("What is Amazon Polly?")
        .id("polly_doc_1")
        .build();

    Document rekognitionDoc = Document
        .builder()
        .s3Path(
            S3Path.builder()
            .bucket("s3-test-bucket")
            .key("what_is_amazon_rekognition.docx")
            .build())
        .title("What is Amazon rekognition?")
        .id("rekognition_doc_1")
        .build();

    BatchPutDocumentRequest batchPutDocumentRequest = BatchPutDocumentRequest
        .builder()
        .indexId(myIndexId)
        .documents(pollyDoc, rekognitionDoc)
        .build();

    BatchPutDocumentResponse result = kendra.batchPutDocument(batchPutDocumentRequest);
    System.out.println(String.format("BatchPutDocument result: %s", result));

    // Once custom data source synced, stop the sync job using the
StopDataSourceSyncJob API
    StopDataSourceSyncJobResponse stopDataSourceSyncJobResponse =
kendra.stopDataSourceSyncJob(
        StopDataSourceSyncJobRequest()
            .indexId(myIndexId)
            .id(dataSourceId)
    );

// List your sync jobs
    ListDataSourceSyncJobsRequest listDataSourceSyncJobsRequest =
ListDataSourceSyncJobsRequest
        .builder()
        .indexId(myIndexId)
        .id(dataSourceId)
        .build();

    while (true) {
```

```
        ListDataSourceSyncJobsResponse listDataSourceSyncJobsResponse =
 kendra.listDataSourceSyncJobs(listDataSourceSyncJobsRequest);
        DataSourceSyncJob job = listDataSourceSyncJobsResponse.history().get(0);
        System.out.println(String.format("Status: %s", job.status()));
    }
  }
}
```

Dropbox

Dropbox 是一项文件托管服务，提供云存储、文档整理和文档模板服务。如果您是 Dropbox 用户，则可以使用 Amazon Kendra 索引您的 Dropbox 文件、Dropbox Paper、Dropbox Paper 模板和存储的网页快捷方式。您还可以配置 Amazon Kendra 为索引特定的 Dropbox 文件、Dropbox Paper、Dropbox Paper 模板和存储的网页快捷方式。

Amazon Kendra 同时支持 Dropbox Business 版 Dropbox 和

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的 Dropbox 数据源。

要对 Amazon Kendra Dropbox 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

Amazon Kendra Dropbox 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Dropbox 数据源之前，请在您的 Dropbox 和 AWS 帐户中进行这些更改。

在 Dropbox 中，请确保：

- 已创建 Dropbox Advanced 账户并设置了管理员用户。
- 为一个 Dropbox 应用配置了唯一的应用程序名称，激活了 Scoped Acces s。请参阅有关创建应用程序的 Dropbox 文档。
- 在 Dropbox 控制台上激活了 Dropbox 的完整权限，并添加了以下权限：
  - files.content.read
  - files.metadata.read
  - sharing.read
  - file_requests.read
  - groups.read
  - team_info.read
  - team_data.content.read
- 已记下您的 Dropbox 应用程序键、Dropbox 应用程序密钥和适用于基本身份验证凭证的 Dropbox 访问令牌

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 为您的 Dropbox 应用配置并复制了临时 Oauth 2.0 访问令牌。此令牌是临时性的，将在 4 小时后过期。请参阅有关 OAuth 身份验证的 Dropbox 文档。

> **Note**
>
> 建议您创建永不过期的 Dropbox 刷新访问令牌，而不是依赖在 4 小时后过期的一次性访问令牌。刷新访问令牌是永久性的，永不过期，因此您将来可以继续同步数据来源。

- 推荐：已配置永不过期的 Dropbox 永久刷新令牌， Amazon Kendra 以便在不中断的情况下继续同步您的数据源。请参阅有关刷新令牌的 Dropbox 文档。

- 在 Dropbox 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Dropbox 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Dropbox 数据源关联到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 Dropbox 数据源，您必须提供 Dropbox 数据源的必要详细信息，这样 Amazon Kendra 他们才能访问您的数据。如果您尚未配置 Dropbox Amazon Kendra，请参阅先决条件。

Console

连接到 Drop Amazon Kendra box

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。

2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Dropbox 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Dropbox 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

   b. 身份验证令牌的类型-选择永久令牌（推荐）或临时访问令牌。

   c. AWS Secrets Manager 密钥-选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Dropbox 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

      i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

         A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Dropbox-” 会自动添加到您的密钥名称中。

         B. 对于应用程序密钥、应用程序密钥和令牌信息（永久或临时），请输入在 Dropbox 中配置的身份验证凭据值。

      ii. 保存并添加您的密钥。

d. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

e. 身份搜寻器-指定是否开启 Amazon Kendra身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

f. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

g. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 对于选择实体或内容类型-选择要抓取的 Dropbox 实体或内容类型。

b. 在其他配置中，对于正则表达式模式 - 添加正则表达式模式以包含或排除某些文件。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率-选择同步数据源内容和更新索引的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 文件、Dropbox Pap er 和 Dropbox Paper 模板-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的默认数据源字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接到 Drop Amazon Kendra box

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构DROPBOX时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅为新增、修改和删除的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 访问令牌类型-指定要对存储身份验证凭据的 AWS Secrets Manager 密钥使用永久访问令牌还是临时访问令牌。

> **Note**
>
> 建议您创建在 Dropbox 中永不过期的刷新访问令牌，而不是依赖在 4 小时后过期的一次性访问令牌。您可以在 Dropbox 开发者控制台中创建应用程序和刷新访问令牌，并在密钥中提供访问令牌。

- 亚马逊秘密资源名称 (ARN)-提供包含您的 Dropbox 帐户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "appKey": "Dropbox app key",
    "appSecret": "Dropbox app secret",
    "accesstoken": "temporary access token or refresh access token"
}
```

- 身份搜寻器-指定是否开启 Amazon Kendra身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。
- IAM 角色-指定您RoleArn何时致电CreateDataSource为 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Dropbox 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Dropbox 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 文档/内容类型-指定是否抓取 Dropbox、Dropbox Paper 文档、Dropbox Paper 模板和存储在 Dropbox 中的网页快捷方式中的文件。
- 包含和排除筛选条件 - 指定是包含还是排除文件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 访问控制列表 (ACL)-如果您有 ACL 并希望将其用于访问控制，则指定是否要搜索文档的 ACL 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Dropbox 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅 Dropbox 模板架构。

了解更多信息

要了解有关将 Amazon Kendra 与 Dropbox 数据来源集成的更多信息，请参阅：

- 使用适用于 Amazon Kendra的 Dropbox 连接器为 Dropbox 内容编制索引

Drupal

Drupal 是一个可用于创建网站和 Web 应用程序的开源内容管理系统（CMS）。你可以在 Drupal 中使用 Amazon Kendra 索引以下内容：

- 内容 - 文章、基本页面、基本数据块、用户定义的内容类型、用户定义的数据块类型、自定义内容类型、自定义数据块类型
- 评论 - 适用于任何内容类型和数据块类型
- 附件 - 适用于任何内容类型和数据块类型

您可以使用Amazon Kendra 控制台或 TemplateConfigurationAPI Amazon Kendra 连接到您的 Drupal 数据源。

要对 Amazon Kendra Drupal 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 注意

支持的特征

Amazon Kendra Drupal 数据源连接器支持以下功能：

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Drupal 数据源之前，请在您的 Drupal 和 AWS 帐户中进行这些更改。

在 Drupal 中，请确保：

- 已创建一个 Drupal（标准）套件账户和一个具有管理员角色的用户。
- 已复制 Drupal 站点名称并配置了主机 URL。例如，*https://<hostname>/<drupalsitename>*。
- 已配置包含用户名（Drupal 网站登录用户名）和密码（Drupal 网站密码）的基本身份验证凭证。
- 推荐：配置 OAuth 2.0 凭证令牌。使用此令牌以及您的 Drupal 密码授权、客户端 ID、客户端密钥、用户名（Drupal 网站登录用户名）和密码（Drupal 网站密码）连接到 Amazon Kendra。
- 使用管理员角色在您的 Drupal 账户中添加了以下权限：
  - 管理数据块
  - 挂你数据块内容显示
  - 管理数据块内容字段
  - 管理数据块内容表达显示
  - 管理视图
  - 查看用户电子邮件地址
  - 查看自己的未发布内容
  - 查看页面修订
  - 查看文章修订
  - 查看所有修订
  - 查看管理主题

- 访问内容
- 访问内容概览
- 访问注释
- 搜索内容
- 访问文件概览
- 访问上下文链接

> **Note**
>
> 如果存在用户定义的内容类型或用户定义的数据块类型，或者任何视图和区块被添加到 Drupal 网站，则必须为其提供管理员访问权限。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Drupal 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Drupal 数据源连接到 Amazon Kendra时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的Drupal数据源，您必须提供您的Drupal凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未配置 Drupal，请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 Drupal

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Drupal 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Drupal 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 在源中，对于主机 URL - 您的 Drupal 网站的主机 URL。例如，*https://<hostname>/<drupalsitename>*。
   b. SSL 证书位置 - 输入您存储在 Amazon S3 存储桶中的 SSL 证书的路径。
   c. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
   d. 对于身份验证 - 根据您的使用案例，选择基本身份验证和 OAuth 2.0 身份验证。

e. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Drupal 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

   i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

      A. 如果已选择基本身份验证，请输入您复制的密钥名称、用户名（Drupal 站点用户名）和密码（Drupal 站点密码），然后选择保存和添加密钥。

      B. 如果您选择 OAuth 2.0 身份验证，请输入在您的 Drupal 账户中生成的密钥名称、用户名（Drupal 站点用户名）、密码（Drupal 站点密码）、客户端 ID 和客户端密钥，然后选择保存和添加密钥。

   ii. 选择保存。

f. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

g. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

h. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 对于同步范围，从以下选项中进行选择：

> **Note**
>
> 当您选择爬取文章、基本页面和基本数据块时，它们的默认字段将自动同步。您也可以选择同步它们的评论、附件、自定义字段和其他自定义实体。

- 对于选择实体：
  - 文章 - 选择是否爬取文章、文章评论、评论和附件。
  - 基本页面 - 选择是否爬取基本页面、页面评论及其附件。
  - 基本数据块 - 选择是否爬取基本数据块、数据块评论及其附件。
  - 您也可以选择添加自定义内容类型和自定义数据块。

b. 对于其他配置 - 可选：

- 对于正则表达式模式 - 添加正则表达式模式以包含或排除特定的实体标题和文件名。最多可以添加 100 个模式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 对于内容、评论和附件-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Drupal

您必须使用 TemplateConfiguration API 指定数据来源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构DRUPAL时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在 Drupal 账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。

  如果您使用基本身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "username": "user name",
    "password": "password"
}
```

  如果您使用 OAuth 2.0 身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "username": "user name",
    "password": "password",
    "clientId": "client id",
    "clientSecret": "client secret"
}
```

> **Note**
>
> > **Note**
> >
> > 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM 角色 —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Drupal 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Drupal 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 `VpcConfiguration`，以便调用 `CreateDataSource`。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除筛选条件 - 您可以指定是否包括内容、评论和附件。您也可以指定正则表达式模式来包含或排除内容、评论和附件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。
- 字段映射 - 选择将 Drupal 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅Drupal 模板架构。

注意

- Drupal API 没有官方节流限制。
- Java 开发工具包不适用于 Drupal。
- 只能使用原生 JSON API 获取 Drupal 数据。
- 无法爬取与任何 Drupal 视图无关的内容类型。
- 您需要管理员访问权限才能从 Drupal 数据块中爬取数据。
- 没有可用于使用 HTTP 动词创建用户定义的内容类型的 JSON API。
- 文章、基本页面、基本数据块、用户定义的内容类型和用户定义的数据块类型的文档正文和评论以 HTML 格式显示。如果 HTML 内容格式不正确，则与 HTML 相关的标签将出现在文档正文和注释中，并在 Amazon Kendra 搜索结果中可见。
- 没有描述或正文的内容类型和区块类型不会被 Amazon Kendra收录。只有此类内容或区块类型的评论和附件才会被提取到您的 Amazon Kendra 索引中。

GitHub

GitHub 是一项用于软件开发的基于 Web 的托管服务，提供带有版本控制的代码存储和管理服务。您可以使用 Amazon Kendra 索引 GitHub 企业云 (SaaS) 和 GitHub 企业服务器（On Prem）存储库文件、议题和拉取请求、议题和拉取请求评论以及议题和拉取请求评论附件。您也可以选择包括或排除某些文件。

> **Note**
>
> Amazon Kendra 现在支持升级后的 GitHub 连接器。

控制台已自动为您升级。您在控制台中创建的任何新连接器都将使用升级后的架构。如果您使用 API，则现在必须使用TemplateConfiguration对象而不是GitHubConfiguration对象来配置您的连接器。
使用较旧的控制台和 API 架构配置的连接器将继续按配置运行。但是，您将无法对其进行编辑或更新。如果要编辑或更新连接器配置，则必须创建新的连接器。
我们建议将您的连接器工作流程迁移到升级版本。对使用旧架构配置的连接器的支持计划于 2024 年 6 月结束。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的 GitHub 数据源。

要对 Amazon Kendra GitHub 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

Amazon Kendra GitHub 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引 GitHub 数据源之前，请在 GitHub 和 AWS 帐户中进行这些更改。

在中 GitHub，请确保你有：

- 已创建具有 GitHub 组织管理权限的 GitHub 用户。

- 在 Git Hub 中配置了个人访问令牌以用作您的身份验证凭证。请参阅有关创建个人访问令牌的 GitHub 文档。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 推荐：为身份验证凭证配置了 OAuth 令牌。使用 OAuth 令牌可获得更好的 API 限制和连接器性能。请参阅有关 OAuth 授权的GitHub 文档。
- 记下了您使用的 GitHub服务类型的 GitHub 主机 URL。例如， GitHub 云的主机 URL 可能是 *https://api.github.com*，GitHub 服务器的主机 URL 可能是 *https: on-prem-host-url //api/v3/*。
- 记下您要连接 GitHub 的 GitHub企业云 (SaaS) 帐户或 GitHub 企业服务器（本地）帐户的组织名称。您可以登录到 GitHub 桌面，然后在个人资料图片下拉列表中选择 “您的组织”，找到您的组织名称。
- 可选（仅限服务器）：生成 SSL 证书并复制存储在存储 Amazon S3 桶中的证书的路径。 GitHub 如果您需要安全的 SSL 连接，则可以使用它进行连接。您只需使用 OpenSSL 在任何计算机上生成自签名 X509 证书。有关使用 OpenSSL 创建 X509 证书的示例，请参阅创建并签署 X509 证书。
- 添加了以下权限：

  适用于 GitHub 企业云 (SaaS)

  - `repo:status`— 授予对公共和私有仓库中提交状态的读/写权限。只有在授予其他用户或服务访问私有存储库提交状态的权限而不授予对代码的访问权限时，才需要使用此作用域。
  - `repo_deployment`— 授予访问公共和私有仓库部署状态的权限。只有在授予其他用户或服务访问部署状态的权限时，才需要使用此作用域，而不授予对代码的访问权限。
  - `public_repo`— 限制对公共存储库的访问权限。这包括对代码、提交状态、存储库项目、协作者以及公共仓库和组织的部署状态的读/写权限。为公共仓库加星标也是必需的。
  - `repo:invite`— 授予接受/拒绝在仓库上进行协作邀请的能力。只有在不授予代码访问权限的情况下授予其他用户或服务访问邀请的权限时，才需要使用此范围。
  - `security_events`— 授权：在代码扫描 API 中读取和写入安全事件的权限。只有在授予其他用户或服务访问安全事件的权限而不授予对代码的访问权限时，才需要使用此范围。
  - `read:org`— 对组织成员资格、组织项目和团队成员资格的只读访问权限。
  - `user:email`— 授予对用户电子邮件地址的读取权限。Amazon Kendra 要求抓取 ACL。

- `user:follow`— 授予关注或取消关注其他用户的权限。Amazon Kendra 要求抓取 ACL。
- `read:user`— 授予读取用户个人资料数据的权限。Amazon Kendra 要求抓取 ACL。
- `workflow`— 授予添加和更新 GitHub 操作工作流程文件的功能。如果同一存储库中的另一个分支上存在相同的文件（路径和内容相同），则可以在没有此范围的情况下提交工作流文件。

有关更多信息，请参阅 Doc s 中的 GitHub OAuth 应用程序的作用域。

适用于 GitHub 企业服务器（本地版）

- `repo:status`— 授予对公共和私有仓库中提交状态的读/写权限。只有在授予其他用户或服务访问私有存储库提交状态的权限而不授予对代码的访问权限时，才需要使用此作用域。
- `repo_deployment`— 授予访问公共和私有仓库部署状态的权限。只有在授予其他用户或服务访问部署状态的权限时，才需要使用此作用域，而不授予对代码的访问权限。
- `public_repo`— 限制对公共存储库的访问权限。这包括对代码、提交状态、存储库项目、协作者以及公共仓库和组织的部署状态的读/写权限。为公共仓库加星标也是必需的。
- `repo:invite`— 授予接受/拒绝在仓库上进行协作邀请的能力。只有在不授予代码访问权限的情况下授予其他用户或服务访问邀请的权限时，才需要使用此范围。
- `security_events`— 授权：在代码扫描 API 中读取和写入安全事件的权限。只有在授予其他用户或服务访问安全事件的权限而不授予对代码的访问权限时，才需要使用此范围。
- `read:user`— 授予读取用户个人资料数据的权限。Amazon Q Business 要求抓取 ACL。
- `user:email`— 授予对用户电子邮件地址的读取权限。Amazon Q Business 要求抓取 ACL。
- `user:follow`— 授予关注或取消关注其他用户的权限。Amazon Q Business 要求抓取 ACL。
- `site_admin`— 授予站点管理员访问 GitHub 企业服务器管理 API 端点的权限。
- `workflow`— 授予添加和更新 GitHub 操作工作流程文件的功能。如果同一存储库中的另一个分支上存在相同的文件（路径和内容相同），则可以在没有此范围的情况下提交工作流文件。

有关更多信息，请参阅GitHub文档中的 OAuth 应用程序的作用域和在开发者中了解 OAuth 应用程序的范围。GitHub

- 已选中每个文档在您计划用于同一索引的其他数据源中 GitHub 以及其他数据源中都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。

- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 GitHub 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下该密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 GitHub 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 GitHub 数据源，您必须提供 GitHub 数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置 Amazon Kendra， GitHub 请参阅先决条件。

Console

要连接 Amazon Kendra 到 GitHub

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。

4. 在添加数据源页面上，选择GitHub 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的GitHub 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. GitHub来源-在GitHub 企业云和GitHub企业服务器之间进行选择。
   b. GitHub 主机 URL —例如， GitHub 云的主机 URL 可能是 *https://api.github.com*，GitHub 服务器的主机 URL 可以是 *https: ///api on-prem-host-url /v3/*。
   c. GitHub 组织名称-输入您的 GitHub组织名称。你可以在你的 GitHub 账户中找到你的组织信息。
   d. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
   e. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 GitHub 身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。
      i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：
         A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-GitHub-” 会自动添加到您的密钥名称中。
         B. 对于GitHub令牌-输入中配置的身份验证凭据值。 GitHub
      ii. 保存并添加您的密钥。
   f. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

g. l Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

h. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 选择存储库-选择对所有存储库进行爬网或选择。

如果您选择对所选存储库进行爬网，请添加存储库的名称，也可以添加任何特定分支的名称。

b. 内容类型-从文件、议题、拉取请求等中选择要抓取的内容类型。

c. 正则表达式模式 - 添加包含或排除某些文件的正则表达式模式。

d. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

e. 在 F requency 的同步运行计划中-选择同步数据源内容和更新索引的频率。

f. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 GitHub

您必须使用 TemplateConfiguration API 指定数据来源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构GITHUB时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- GitHubtyp e —将类型指定为SAAS或ON_PREMISE。
- 主机 URL-指定 GitHub 主机 URL 或 API 端点网址。例如，如果您使用 GitHub SaaS/Enterprise Cloud，则主机 URL 可能是https://api.github.com，对于 GitHub 本地/企业服务器，主机 URL 可能是。https://on-prem-host-url/api/v3/
- 组织名称-指定 GitHub 账户所在组织的名称。您可以登录到 GitHub 桌面，然后在个人资料图片下拉列表中选择 “您的组织”，找到您的组织名称。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访

问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

- 亚马逊秘密资源名称 (ARN)-提供包含您账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。 GitHub 密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "personalToken": "token"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 GitHub 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 GitHub 数据源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。

> **Note**
>
> 如果您使用 GitHub 服务器，则必须使用 Amazon VPC 才能连接到 GitHub服务器。

- 存储库筛选器-按名称和分支名称筛选存储库。
- 文档/内容类型-指定是否抓取存储库文档、议题、议题评论、议题评论附件、拉取请求、拉取请求评论、拉取请求评论附件。
- 包含和排除筛选器-指定是包括还是排除某些文件和文件夹。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 访问控制列表 (ACL)-如果您有 ACL 并希望将其用于访问控制，则指定是否要搜索文档的 ACL 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

- 字段映射-选择将 GitHub 数据源字段映射到 Amazon Kendra 索引字段。您可以包括文档、提交、议题、议题附件、议题评论、拉取请求、拉取请求附件、拉取请求评论等字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 为了让 Amazon Kendra 搜索您的文档，必须填写文件正文字段或与您的文档相同的正文。您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅 GitHub 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与 GitHub 数据源集成的更多信息，请参阅：

- 借助连接器的强大功能，重新构想对 GitHub 存储库的 Amazon Kendra GitHub 搜索

Gmail

Gmail 是 Google 开发的电子邮件客户端，可用于发送包含文件附件的电子邮件。您可以使用文件夹和标签对 Gmail 邮件进行排序并将其存储在您的电子邮件收件箱中。您可以使用索 Amazon Kendra 引电子邮件和邮件附件。您还可以配置 Amazon Kendra 为包括或排除特定的电子邮件、邮件附件和标签以进行索引。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI 连接 Amazon Kendra 到 Gmail 数据源。

要对 Amazon Kendra Gmail 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Gmail 数据源之前，请在您的 Gmail 和 AWS 帐号中进行这些更改。

在 Gmail 中，请确保：

- 已创建 Google Cloud Platform 管理员账号并创建 Google Cloud 项目。
- 在您的管理员账户中激活了 Gmail API 和管理开发工具包 API。
- 为您的 Gmail 创建了一个服务帐号并下载了一个 JSON 私有密钥。有关如何创建和访问私有密钥的信息，请参阅有关如何创建服务帐号密钥和服务帐号凭证的 Google Cloud 文档。
- 已复制您的管理员账户电子邮件、服务帐号电子邮件地址和用作身份验证凭据的私钥。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 为您的用户和要索引的共享目录添加了以下 Oauth 作用域（使用管理员角色）：
  - https://www.googleapis.com/auth/admin.directory.user.readonly
  - https://www.googleapis.com/auth/gmail.readonly
- 在 Gmail 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。

- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Gmail 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Gmail 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 Gmail 数据源，您必须提供 Gmail 凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为 Gmail 配置 Amazon Kendra，请参阅先决条件。

Console

连接 Amazon Kendra 到 Gmail

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。

4. 在添加数据源页面上，选择 Gmail 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Gmail 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

   b. 在AWS Secrets Manager 密钥身份验证中-选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Gmail 身份验证凭据。如果您选择创建新密钥，则会打开一个 AWS Secrets Manager 秘密窗口。

      - 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

        A. 密钥名称 - 密钥的名称。

        B. 客户端电子邮件地址 - 您从 Google 服务帐号中复制的客户端电子邮件地址。

        C. 管理员账户电子邮件地址 - 您要使用的管理员账户电子邮件地址。

        D. 私有密钥 - 您从 Google 服务帐号中复制的私有密钥。

        E. 保存并添加您的密钥。

   c. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

   d. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

e. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 对于实体类型-选择同步邮件附件。

b. （可选）对于其他配置，输入以下信息：

i. 日期范围-输入日期范围以指定要抓取的电子邮件的开始和结束日期。

ii. 电子邮件域-根据 “收件人”、“发件人”、“抄送” 和 “密件抄送” 电子邮件域名包含或排除某些电子邮件。

iii. 主题中的关键字-根据电子邮件主题中的关键字包含或排除电子邮件。

> **Note**
>
> 您也可以选择包含与您输入的所有主题关键字相匹配的任何文档。

iv. 标签-添加正则表达式模式以包含或排除某些电子邮件标签。

v. 附件-添加正则表达式模式以包含或排除某些电子邮件附件。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

> **Important**
>
> 由于没有 API 可以更新永久删除的 Gmail 邮件，所以新增内容、修改内容或已删除内容会同步：
>
> - 不会从您的 Amazon Kendra 索引中移除从 Gmail 中永久删除的邮件
> - 无法同步 Gmail 电子邮件标签中的更改

要将您的 Gmail 数据来源标签更改和永久删除的电子邮件同步到您的 Amazon Kendra 索引，您必须定期进行全面爬取。

d. 在同步运行计划中，对于频率-选择同步数据源内容和更新索引的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

> **Note**
>
> Amazon Kendra 由于 API 限制，Gmail 数据源连接器不支持创建自定义索引字段。

b. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接 Amazon Kendra 到 Gmail

您必须使用 TemplateConfiguration API 指定数据来源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构GMAIL时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅为新增、修改和删除的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

> **Important**
>
> 由于没有 API 可以更新永久删除的 Gmail 邮件，所以新增内容、修改内容或已删除内容会同步：
>
> - 不会从您的 Amazon Kendra 索引中移除从 Gmail 中永久删除的邮件
> - 无法同步 Gmail 电子邮件标签中的更改
>
> 要将您的 Gmail 数据源标签更改和永久删除的电子邮件同步到您的 Amazon Kendra 索引，您必须定期进行全面抓取。

- 亚马逊秘密资源名称 (ARN)-提供包含您的 Gmail 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "adminAccountEmailId": "service account email",
    "clientEmailId": "user account email",
    "privateKey": "private key"
}
```

- IAM 角色-指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Gmail 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Gmail 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 `VpcConfiguration`，以便调用 `CreateDataSource`。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器-指定是包含还是排除某些 “收件人”、“发件人”、“抄送”、“密件抄送” 电子邮件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Gmail 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

> **Note**
>
> Amazon Kendra 由于 API 限制，Gmail 数据源连接器不支持创建自定义索引字段。

有关要配置的其他重要 JSON 键的列表，请参阅 Gmail 模板架构。

了解更多信息

要详细了解如何 Amazon Kendra 与 Gmail 数据源集成，请参阅：

- 使用适用于 Amazon Kendra的 Gmail 连接器在 Google 工作区中对电子邮件进行智能搜索.

注意

- 由于没有 API 可以更新永久删除的 Gmail 邮件，所以FULL_CRAWL/同步新增、修改或删除的内容：
  - 不会从您的 Amazon Kendra 索引中移除从 Gmail 中永久删除的邮件
  - 无法同步 Gmail 电子邮件标签中的更改

  要将您的 Gmail 数据源标签更改和永久删除的电子邮件同步到您的 Amazon Kendra 索引，您必须定期进行全面抓取。
- Amazon Kendra 由于 API 限制，Gmail 数据源连接器不支持创建自定义索引字段。

Google Drive

Google Drive 是一项基于云的文件存储服务。您可以使用 Amazon Kendra 为存储在 Google Drive 数据来源中的共享云端硬盘、我的云端硬盘和与我共享文件夹中的文档编制索引。您可以为 Google Workspace 文档以及文档类型中列出的文档编制索引。您还可以使用包含和排除筛选条件按文件名、文件类型和文件路径对内容进行索引。

您可以使用Amazon Kendra 控制台、API 或 TemplateConfigurationAPI Amazon Kendra 连接到您的 Google 云端硬盘数据源。GoogleDriveConfiguration

Amazon Kendra 有两个版本的 Google 云端硬盘连接器。每个版本支持的功能包括：

谷歌云端硬盘连接器 V1.0/API GoogleDriveConfiguration

- 字段映射
- 用户访问控制
- 包含/排除筛选条件

谷歌云端硬盘连接器 V2.0/API TemplateConfiguration

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

> **Note**
>
> 对谷歌云端硬盘连接器 V1.0/Google DriveConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 Google 云端硬盘连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra Google 云端硬盘数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

Google Drive 连接器 V1.0

Google Drive 是一项基于云的文件存储服务。您可以使用 Amazon Kendra 将存储在 Google 云端硬盘数据源中的共享云端硬盘、“我的云端硬盘” 和 “与我共享” 文件夹中的文档和评论编入索引。您可以为 Google Workspace 文档以及文档类型中列出的文档编制索引。您还可以使用包含和排除筛选条件按文件名、文件类型和文件路径对内容进行索引。

> **Note**
>
> 对谷歌云端硬盘连接器 V1.0/Google DriveConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 Google 云端硬盘连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra Google 云端硬盘数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

- 字段映射
- 用户访问控制
- 包含/排除筛选条件

先决条件

在使用 Amazon Kendra 索引 Google 云端硬盘数据源之前，请在您的 Google 云端硬盘和 AWS 帐号中进行这些更改。

在 Google Drive 中，请确保：

- 要么超级管理员角色已授予访问权限，要么是具有管理权限的用户。如果您已获得超级管理员角色的访问权限，则无需为自己设置超级管理员角色。

- 使用该账号创建了一个服务帐号，并激活了启用 G Suite 全网域授权，并使用该帐号将 JSON 密钥作为私有密钥。
- 已复制您的用户账户电子邮件和服务账户电子邮件。当您与您建立联系时，将您的用户帐户电子邮件作为管理员帐户电子邮件输入，在您的 AWS Secrets Manager 密钥中将服务帐户电子邮件作为客户端电子邮件输入。 Amazon Kendra

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 在您的账户中添加了管理员 SDK API 和 Google Drive API。
- 使用超级管理员角色向您的服务帐号添加（或要求具有超级管理员角色的用户添加）以下权限：
  - https://www.googleapis.com/auth/drive.readonly
  - https://www.googleapis.com/auth/drive.metadata.readonly
  - https://www.googleapis.com/auth/admin.directory.user.readonly
  - https://www.googleapis.com/auth/admin.directory.group.readonly
- 在 Google Drive 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Google Drive 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Google 云端硬盘数据源关联到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到您的 Google 云端硬盘数据源，您必须提供 Google 云端硬盘数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未配置 Google 云端硬盘， Amazon Kendra 请参阅先决条件。

Console

连接 Amazon Kendra 到 Google 云端硬盘

1. 登录 AWS 管理控制台并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

   > **Note**
   >
   > 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据来源页面上，选择 Google Drive 连接器 V1.0，然后选择添加连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

a. 对于身份验证类型 - 选择现有和新建。如果您选择使用现有密钥，请使用选择密钥来选择您的密钥。

b. 如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥选项。

- 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Google 云端硬盘” 会自动添加到您的密钥名称中。

B. 对于管理员账户电子邮件地址、客户端电子邮件地址和私有密钥，请输入您生成并从 Google Drive 账户下载的身份验证凭证值。

C. 选择保存身份验证。

c. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

d. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 排除用户账户 - 您要从索引中排除的 Google Drive 用户。您最多可以添加 100 个用户账户。

b. 排除共享云端硬盘 - 要从索引中排除的 Google Drive 共享云端硬盘。您最多可以添加 100 个共享驱动器。

c. 排除文件类型驱动器 - 要从索引中排除的 Google Drive 文件类型。您也可以选择编辑 MIME 类型选择。

d. 对于其他配置 - 添加正则表达式模式以包含或排除某些内容。最多可以添加 100 个模式。

e. 频率 - Amazon Kendra 与您的数据来源同步的频率。

f. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 对于GoogleDrive 字段名称和其他建议的字段映射-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接 Amazon Kendra 到 Google 云端硬盘

您必须使用 GoogleDriveConfigurationAPI 指定以下内容：

- 亚马逊秘密资源名称 (ARN)-提供包含您的 Google 云端硬盘账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "clientAccount": "service account email",
    "adminAccount": "user account email"",
    "privateKey": "private key"
}
```

- IAM 角色-指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥的权限以及调用 Google 云端硬盘连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Google Drive 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 包含和排除筛选条件 - 默认情况下， Amazon Kendra 会为 Google Drive 中的所有文档编制索引。您可以指定是包含还是排除共享云端硬盘、用户账户、文档 MIME 类型和文件中的某些内容。如果您选择排除用户账户，则该账户拥有的 “我的云端硬盘” 中的所有文件都不会被编入索引。除非文件的所有者也被排除在外，否则将对与用户共享的文件编制索引。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件

不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 字段映射 - 选择将 Google Drive 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

了解更多信息

要详细了解如何 Amazon Kendra 与 Google 云端硬盘数据源集成，请参阅：

- 开始使用 Amazon Kendra Google 云端硬盘连接器

Google Drive 连接器 V2.0

Google Drive 是一项基于云的文件存储服务。您可以使用 Amazon Kendra 将存储在 Google 云端硬盘数据源中的共享云端硬盘、“我的云端硬盘” 和 “与我共享” 文件夹中的文档和评论编入索引。您可以为 Google Workspace 文档以及文档类型中列出的文档编制索引。您还可以使用包含和排除筛选条件按文件名、文件类型和文件路径对内容进行索引。

> **Note**
>
> 对谷歌云端硬盘连接器 V1.0/Google DriveConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 Google 云端硬盘连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra Google 云端硬盘数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Google 云端硬盘数据源之前，请在您的 Google 云端硬盘和 AWS 帐号中进行这些更改。

在 Google Drive 中，请确保：

- 要么超级管理员角色已授予访问权限，要么是具有管理权限的用户。如果您已获得超级管理员角色的访问权限，则无需为自己设置超级管理员角色。
- 已配置 Google Drive 服务帐号连接凭证，其中包含您的管理员帐号电子邮件、客户电子邮件（服务帐号电子邮件）和私有密钥。请参阅有关创建和删除服务帐号密钥的 Google Cloud 文档。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 创建了一个 Google Cloud 服务帐号（具有使用用户身份授权的帐号），并激活了 “启用 G Suite 全网域授权” 进行 server-to-server身份验证，然后使用该帐号生成了 JSON 私钥。

> **Note**
>
> 私有密钥应在创建服务帐号之后生成。

- 在您的用户账户中添加了管理员 SDK API 和 Google Drive API。
- 可选：将 Google Drive OAuth 2.0 连接凭证配置为特定用户的连接凭证，其中包含客户端 ID、客户端密钥和刷新令牌。您需要它来爬取个人账户数据。请参阅有关如何使用 OAuth 2.0 访问 API 的 Google 文档。
- 使用超级管理员角色向您的服务帐号添加（或要求具有超级管理员角色的用户添加）以下 OAuth 范围。要爬取 Google Workspace 网域中所有用户的所有文档和访问控制（ACL）信息，需要这些 API 范围：
  - https://www.googleapis.com/auth/drive.readonly - 查看和下载您所有的 Google Drive 文件
  - https://www.googleapis.com/auth/drive.metadata.readonly - 查看您的 Google Drive 中文件的元数据
  - https://www.googleapis.com/auth/admin.directory.group.readonly - 范围仅用于检索组、组别名和成员信息。这是 Amazon Kendra 身份搜寻器所必需的。
  - https://www.googleapis.com/auth/admin.directory.user.readonly - 范围仅用于检索用户或用户别名。在 Ident Amazon Kendra ity Crawler 中列出用户和设置 ACL 时需要这样做。
  - https://www.googleapis.com/auth/cloud-platform - 范围用于生成访问令牌以获取大型 Google Drive 文件内容。
  - https://www.googleapis.com/auth/forms.body.readonly - 范围用于从 Google 表单中获取数据。

  要支持 Forms API，请添加以下附加范围：

  - https://www.googleapis.com/auth/forms.body.readonly
- 在 Google Drive 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Google Drive 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Google 云端硬盘数据源关联到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到您的 Google 云端硬盘数据源，您必须提供 Google 云端硬盘数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未配置 Google 云端硬盘， Amazon Kendra 请参阅先决条件。

Console

连接 Amazon Kendra 到 Google 云端硬盘

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。

4. 在添加数据源页面上，选择 Google 云端硬盘连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Google 云端硬盘连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

   b. 对于身份验证 - 根据您的使用案例，选择 Google 服务帐号或 OAuth 2.0 身份验证。

   c. AWS Secrets Manager s@@ ecret —选择现有密钥，或创建一个新 Secrets Manager 密钥来存储您的 Google 云端硬盘身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

      i. 如果您选择了 Google 服务帐号，请输入您的密钥名称、服务帐号配置中的管理员用户或 “服务帐号用户” 的电子邮件 ID（管理员电子邮件）、服务帐号的电子邮件 ID（客户端电子邮件）以及您在服务帐号中创建的私钥。

         保存并添加您的密钥

      ii. 如果您选择 OAuth 2.0 身份验证，请输入您在 OAuth 账户中创建的密钥名称、客户端 ID、客户端密钥和刷新令牌。

         保存并添加您的密钥。

   d. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

   e. （仅适用于 Google 服务帐号身份验证用户）

      l Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结

果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

f. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

g. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 同步内容-选择要抓取的选项或内容。您可以选择抓取 “我的云端硬盘”（个人文件夹）、“共享云端硬盘”（与您共享的文件夹）或两者兼而有之。您也可以添加文件注释。

b. 在 “其他配置-可选” 中您还可以输入以下可选信息：

i. 目标受众-为要抓取的文档添加特定的目标受众。

ii. 最大文件大小-设置要抓取的文件的最大大小限制（以 MB 为单位）。

iii. 用户电子邮件-添加要包含或排除的用户电子邮件。

iv. 共享云端硬盘-添加要包含或排除的共享云端硬盘名称。

v. MIME 类型-添加要包含或排除的 MIME 类型。

vi. 实体正则表达式模式-添加正则表达式模式以包含或排除所有支持的实体的某些附件。最多可以添加 100 个模式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改后的全新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

> **Important**
>
> Google Drive API 不支持从永久删除的文件中检索评论。可以检索已丢弃文件中的评论。当文件被丢弃时，连接器将从 Amazon Kendra 索引中删除注释。

d. 在“同步运行计划”中，对于“频率”，选择同步数据源内容和更新索引的频率。

e. 在“同步运行历史记录”中，选择在同步数据源 Amazon S3 时将自动生成的报告存储在中。这对于跟踪同步数据源时出现的问题非常有用。

f. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 对于文件-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

> **Note**
>
> Google Drive API 不支持创建自定义字段。自定义字段映射不适用于 Google Drive 连接器。

b. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接 Amazon Kendra 到 Google 云端硬盘

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构GOOGLEDRIVEV2时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 身份验证类型-指定是使用服务帐户身份验证还是 OAuth 2.0 身份验证。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。

- FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

> **Important**
>
> Google Drive API 不支持从永久删除的文件中检索评论。可以检索已丢弃文件中的评论。当文件被丢弃时，连接器将从 Amazon Kendra 索引中删除注释。

- 亚马逊秘密资源名称 (ARN)-提供包含您在 Google 云端硬盘账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。如果您使用 Google 服务账户身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "clientEmail": "user account email",
    "adminAccountEmail": "service account email",
    "privateKey": "private key"
}
```

如果您使用 OAuth 2.0 身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "clientID": "OAuth client ID",
    "clientSecret": "client secret",
    "refreshToken": "refresh token"
}
```

- IAM 角色-指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥的权限以及调用 Google 云端硬盘连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Google Drive 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 我的云端硬盘、共享云端硬盘、评论-您可以指定是否抓取这些类型的内容。
- 包含和排除过滤器-您可以指定是包含还是排除某些用户帐户、共享云端硬盘和 MIME 类型。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 访问控制列表 (ACL)-如果您有 ACL 并希望将其用于访问控制，则指定是否要搜索文档的 ACL 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。
- 字段映射 - 选择将 Google Drive 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅 Google Drive 模板架构。

注意

- 由于 Google Drive UI 不支持创建自定义字段，因此自定义字段映射不适用于 Google Drive 连接器。
- Google Drive API 不支持从永久删除的文件中检索评论。但是，对于已丢弃的文件，可以检索评论。当文件被丢弃时， Amazon Kendra 连接器将从 Amazon Kendra 索引中删除注释。
- Google Drive API 不会返回 .docx 文件中存在的评论。

IBM DB2

IBM DB2 是 IBM 开发的一个关系数据库管理系统。如果您是 IBM DB2 用户，则可以使用 Amazon Kendra 为您的 IBM DB2 数据来源编制索引。 Amazon Kendra IBM DB2数据源连接器支持 DB2 11.5.7。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的IBM DB2数据源。

要对 Amazon Kendra IBM DB2数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引IBM DB2数据源之前，请在IBM DB2和 AWS 帐户中进行这些更改。

在 IBM DB2 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 最佳做法是提供 Amazon Kendra 只读数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。
- 在 IBM DB2 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 IBM DB2 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将IBM DB2数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的IBM DB2数据源，您必须提供IBM DB2凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置，IBM DB2请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 IBM DB2

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。

2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择IBM DB2连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的IBM DB2连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 在源中，输入以下信息：

   b. 主机 - 输入数据库主机名。

   c. 端口 - 输入数据库端口。

   d. 实例 - 输入数据库实例。

   e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

   f. 在身份验证中 - 请输入以下信息：

      - AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的IBM DB2身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

        A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

           I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-IBM DB2-” 会自动添加到您的密钥名称中。

           II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

B. 选择保存。

g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

h. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。
- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 IBM DB2

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 db2。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。

  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。IBM DB2密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用IBM DB2连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 IBM DB2 S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 IBM DB2 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅IBM DB2 模板架构。

注意

- 在 Amazon Kendra 检查更新内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。
- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 最佳做法是提供 Amazon Kendra 只读数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

Jira

Jira 是一款用于软件开发、产品管理和错误跟踪的项目管理工具。您可以使用 Amazon Kendra 索引 Jira 项目、议题、评论、附件、工作日志和状态。

Amazon Kendra 目前仅支持 Jira Cloud。

您可以使用Amazon Kendra 控制台或 JiraConfiguration API Amazon Kendra 连接到您的 Jira 数据源。有关每个版本支持的功能列表，请参阅 支持的特征。

要对 Amazon Kendra Jira 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

Amazon Kendra Jira 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Jira 数据源之前，请在您的 Jira 和 AWS 账户中进行这些更改。

在 Jira 中，请确保：

- 已配置的 API 令牌身份验证凭证，其中包括 Jira ID（用户名或电子邮件）和 Jira 凭据（Jira API 令牌）。请参阅有关管理 API 令牌的 Atlassian 文档。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 已记下您的 Jira 账户设置中的 Jira 账户 URL。例如，*https://company.atlassian.net/*。
- 在 Jira 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Jira 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Jira 数据源连接到 Amazon Kendra时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 Jira 数据源，您必须提供 Jira 数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为 Jira 配置 Amazon Kendra，请参阅先决条件。

Console

连接 Amazon Kendra 到 Jira

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Jira 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Jira 连接器。

5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. Jira 账户网址-输入你的 Jira 账户网址。例如：*https://company.atlassian.net/*。

   b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

   c. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Jira 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

      i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

         A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Jira-” 会自动添加到您的密钥名称中。

         B. 对于 Jira ID - 输入 Jira 用户名或电子邮件地址。

         C. 对于密码/令牌-输入在 Jira 中配置的 Jira API 令牌。

      ii. 保存并添加您的密钥。

   d. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

   e. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

   f. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

g. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 选择要为哪个 Jira 项目编制索引-选择对所有项目或特定项目进行爬网。

   b. 其他配置-指定某些状态和问题类型。选择抓取评论、附件和工作日志。使用正则表达式模式来包含或排除某些内容。

   c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

      - 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
      - 修改后的全新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
      - 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

   d. 在 “同步” 运行计划中，“频率”-选择同步数据源内容和更新索引的频率。

   e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

   a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

   b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

   c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接 Amazon Kendra 到 Jira

您必须使用 JiraConfiguration API 指定以下内容：

- 数据来源 URL - 指定您的 Jira 账户 URL。例如，*company.atlassian.net*。
- 亚马逊秘密资源名称 (ARN)-提供包含您的 Jira 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "jiraId": "Jira user name or email",
    "jiraCredential": "Jira API token"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Jira 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Jira 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 在数据来源配置中指定 VpcConfiguration。请参阅配置 Amazon Kendra 以使用 VPC。
- 更改日志-是否 Amazon Kendra 应使用 Jira 数据源更改日志机制来确定是否必须更新索引中的文档。

> **Note**
>
> 如果您不想让 Amazon Kendra 扫描所有文档，请使用更改日志。如果您的更改日志很大，则扫描 Jira 数据源中的文档所花费的时间可能比处理更改日志所需的时间 Amazon Kendra 少。如果您是首次将 Jira 数据来源与索引同步，则会扫描所有文档。

- 包含和排除筛选器-您可以指定是包含还是排除某些文件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 评论、附件和工作日志-您可以指定是否抓取问题的某些评论、附件和工作日志。

- 项目、问题、状态-您可以指定是否对某些项目 ID、问题类型和状态进行爬网。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 抓取文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Jira 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

了解更多信息

要了解有关 Amazon Kendra 与 Jira 数据源集成的更多信息，请参阅：

- 使用 Jira 云连接器智能搜索你的 Amazon Kendra Jira 项目

Microsoft Exchange

Microsoft Exchange 是一款用于消息、会议和文件共享的企业协作工具。如果你是微软 Exchange 用户，你可以使用 Amazon Kendra 索引你的微软 Exchange 数据源。

你可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到你的 Microsoft Exchange 数据源。

有关 Amazon Kendra 微软 Exchange 数据源连接器的疑难解答，请参阅数据来源故障排除。

支持的特征

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引微软 Exchange 数据源之前，请在微软 Exchange 和 AWS 账户中进行这些更改。

在 Microsoft Exchange 中，请确保：

- 在 Office 365 中创建了一个 Microsoft Exchange 账户。
- 记下了您的 Microsoft 365 租户 ID。在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中可以找到您的租户 ID。
- 在 Azure 门户中配置了 OAuth 应用程序，并记下了客户端 ID 和客户端密钥或客户端凭据。有关更多信息，请参阅 Microsoft 教程和注册应用程序示例。

> **Note**
>
> 在 Azure 门户中创建或注册应用程序时，密钥 ID 代表实际的密钥值。在创建密钥和应用程序时，您必须立即记下或保存实际的密钥值。你可以通过在 Azure 门户中选择应用程序的名称，然后导航到证书和密钥上的菜单选项来访问你的密钥。
> 你可以通过在 Azure 门户中选择应用程序的名称，然后导航到概述页面来访问你的客户端 ID。应用程序（客户端）ID 是客户端 ID。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 为连接器应用程序添加了以下权限：

Microsoft Graph	Office 365 Exchange Online
• Mail.Read（应用程序） • 邮件。 ReadBasic （应用程序） • 邮件。 ReadBasic.All（应用程序） • Calendars.Read（应用程序） • User.Read.All（应用程序）	full_access_as_app（应用程序）

Microsoft Graph	Office 365 Exchange Online
• Contacts.Read（应用程序） • Notes.Read.All（应用程序） • Directory.Read.All（应用程序） • 新闻。 AccessAsUser.All（已委托）	

- 在 Microsoft Exchange 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Microsoft Exchange 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果你没有现有的 IAM 角色或密钥，则可以在将 Microsoft Exchange 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到你的微软 Exchange 数据源，你必须提供微软 Exchange 数据源的必要细节，这样 Amazon Kendra 才能访问你的数据。如果你尚未为 Microsoft Exchange 进行配置 Amazon Kendra，请参阅先决条件。

Console

Amazon Kendra 连接到微软 Exchange

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在 “添加数据源” 页面上，选择 Microsoft Exchange 连接器，然后选择 “添加连接器”。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Microsoft Exchange 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 租户 ID-输入你的微软 365 租户 ID。在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中可以找到您的租户 ID。
   b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

c. AWS Secrets Manager 密钥 — 选择现有密钥或创建新 Secrets Manager 密钥来存储你的 Microsoft Exchange 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

   i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

      A. 密钥名称 - 密钥的名称。前缀 'AmazonKendra-微软 Exchange

      B. 对于客户端 ID，客户端密钥-在 Azure 门户中输入在 Microsoft Exchange 中配置的身份验证凭据。

   ii. 保存并添加您的密钥。

d. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

e. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

f. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 用户 ID-如果您想按某些电子邮件筛选内容，请提供用户电子邮件。

b. 其他配置-指定要抓取的内容类型。

- 实体类型-您可以选择抓取日历或联系 OneNotes人内容。
- 日历搜寻-输入开始和结束日期，以便在特定日期之间搜寻内容。
- 包括电子邮件-输入 “收件人”、“发件人” 和电子邮件主题行以筛选要抓取的某些电子邮件。
- 共享文件夹访问权限-选择启用对访问控制列表的抓取，以控制您的 Microsoft Exchange 数据源的访问控制。
- 域名正则表达式-添加正则表达式模式以包含或排除某些电子邮件域。
- 正则表达式模式 - 添加包含或排除某些文件的正则表达式模式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率-选择同步数据源内容和更新索引的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

> **Note**
>
> Amazon Kendra 微软 Exchange 数据源连接器不支持自定义字段映射。

b. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

Amazon Kendra 连接到微软 Exchange

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构MSEXCHANGE时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 租户 ID - 在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中可以找到您的租户 ID。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。

  - FULL_CRAWL每次数据源与索引同步时，仅为新增、修改和删除的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含微软 Exchange 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "clientId": "client ID",
    "clientSecret": "client secret"
}
```

- IAM 角色 —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Microsoft Exchange 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Microsoft Exchange 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器-指定是包含还是排除某些内容。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 访问控制列表 (ACL)-如果您有 ACL 并希望将其用于访问控制，则指定是否要搜索文档的 ACL 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Microsoft Exchange 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅微软 Exchange 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与微软 Exchange 数据源集成的更多信息，请参阅：

- 使用适用于 Amazon Kendra的 Exchange 连接器为 Microsoft Exchange 内容编制索引

微软 OneDrive

Microsoft OneDrive 是一项基于云的存储服务，你可以用它来存储、共享和托管你的内容。您可以使用 Amazon Kendra 为 OneDrive 数据源编制索引。

您可以使用Amazon Kendra 控制台和 OneDriveConfigurationAPI Amazon Kendra 连接到您的 OneDrive 数据源。

Amazon Kendra 有两个版本的 OneDrive 连接器。每个版本支持的功能包括：

微软 OneDrive 连接器 V1.0/API OneDriveConfiguration

- 字段映射
- 包含/排除筛选条件

微软 OneDrive 连接器 V2.0/API TemplateConfiguration

- 用户上下文筛选
- 用户身份搜寻器
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

> **Note**
>
> 对 OneDrive 连接器 V1.0/ OneDriveConfiguration API 的支持计划于 2023 年 6 月结束。我们建议使用 OneDrive 连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra OneDrive 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

微软 OneDrive 连接器 V1.0

Microsoft OneDrive 是一项基于云的存储服务，你可以用它来存储、共享和托管你的内容。你可以使用索引你 Amazon Kendra 的 Microsoft OneDrive 数据源。

> **Note**
>
> 对 OneDrive 连接器 V1.0/Microsft OneDrive API 的支持计划于 2023 年 6 月结束。我们建议使用 OneDrive 连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra OneDrive 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 包含/排除筛选条件

先决条件

在使用索 Amazon Kendra 引 OneDrive 数据源之前，请在 OneDrive 和 AWS 帐户中进行这些更改。

在 Azure Active Directory（AD）中，请确保：

- 已创建 Azure Active Directory（AD）应用程序
- 使用 AD 应用程序 ID 在 AD 站点上注册应用程序的密钥。该密钥必须包含应用程序 ID 和密钥。
- 已复制组织的 AD 域。
- 在 Microsoft Graph 选项上为你的 AD 应用程序添加了以下应用程序权限：
  - 读取所有网站集中的文件（File.Read.All）
  - 阅读所有用户的完整个人资料（User.Read.All）
  - 读取目录数据（Directory.Read.All）
  - 阅读所有组（Group.Read.All）
  - 阅读所有网站集中的项目（Site.Read.All）
- 复制必须为其文档编制索引的用户列表。您可以选择提供用户名列表，也可以在存储在 Amazon S3 中的文件中提供用户名。创建数据来源后，您可以：
  - 修改用户列表。
  - 从用户列表更改为存储在存储 Amazon S3 桶中的列表。
  - 更改用户列表的 Amazon S3 存储桶位置。如果您更改存储桶位置，则还必须更新数据源的 IAM 角色，使其能够访问存储桶。

> **Note**
>
> 如果您将用户名列表存储在 Amazon S3 存储桶中，则数据源的 IAM 策略必须提供对存储桶的访问权限以及对存储桶加密时使用的密钥（如果有）的访问权限。

- 已选中每个文档在 OneDrive 您计划用于同一索引的其他数据源中都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 OneDrive 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下该密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 OneDrive 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 OneDrive 数据源，您必须提供 OneDrive 凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置， OneDrive 请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 OneDrive

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择OneDrive 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的OneDrive 连接器。

5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. OneDrive 租户 ID-输入不带协议的 OneDrive 租户 ID。

   b. 身份验证类型 - 选择新建或现有。

   c. i. 如果您选择现有，请为选择密钥选择现有密钥。

      ii. 如果您选择新建，请在新的 AWS Secrets Manager 密钥部分中输入以下信息：

         A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-OneDrive-” 会自动添加到您的密钥名称中。

         B. 对于应用程序 ID 和应用程序密码-输入您 OneDrive 帐户中的身份验证凭据值，然后选择保存身份验证。

   d. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

   > **Note**
   >
   > IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

   e. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 根据您的使用案例，选择列表文件和名称列表。

      i. 如果选择列表文件，请输入以下信息：

         - 选择位置 - 输入 Amazon S3 存储桶的路径。

将用户列表文件添加到 Amazon S3-选择将您的用户列表文件添加到您的 Amazon S3 存储桶。

用户本地组映射 - 选择使用本地组映射来筛选您的内容。

ii. 如果选择名称列表，请输入以下信息：

- 用户名 - 输入最多 10 个要索引的用户驱动器。要添加 10 个以上的用户，请创建一个包含用户名的文件。

添加另一个 - 选择添加更多用户。

用户本地组映射 - 选择使用本地组映射来筛选您的内容。

b. 对于其他配置 - 添加正则表达式模式以包含或排除某些文件。最多可以添加 100 个模式。

c. 在 “同步运行计划” 中，“频率”-选择与数据源同步的频率。 Amazon Kendra

d. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 对于默认数据源字段和其他建议的字段映射-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 OneDrive

您必须使用 OneDriveConfigurationAPI 指定以下内容：

- 租户 ID - 指定组织的 Azure Active Directory 域。
- OneDrive 用户-指定应为其文档编制索引的用户帐户列表。
- 亚马逊秘密资源名称 (ARN)-提供包含您账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。 OneDrive 密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "username": "OAuth client ID",
    "password": "client secret"
```

```
}
```

- IAM ro le —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 OneDrive 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 OneDrive 数据源的IAM 角色。

您还可以添加以下可选功能：

- 包含和排除筛选条件 - 指定是包含还是排除文档。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 字段映射-选择将 OneDrive 数据源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 必须将数据源中的文档正文字段名称映射到索引字段名称`_document_body`。其他所有字段均为可选字段。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

微软 OneDrive 连接器 V2.0

Microsoft OneDrive 是一项基于云的存储服务，你可以用它来存储、共享和托管你的内容。您可以使用 Amazon Kendra 为 OneDrive数据源编制索引。

您可以使用Amazon Kendra 控制台和 OneDriveConfigurationAPI Amazon Kendra 连接到您的 OneDrive 数据源。

> **Note**
>
> 对 Conn OneDrive ector V1.0/ OneDriveConfiguration API 的支持计划于 2023 年 6 月结束。我们建议使用 OneDrive 连接器 V2.0/ TemplateConfiguration API。版本 2.0 提供了其他 ACL 和身份爬网程序功能。

要对 Amazon Kendra OneDrive 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明

支持的特征

Amazon Kendra OneDrive 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引 OneDrive 数据源之前，请在 OneDrive 和 AWS 帐户中进行这些更改。

在中 OneDrive，请确保你有：

- 已在 Office 365 中创建了一个 OneDrive 账户。
- 记下了您的 Microsoft 365 租户 ID。在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中可以找到您的租户 ID。
- 在 Azure 门户中创建了一个 OAuth 应用程序，并记下了用于使用密钥进行身份验证的客户端 AWS Secrets Manager ID 和客户端密钥或客户端凭据。有关更多信息，请参阅 Microsoft 教程和注册应用程序示例。

> **Note**
>
> 在 Azure 门户中创建或注册应用程序时，密钥 ID 代表实际的密钥值。在创建密钥和应用程序时，您必须立即记下或保存实际的密钥值。你可以通过在 Azure 门户中选择应用程序的名称，然后导航到证书和密钥上的菜单选项来访问你的密钥。
> 你可以通过在 Azure 门户中选择应用程序的名称，然后导航到概述页面来访问你的客户端 ID。应用程序（客户端）ID 是客户端 ID。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 使用 AD 应用程序 ID 在 AD 站点上注册应用程序的密钥。该密钥必须包含应用程序 ID 和密钥。
- 已复制组织的 AD 域。
- 在 Microsoft Graph 选项上为您的 AD 应用程序添加了以下权限：
  - 读取所有网站集中的文件（File.Read.All）
  - 读取所有用户的完整个人资料（User.Read.All）
  - 阅读所有组（Group.Read.All）
  - 读取所有笔记（Notes.Read.All）
- 复制必须为其文档编制索引的用户列表。您可以选择提供用户名列表，也可以在存储在 Amazon S3 中的文件中提供用户名。创建数据来源后，您可以：
  - 修改用户列表。
  - 从用户列表更改为存储在存储 Amazon S3 桶中的列表。
  - 更改用户列表的 Amazon S3 存储桶位置。如果您更改存储桶位置，则还必须更新数据源的 IAM 角色，使其能够访问存储桶。

> **Note**
>
> 如果您将用户名列表存储在 Amazon S3 存储桶中，则数据源的 IAM 策略必须提供对存储桶的访问权限以及对存储桶加密时使用的密钥（如果有）的访问权限。

OneDrive 连接器使用 O nedrive 用户属性中显示的 “来自联系人信息的电子邮件”。确保您要爬取其数据的用户在联系信息页面中配置了电子邮件字段，因为对于新用户，该字段可能为空。

在您的 AWS 账户中，请确保您有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM
- 将您的 OneDrive 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下该密钥的 ARN。

如果您没有现有的 IAM 角色或密钥，则可以在将 OneDrive 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 OneDrive 数据源，您必须提供 OneDrive 凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置 Amazon Kendra， OneDrive 请参阅先决条件。

Console

要连接 Amazon Kendra 到 OneDrive

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择OneDrive 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的OneDrive 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

b. （可选）说明 - 为数据来源输入说明。

c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

a. OneDrive 租户 ID-输入不带协议的 OneDrive 租户 ID。

b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

c. 在身份验证 中 - 选择新建或现有。

d. i. 如果您选择现有，请为选择密钥选择现有密钥。

ii. 如果您选择新建，请在新的 AWS Secrets Manager 密钥部分中输入以下信息：

A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-OneDrive-” 会自动添加到您的密钥名称中。

B. 对于客户端 ID 和客户机密钥-输入客户端 ID 和客户机密钥。

e. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

f. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

g. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

h. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

8. a. 对于同步范围-选择要为哪些用户 OneDrive 的数据编制索引。您可以手动添加最多 10 个用户。

   b. 对于其他配置 - 添加正则表达式模式以包含或排除某些内容。最多可以添加 100 个模式。

   c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

      - 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
      - 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
      - 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

   d. 在同步运行计划中，对于频率-选择同步数据源内容和更新索引的频率。

   e. 选择下一步。

9. 在设置字段映射页面上，请输入以下信息：

   a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

   b. 选择下一步。

10. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 OneDrive

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构`ONEDRIVEV2`时的类型。还要像调用 CreateDataSourceAPI `TEMPLATE` 时一样指定数据源。
- 租户 ID - 指定 Microsoft 365 租户 ID。在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中可以找到您的租户 ID。

- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅为新增、修改和删除的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。 OneDrive

  如果您使用 OAuth 2.0 身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "clientId": "client ID",
    "clientSecret": "client secret"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 OneDrive 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 OneDrive 数据源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（ VPC ）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除筛选器-您可以指定是包含还是排除某些文件、 OneNote 章节和 OneNote 页面。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。
- 字段映射-只能映射连接器的内置或常用索引字段。 Amazon Kendra OneDrive 由于 API 限制，自定义字段映射不适用于 OneDrive 连接器。有关更多信息，请参阅映射数据来源字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅OneDrive 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与 OneDrive 数据源集成的更多信息，请参阅：

- 宣布推出更新的 Microsoft OneDrive 连接器 (V2)。 Amazon Kendra

微软 SharePoint

SharePoint 是一项协作建站服务，可用于自定义 Web 内容以及创建页面、网站、文档库和列表。您可以使用索 Amazon Kendra 引您的 SharePoint 数据源。

Amazon Kendra 目前支持 SharePoint 在线版和 SharePoint服务器版（版本 2013、2016、2019 和订阅版）。

您可以使用Amazon Kendra 控制台、API 或 TemplateConfigurationAPI 连接到 Amazon Kendra 您的 SharePoint 数据源。SharePointConfiguration

Amazon Kendra 有两个版本的 SharePoint 连接器。每个版本支持的功能包括：

SharePoint 连接器 V1.0/API SharePointConfiguration

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 更改日志

- Virtual Private Cloud (VPC)

SharePoint 连接器 V2.0/API TemplateConfiguration

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

> **Note**
>
> 对 SharePoint 连接器 V1.0/ SharePointConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 SharePoint 连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra SharePoint 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- SharePoint 连接器 V1.0
- SharePoint 连接器 V2.0

SharePoint 连接器 V1.0

SharePoint 是一项协作建站服务，可用于自定义 Web 内容以及创建页面、网站、文档库和列表。如果您是 SharePoint 用户，则可以使用索 Amazon Kendra 引您的 SharePoint 数据源。

> **Note**
>
> 对 SharePoint 连接器 V1.0/ SharePointConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 SharePoint 连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra SharePoint 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 更改日志
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引 SharePoint 数据源之前，请在 SharePoint 和 AWS 帐户中进行这些更改。

您需要提供身份验证凭证，这些凭据可以安全地存储在 AWS Secrets Manager 密钥中。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

在中 SharePoint，请确保你有：

- 记下了您要编制索引的 SharePoint 网站的网址。
- SharePoint 在线版：
  - 已记下您的基本身份验证凭证，其中包含具有站点管理员权限的用户名和密码。
  - 可选：生成的 OAuth 2.0 凭证，其中包含用户名、密码、客户端 ID 和客户端密钥。
  - 使用管理用户在 Azure 门户中停用安全默认值。有关在 Azure 门户中管理安全默认设置的更多信息，请参阅 Microsoft 关于如何启用/禁用安全默认设置的文档。
- 对于 SharePoint 服务器：

- 记下您的 SharePoint 服务器域名（活动目录中的 NetBIOS 名称）。您可以使用它以及您的 SharePoint 基本身份验证用户名和密码将 SharePoint 服务器连接到 Amazon Kendra。

> **Note**
>
> 如果您使用 SharePoint 服务器并且需要将访问控制列表 (ACL) 转换为电子邮件格式以便根据用户上下文进行筛选，请提供 LDAP 服务器 URL 和 LDAP 搜索库。或者，您也可以使用目录域覆盖。LDAP 服务器 URL 是完整的域名和端口号（例如，ldap://example.com:389）。LDAP 搜索库是域控制器的“example”和“com”。凭借目录域覆盖，您可以使用电子邮件域来代替 LDAP 服务器 URL 和 LDAP 搜索库。例如，username@example.com 的电子邮件域名是“example.com”。如果您不想验证域名，而只想使用您的电子邮件域名，则可以使用此替代方法。

- 为您的 SharePoint 账户添加了以下权限：

  对于 SharePoint 清单

  - 打开项目 - 使用服务器端文件处理程序查看文档的来源。
  - 查看应用程序页面 - 查看表单、视图和应用程序页面。枚举列表。
  - 查看项目 - 查看列表中的项目和文档库中的文档。
  - 查看版本 - 查看列表项或文档的过去版本。

  对于 SharePoint 网站

  - 浏览目录-使用 SharePoint 设计器和 Web DAV 界面枚举网站中的文件和文件夹。
  - 浏览用户信息 - 查看有关网站用户的信息。
  - 枚举权限 - 枚举网站、列表、文件夹、文档或列表项的权限。
  - 打开 - 打开网站、列表或文件夹以访问容器内的项目。
  - 使用客户端集成功能-使用 SOAP、WebDAV、客户端对象模型或 SharePoint 设计器界面访问网站。
  - 使用远程接口 - 使用启动客户端应用程序的功能。
  - 查看页面 - 查看网站上的页面。
- 已选中每个文档在您计划用于同一索引的其他数据源中 SharePoint 以及其他数据源中都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 SharePoint 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下该密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 SharePoint 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 SharePoint 数据源，您必须提供 SharePoint 凭据的详细信息，Amazon Kendra 以便访问您的数据。如果您尚未进行配置， SharePoint 请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 SharePoint

1. 登录 AWS 管理控制台并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择SharePoint 连接器 v1.0，然后选择添加数据源。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 对于托管方式-在 “SharePoint 联机” 和 “SharePoint服务器” 之间进行选择。

      i. 对于SharePoint联机-输入特定于您的 SharePoint存储库的站点 URL。

      ii. 对于SharePoint服务器-选择您的SharePoint 版本，输入特定于 SharePoint 存储库的站点 URL，然后输入 SSL 证书位置的 Amazon S3 路径。

   b. （仅限SharePoint 服务器）对于 Web 代理-输入内部 SharePoint 实例的主机名和端口号。端口号应为介于 0 到 65535 之间的数值。

   c. 对于身份验证 - 根据您的使用案例选择以下选项：

      i. 对于 SharePoint 联机-在基本身份验证和 OAuth 2.0 身份验证之间进行选择。

      ii. 对于 SharePoint 服务器-在 “无”、“LDAP” 和 “手动” 之间进行选择。

   d. fo AWS Secrets Manager r secre t-选择现有密钥或创建新 Secrets Manager 密钥来存储您的 SharePoint身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。您必须输入密钥名称。前缀 “AmazonKendra-SharePoint-” 会自动添加到您的密钥名称中。

   e. 在创建 AWS Secrets Manager 密钥窗口中输入其他信息：

      i. 根据您的用例，从以下 SharePoint Cloud 身份验证选项中进行选择：

         A. 基本身份验证-输入您的 SharePoint 帐户用户名作为用户名，将 SharePoint 帐户密码输入为密码。

         B. OAuth 2.0 身份验证 -输入您的 SharePoint 账户用户名作为用户名，将SharePoint帐户密码输入为密码，将自动生成的唯一 SharePoint ID 作为客户端

ID， SharePoint 以及 Amazon Kendra 两者使用的共享密钥字符串作为客户机密。

ii. 根据您的用例，从以下 SharePoint服务器身份验证选项中进行选择：

A. 无-输入您的 SharePoint 帐户用户名作为用户名，将您的 SharePoint 帐户密码输入为密码，并输入您的服务器域名。

B. LDAP —*#### SharePoint ############ SharePoint############## LDAP ################ ldap: //example.com: 389#### LDAP ## #####dc=example#dc=com##*

C. 手动-输入您的 SharePoint 帐户用户名作为用户名，将您的 SharePoint 帐户密码输入为密码，并输入您的电子邮件域覆盖（目录用户或群组的电子邮件域）。

iii. 选择保存。

f. 虚拟私有云（VPC）- 您还必须添加子网和 VPC 安全组。

> **Note**
>
> 如果您使用 SharePoint 服务器，则必须使用 VPC。 Amazon VPC 对于其他 SharePoint版本是可选的。

g. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

h. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 使用更改日志 - 选择更新索引，而不是同步所有文件。

b. 爬取附件 - 选择此选项可爬取附件。

c. 使用本地组映射 - 选择此选项可确保正确筛选文档。

d. 其他配置 - 添加正则表达式模式以包含或排除某些文件。最多可以添加 100 个模式。

e. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

f. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. Amazon Kendra 默认字段映射-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 对于自定义字段映射 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 SharePoint

必须使用 SharePointConfigurationAPI 指定以下内容：

- SharePoint版本-指定配置 SharePoint时使用的 SharePoint版本。无论你使用的是 Server 2013、SharePoint Server 2016、S SharePoint erver 2019 还是 O SharePoint nlin SharePoint e，情况都是如此。
- 亚马逊秘密资源名称 (ARN)-提供包含您在 SharePoint 账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥存储在 JSON 结构中。

  对于SharePoint 在线基本身份验证，以下是您的密钥中必须包含的最低 JSON 结构：

```
{
    "userName": "user name",
    "password": "password"
}
```

  对于SharePoint 在线 OAuth 2.0 身份验证，以下是您的密钥中必须包含的最低 JSON 结构：

```
{
    "userName": "SharePoint account user name"",
    "password": "SharePoint account password",
    "clientId": "SharePoint auto-generated unique client id",
    "clientSecret": "secret string shared by Amazon Kendra and SharePoint to
 authorize communications"
}
```

  对于SharePoint 服务器基本身份验证，以下是您的密钥中必须包含的最低 JSON 结构：

```
{
    "userName": "user name",
    "password": "password",
    "domain": "server domain name"
}
```

对于SharePoint 服务器 LDAP 身份验证（如果您需要将访问控制列表 (ACL) 转换为电子邮件格式以便根据用户上下文进行筛选，则可以在密钥中包含 LDAP 服务器 URL 和 LDAP 搜索库），以下是您的密钥中必须包含的最低 JSON 结构：

```
{
    "userName": "user name",
    "password": "password",
    "domain": "server domain name"
    "ldapServerUrl": "ldap://example.com:389",
    "ldapSearchBase": "dc=example,dc=com"
}
```

对于SharePoint 服务器手动身份验证，以下是您的密钥中必须包含的最低 JSON 结构：

```
{
    "userName": "user name",
    "password": "password",
    "domain": "server domain name",
    "emailDomainOverride": "example.com"
}
```

- IAM ro le —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 SharePoint 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 SharePoint 数据源的IAM 角色。
- Amazon VPC— 如果您使用 SharePoint Server`VpcConfiguration`，请在数据源配置中指定。请参阅配置 Amazon Kendra 以使用 VPC。

您还可以添加以下可选功能：

- Web 代理-是否通过 Web 代理连接到您的 SharePoint 网站 URL。此选项只能用于 SharePoint服务器。
- 索引列表-是否 Amazon Kendra 应将附件的内容编入 SharePoint 列表项的索引。

- 更改日志-是否 Amazon Kendra 应使用 SharePoint 数据源更改日志机制来确定是否必须更新索引中的文档。

  > **Note**
  >
  > 如果您不想让 Amazon Kendra 扫描所有文档，请使用更改日志。如果您的更改日志很大，则扫描 SharePoint 数据源中的文档所花费的时间可能比处理更改日志所需的时间 Amazon Kendra 少。如果您是首次将 SharePoint数据源与索引同步，则会扫描所有文档。

- 包含和排除筛选条件 - 您可以指定是包含还是排除某些内容。

  > **Note**
  >
  > 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 字段映射-选择将 SharePoint 数据源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

  > **Note**
  >
  > 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

了解更多信息

要了解有关 Amazon Kendra 与 SharePoint数据源集成的更多信息，请参阅：

- Amazon Kendra SharePoint 在线连接器入门

SharePoint 连接器 V2.0

SharePoint 是一项协作建站服务，可用于自定义 Web 内容以及创建页面、网站、文档库和列表。您可以使用索 Amazon Kendra 引您的 SharePoint 数据源。

Amazon Kendra 目前支持 SharePoint 在线版和 SharePoint服务器版（2013 年、2016 年、2019 年和订阅版）。

> **Note**
>
> 对 SharePoint 连接器 V1.0/ SharePointConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 SharePoint 连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra SharePoint 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

Amazon Kendra SharePoint 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引 SharePoint 数据源之前，请在 SharePoint 和 AWS 帐户中进行这些更改。

您需要提供身份验证凭证，这些凭据可以安全地存储在 AWS Secrets Manager 密钥中。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

在 “在 SharePoint 线” 中，请确保你有：

- 已复制您的 SharePoint 实例 URL。您输入的主机 URL 的格式为 *https://yourdomain.sharepoint.com/sites/mysite*。您的 URL 必须以 `https` 开头并包含 `sharepoint.com`。
- 已复制您的 SharePoint 实例 URL 的域名。
- 已记下您的基本身份验证凭据，其中包含用户名和密码，具有连接到 On SharePoint line 的站点管理员权限。
- 使用管理用户在 Azure 门户中停用安全默认值。有关在 Azure 门户中管理安全默认设置的更多信息，请参阅 Microsoft 关于如何启用/禁用安全默认设置的文档。
- 已在您的 SharePoint 账户中停用多因素身份验证 (MFA)，这样就不会阻止 Amazon Kendra 它抓取您的内容。 SharePoint
- 如果使用基本身份验证以外的身份验证类型：已复制 SharePoint 实例的租户 ID。有关如何查找租户 ID 的详细信息，请参阅查找您的 Microsoft 365 租户 ID。
- 如果你需要使用 Microsoft Entra 迁移到云用户身份验证，请参阅微软关于云身份验证的文档。
- 对于 OAuth 2.0 身份验证和 OAuth 2.0 刷新令牌身份验证：请注意您的基本身份验证凭据，其中包含用于连接到 On SharePoint line 的用户名和密码以及注册 SharePoint 到 Azure AD 后生成的客户端 ID 和客户端密钥。
  - 如果您未使用 ACL，请添加以下权限：

Microsoft Graph	SharePoint
• Notes.Read.All（应用程序）-阅读所有笔记本 OneNote • Sites.Read.All（应用程序）- 读取所有网站集合的项目	• AllSites.Read（委托）-读取所有网站集中的项目

> **Note**
>
> Note.Read.All 和 Sites.Read.All 只有在你想要抓取 Documents 时才是必填的。 OneNote 如果您想抓取特定的网站，则权限可以仅限于特定的网站，而不是域中所有可用的网站。您可以配置 “站点”。选定（应用程序）权限。有了这个 API 权限，你需要通过 Microsoft Graph API 明确设置每个网站的访问权限。有关更多信息，请参阅 Microsoft 在 “站点” 上发布的博客。所选权限。

- 如果您使用 ACL，请添加以下权限：

Microsoft Graph	SharePoint
• Group.Member.Read.All（应用程序）- 读取所有组成员资格 • Notes.Read.All（应用程序）-阅读所有笔记本 OneNote • 网站。 FullControl.All（已授权）-需要检索文档的 ACL • Sites.Read.All（应用程序）- 读取所有网站集合的项目 • User.Read.All（应用程序）- 读取所有用户的完整个人资料	• AllSites.Read（委托）-读取所有网站集中的项目

> **Note**
>
> GroupMember只有激活 Identity Crawler 时，才需要.Read.All 和 User.Read.All。
> 如果您想抓取特定的网站，则权限可以仅限于特定的网站，而不是域中所有可用的网站。您可以配置 “站点”。选定（应用程序）权限。有了这个 API 权限，你需要通过 Microsoft Graph API 明确设置每个网站的访问权限。有关更多信息，请参阅 Microsoft 在 “站点” 上发布的博客。所选权限。

- 对于 Azure AD 仅限应用程序的身份验证：私钥和注册 SharePoint 到 Azure AD 后生成的客户端 ID。还要注意 X.509 证书。
  - 如果您未使用 ACL，请添加以下权限：

SharePoint
• Sites.Read.All（应用程序）-需要访问所有网站集中的项目和列表

> **Note**
>
> 如果您想抓取特定的网站，则权限可以仅限于特定的网站，而不是域中所有可用的网站。您可以配置 “站点”。选定（应用程序）权限。有了这个 API 权限，你需要通过 Microsoft Graph API 明确设置每个网站的访问权限。有关更多信息，请参阅 Microsoft 在 “站点” 上发布的博客。所选权限。

- 如果您使用 ACL，请添加以下权限：

SharePoint
• 网站。 FullControl.All（应用程序）-需要检索文档的 ACL

> **Note**
>
> 如果您想抓取特定的网站，则权限可以仅限于特定的网站，而不是域中所有可用的网站。您可以配置 “站点”。选定（应用程序）权限。有了这个 API 权限，你需要通过 Microsoft Graph API 明确设置每个网站的访问权限。有关更多信息，请参阅 Microsoft 在 “站点” 上发布的博客。所选权限。

- 对于 SharePoint 仅限应用程序的身份验证：记下在授予 “仅限 SharePoint 应用程序” 权限时生成的客户端 ID 和客户端密钥，以及向 Azure AD 注册 SharePoint 应用程序时生成的客户端 ID 和客户端密钥。 SharePoint

> **Note**
>
> SharePoint SharePoint 2013 版本不支持仅限应用程序的身份验证。

- （可选）如果您正在搜索 OneNote文档并使用 Id entity Crawler，请添加以下权限：

Microsoft Graph
• GroupMember.Read.All（应用程序）-读取所有群组成员资格 • Notes.Read.All（应用程序）-阅读所有笔记本 OneNote • Sites.Read.All（应用程序）- 读取所有网站集合的项目 • User.Read.All（应用程序）- 读取所有用户的完整个人资料

> **Note**
>
> 使用基本身份验证和仅限 SharePoint应用程序的身份验证来抓取实体不需要 AP I 权限。

在 SharePoint 服务器中，请确保你有：

- 已复制您的 SharePoint 实例 URL 和 SharePoint URL 的域名。您输入的主机 URL 的格式为 *`https://yourcompany/sites/mysite`*。您的 URL 必须以 `https` 开头。

> **Note**
>
> （本地/服务器）Amazon Kendra 会检查中 AWS Secrets Manager 包含的端点信息是否与数据源配置详细信息中指定的端点信息相同。这有助于防止出现混淆代理人问题，这是一个安全问题，即用户无权执行操作，但可以将 Amazon Kendra 作为代理来访问配置的密钥和执行操作。如果以后更改端点信息，则必须创建一个新密钥来同步此信息。

- 已在您的 SharePoint 账户中停用多因素身份验证 (MFA)，这样就不会阻止 Amazon Kendra 它抓取您的内容。 SharePoint
- 如果使用SharePoint 仅限应用程序的身份验证进行访问控制：

- 已复制您在站点级别注册 “仅限应用程序” 时生成的 SharePoint 客户端 ID。客户端 ID 格式为 ClientId @ TenantId。例如，*ffa956f3-8f89-44e7-b0e4-49670756342c@888d0b57-69f1-4fb8-957f-e1f0bedf82fe*。
- 已复制您在站点级别注册 “仅限应用程序” 时生成的 SharePoint 客户端密钥。

注意：由于只有在注册 SharePoint 服务器进行仅限应用程序身份验证时，才会为单个站点生成客户端 ID 和客户端密钥，因此仅 SharePoint 应用程序身份验证仅支持一个站点 URL。

> **Note**
>
> SharePoint SharePoint 2013 版本不支持仅限应用程序的身份验证。

- 如果使用带有自定义域名的电子邮件 ID 进行访问控制：
  - 记下了您的自定义电子邮件域名值，例如：*“amazon.com”*。
- 如果使用带有来自 IDP 的域名的电子邮件 ID 身份验证，请复制您的：
  - LDAP 服务器端点（LDAP 服务器的端点，包括协议和端口号）。例如：*ldap://example.com:389*。
  - LDAP 搜索库（LDAP 用户的搜索库）。例如：*CN=Users,DC=sharepoint,DC=com*。
  - LDAP 用户名和 LDAP 密码。
- 要么配置了 NTLM 身份验证凭据，要么配置了包含用户名（SharePoint 帐户用户名）和密码（SharePoint 帐户密码）的 Kerberos 身份验证凭据。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 SharePoint 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下该密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 SharePoint 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 SharePoint 数据源，您必须提供 SharePoint 凭据的详细信息，Amazon Kendra 以便访问您的数据。如果您尚未进行配置， SharePoint 请参 Amazon Kendra 阅先决条件。

Console: SharePoint Online

连接 Amazon Kendra 到 “在 SharePoint 线”

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

   > **Note**
   >
   > 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择SharePoint 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的SharePoint 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

a. 托管方式-选择SharePoint 在线。

b. 特定于您的 SharePoint存储库的站点 URL-输入 SharePoint 主机 URL。您输入的主机 URL 的格式为 *https://yourdomain.sharepoint.com/sites/mysite*。URL 必须以 `https` 协议开头。使用分行符分隔 URL。最多可以添加 100 个 URL。

c. 域-输入 SharePoint 域。例如，URL *https://yourdomain.sharepoint.com/sites/mysite* 中的域名是 *yourdomain*。

d. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

你也可以选择用户 ID 的类型，无论是用户主体名称还是从 Azure 门户获取的用户电子邮件。如果您未指定，则默认使用电子邮件。

e. 身份验证-选择基本、OAuth 2.0、仅限 Azure AD 应用程序身份验证、仅限应用程序身份验证或 O SharePoint Auth 2.0 刷新令牌身份验证。您可以选择现有 AWS Secrets Manager 密钥来存储您的身份验证凭证，也可以创建密钥。

i. 如果使用基本身份验证，则您的密钥必须包含机密名称、 SharePoint 用户名和密码。

ii. 如果使用 OAuth 2.0 身份验证，则您的密钥必须包括 SharePoint 租户 ID、密钥名称、 SharePoint 用户名、密码、在 Azure AD 中注册时生成的 Azure AD 客户端 ID 以及 SharePoint 在 Azure AD SharePoint 中注册时生成的 Azure AD 客户端密钥。

iii. 如果使用仅限 Azure AD 应用程序的身份验证，则您的密钥必须包括 SharePoint 租户 ID、Azure AD 自签名 X.509 证书、密钥名称、在 Azure AD SharePoint 中注册时生成的 Azure AD 客户端 ID 以及用于对 Azure AD 连接器进行身份验证的私钥。

iv. 如果使用SharePoint仅限应用程序的身份验证，则密钥必须包括 SharePoint 租户 ID、密钥名称、在租户级别注册仅限应用程序时生成的 SharePoint 客户端 ID、在租户级别注册仅限应用程序时生成的 SharePoint 客户端密钥、在 Azure AD SharePoint 中注册时生成的 Azure AD 客户端 ID 以及注册 SharePoint 到 Azure AD 时生成的 Azure AD 客户端密钥。

SharePoint 客户端 ID 格式为 *clientId@ TenantId*。例如，*ffa956f3-8f89-44e7-b0e4-49670756342c@888d0b57-69f1-4fb8-957f-e1f0bedf82fe*。

v. 如果使用 OAuth 2.0 刷新令牌身份验证，则您的密钥必须包括 SharePoint 租户 ID、密钥名称、在 Azure AD SharePoint 中注册时生成的唯一 Azure AD 客户端 ID、注册到 Azure AD 时生成的 Azure AD 客户端密钥、 SharePoint 为连接 Amazon Kendra 而生成的刷新令牌。 SharePoint

f. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

g. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

你也可以选择抓取本地组映射或 Azure Active Directory 组映射。

> **Note**
>
> AD 组映射抓取仅适用于 OAuth 2.0、OAuth 2.0 刷新令牌和仅限应用程序的身份验证。 SharePoint

h. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

i. 选择实体 - 选择要爬取的实体。您可以选择爬取所有实体或文件、附件、链接、页面、事件、备注和列表数据的任意组合。

ii. 在其他配置中，对于实体正则表达式模式 - 为链接、页面和事件添加正则表达式模式以包含特定实体，而不是同步所有文档。

iii. 正则@@ 表达式模式-添加正则表达式模式，通过文件路径、文件名、文件类型、OneNote 章节名称和OneNote 页面名称来包含或排除文件，而不是同步所有文档。最多可以添加 100 个。

> **Note**
>
> OneNote 抓取功能仅适用于 OAuth 2.0、OAuth 2.0 刷新令牌和仅限应用程序的身份验证。 SharePoint

b. 对于同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次将数据来源与 Amazon Kendra 同步时，所有内容都会默认同步。

- 完全同步 - 无论之前的同步状态如何，都同步所有内容。
- 同步新增或修改的文档 - 仅同步新增或修改的文档。
- 同步新增、修改或删除的文档 - 仅同步新增、修改和删除的文档。

c. 在同步运行计划中，对于频率-选择同步数据源内容和更新索引的频率。

d. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

Console: SharePoint Server

要连接 Amazon Kendra 到 SharePoint

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。

4. 在添加数据源页面上，选择SharePoint 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的SharePoint 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 托管方法-选择SharePoint服务器。
   b. 选择 SharePoint版本-选择 SharePoint 2013 年、SharePoint 201 6 年、201 SharePoint 9 年和 SharePoint （订阅版）。
   c. 特定于您的 SharePoint存储库的站点 URL-输入 SharePoint 主机 URL。您输入的主机 URL 的格式为 *`https://yourcompany/sites/mysite`*。URL 必须以 `https` 协议开头。使用分行符分隔 URL。最多可以添加 100 个 URL。
   d. 域-输入 SharePoint 域。例如，URL *`https://yourcompany/sites/mysite`* 中的域名是 *`yourcompany`*。
   e. SSL 证书位置-输入 SSL 证书文件的 Amazon S3 路径。
   f. （可选）对于 Web 代理 - 输入主机名（不带 `http://` 或 `https://` 协议）和主机 URL 传输协议使用的端口号。端口号的数值必须介于 0 和 65535 之间。
   g. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

      对于 SharePoint 服务器，您可以从以下 ACL 选项中进行选择：

      i. 域名来自 IDP 的电子邮件 ID-用户 ID 基于电子邮件 ID，其域名从底层身份提供商 (IDP) 获取。作为身份验证的一部分，您在 Secrets Manager 密钥中提供 IDP 连接详细信息。

ii. 带有自定义域的电子邮件 ID-用户 ID 基于自定义电子邮件域值。例如，*“amazon.com”*。电子邮件域名将用于构造用于访问控制的电子邮件 ID。您必须输入您的自定义电子邮件域名。

iii. 域\ 带域的用户-用户 ID 是使用 “域\ 用户 ID” 格式构造的。您需要提供有效的域名。例如：*“sharepoint2019”*，以便构造访问控制。

h. 对于身份验证，请选择 SharePoint 仅限应用程序身份验证、NTLM 身份验证或 Kerberos 身份验证。您可以选择现有 AWS Secrets Manager 密钥来存储您的身份验证凭证，也可以创建密钥。

i. 如果使用 NTLM 身份验证或 Kerberos 身份验证，则您的密钥必须包含密钥名称、用户名和密码。

如果使用带有来自 IDP 的域名的电子邮件 ID，请同时输入您的：

- LDAP 服务器端点 - LDAP 服务器的端点，包括协议和端口号。例如：*ldap://example.com:389*。
- LDAP 搜索库 - LDAP 用户的搜索库。例如：*CN=Users,DC=sharepoint,DC=com*。
- LDAP 用户名 - 您的 LDAP 用户名。
- LDAP 密码 - 您的 LDAP 密码。

ii. 如果使用SharePoint 仅限应用程序的身份验证，则您的密钥必须包含密钥名称，即您在站点级别注册仅限应用程序时生成的客户 SharePoint 端 ID，在站点级别注册仅限应用程序时生成的 SharePoint 客户端密钥。

SharePoint 客户端 ID 格式为 *clientId@ TenantId*。例如，*ffa956f3-8f89-44e7-b0e4-49670756342c@888d0b57-69f1-4fb8-957f-e1f0bedf82fe*。

注意：由于只有在注册 SharePoint 服务器进行仅限应用程序身份验证时，才会为单个站点生成客户端 ID 和客户端密钥，因此仅 SharePoint 应用程序身份验证仅支持一个站点 URL。

如果使用带有来自 IDP 的域名的电子邮件 ID，请同时输入您的：

- LDAP 服务器端点 - LDAP 服务器的端点，包括协议和端口号。例如：*ldap://example.com:389*。

- LDAP 搜索库 - LDAP 用户的搜索库。例如：*CN=Users,DC=sharepoint,DC=com*。
- LDAP 用户名 - 您的 LDAP 用户名。
- LDAP 密码 - 您的 LDAP 密码。

i. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

j. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

你也可以选择抓取本地组映射或 Azure Active Directory 组映射。

> **Note**
>
> AD 组映射抓取仅适用于 SharePoint 应用程序身份验证。

k. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

l. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

i. 选择实体 - 选择要爬取的实体。您可以选择爬取所有实体或文件、附件、链接、页面、事件和列表数据的任意组合。

ii. 在其他配置中，对于实体正则表达式模式 - 为链接、页面和事件添加正则表达式模式以包含特定实体，而不是同步所有文档。

iii. 正则@@ 表达式模式-添加正则表达式模式，通过文件路径文件名文件类型、OneNote章节名称和OneNote页面名称来包含或排除文件，而不是同步所有文档。最多可以添加 100 个。

> **Note**
>
> OneNote 抓取仅适用于 “仅限 SharePoint 应用程序” 的身份验证。

b. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改后的全新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

c. 在同步运行计划中，对于频率-选择同步数据源内容和更新索引的频率。

d. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 SharePoint

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构SHAREPOINTV2时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。

- 存储库端点元数据-指定 SharePoint 实例`siteUrls`的`tenantID`和`domain`。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - `FORCED_FULL_CRAWL`对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - `FULL_CRAWL`每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - `CHANGE_LOG`每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

  > **Note**
  >
  > 只有当您将设置为时，身份搜寻器才可用`crawlAcl`。`true`

- 存储库其他属性 - 指定：
  - （适用于 Azure AD）`s3bucketName`，`s3certificateName`你可以用来存储 Azure AD 自签名 X.509 证书。
  - 您使用的身份验证类型 (`auth_Type`)，是`OAuth2`OAuth2App、`OAuth2Certificate`、`Basic`、`OAuth2_RefreshToken`、`NTLM`、和`Kerberos`。
  - 您使用的版本 (`version`)，`Server`无论是`Online`。如果使用 `Server`，您可以进一步将 `onPremVersion` 指定为 `2013`、`2016`、`2019` 或 `SubscriptionEdition`。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。 SharePoint

  如果您使用 SharePoint 联机，则可以在基本身份验证、OAuth 2.0、仅限 Azure AD 应用程序身份验证和 SharePoint 仅限应用程序身份验证之间进行选择。以下是每个身份验证选项的密钥中必须包含的最简单 JSON 结构：

- 基本身份验证

```
{
    "userName": "SharePoint account user name",
    "password": "SharePoint account password"
}
```

- OAuth 2.0 身份验证

```
{
    "clientId": "client id generated when registering SharePoint with Azure AD",
    "clientSecret": "client secret generated when registering SharePoint with
 Azure AD",
    "userName": "SharePoint account user name",
    "password": "SharePoint account password"
}
```

- 仅限 Azure AD 应用程序的身份验证

```
{
    "clientId": "client id generated when registering SharePoint with Azure AD",
    "privateKey": "private key to authorize connection with Azure AD"
}
```

- SharePoint 仅限应用程序的身份验证

```
{
    "clientId": "client id generated when registering SharePoint for App Only at
 Tenant Level",
    "clientSecret": "client secret generated when registering SharePoint for App
 Only at Tenant Level",
    "adClientId": "client id generated while registering SharePoint with Azure
 AD",
    "adClientSecret": "client secret generated while registering SharePoint with
 Azure AD"
}
```

- OAuth 2.0 刷新令牌身份验证

```
{
    "clientId": "client id generated when registering SharePoint with Azure AD",
```

```
    "clientSecret": "client secret generated when registering SharePoint with
 Azure AD",
    "refreshToken": "refresh token generated to connect to SharePoint"
}
```

如果您使用 SharePoint 服务器，则可以在 SharePoint 仅限应用程序身份验证、NTLM 身份验证和 Kerberos 身份验证之间进行选择。以下是每个身份验证选项的密钥中必须包含的最简单 JSON 结构：

- SharePoint 仅限应用程序的身份验证

```
{
    "siteUrlsHash": "Hash representation of SharePoint site URLs",
    "clientId": "client id generated when registering SharePoint for App Only at
 Site Level",
    "clientSecret": "client secret generated when registering SharePoint for App
 Only at Site Level"
}
```

- SharePoint 使用来自 IDP 授权的域进行仅限应用程序的身份验证

```
{
    "siteUrlsHash": "Hash representation of SharePoint site URLs",
    "clientId": "client id generated when registering SharePoint for App Only at
 Site Level",
    "clientSecret": "client secret generated when registering SharePoint for App
 Only at Site Level",
    "ldapUrl": "LDAP Account url eg. ldap://example.com:389",
    "baseDn": "LDAP Account base dn eg. CN=Users,DC=sharepoint,DC=com",
    "ldapUser": "LDAP account user name",
    "ldapPassword": "LDAP account password"
}
```

- （仅限服务器）NTLM 或 Kerberos 身份验证

```
{
    "siteUrlsHash": "Hash representation of SharePoint site URLs",
    "userName": "SharePoint account user name",
    "password": "SharePoint account password"
}
```

- （仅限服务器）带有来自 IDP 授权的域名的 NTLM 或 Kerberos 身份验证

```
{
    "siteUrlsHash": "Hash representation of SharePoint site URLs",
    "userName": "SharePoint account user name",
    "password": "SharePoint account password",
    "ldapUrl": "ldap://example.com:389",
    "baseDn": "CN=Users,DC=sharepoint,DC=com",
    "ldapUser": "LDAP account user name",
    "ldapPassword": "LDAP account password"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 SharePoint 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 SharePoint 数据源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（ VPC ）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器-您可以指定是包含还是排除某些文件和其他内容。 OneNotes

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 字段映射-选择将 SharePoint 数据源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅SharePoint 模板架构。

注意

- 连接器仅支持文件实体的自定义字段映射。
- 对于所有 SharePoint 服务器版本，ACL 令牌必须为小写。对于带有来自 IDP 的域名的电子邮件和带有自定义域 ACL 的电子邮件 ID ACL，例如：*user@sharepoint2019.com*。对于域名\带有域名的用户，例如：*sharepoint2013\user*。
- 连接器不支持 SharePoint 2013 年的更改日志模式/ 新内容或修改内容同步。
- 如果实体名称中包含字 % 符，则由于 API 限制，连接器将跳过这些文件。
- OneNote 只能由连接器使用租户 ID 进行抓取，并启用 OAuth 2.0、OAuth 2.0 刷新令牌或仅限 SharePoint 应用程序的身份验证以进行联机。 SharePoint
- 连接器仅使用 OneNote 文档的默认名称抓取文档的第一部分，即使文档已重命名也是如此。
- 连接器会抓取 SharePoint 2019 年、 SharePoint在线版和订阅版中的链接，前提是除了链接之外还选择 “页面” 和 “文件” 作为要抓取的实体。
- 如果选择链接作为要抓取的实体，则连接器会在 SharePoint SharePoint 2013 年和 2016 年抓取链接。
- 仅当列表数据也被选为要爬取的实体时，连接器才会爬取列表附件和评论。
- 仅当事件也被选为要爬取的实体时，连接器才会爬取事件附件。
- 对于 SharePoint 在线版本，ACL 令牌将使用小写。例如，如果 Azure 门户中的用户主体名称为 *MaryMajor@domain .com*，则 SharePoint 连接器中的 ACL 令牌将*# marymajor@domain.com*。
- 在适用于 SharePoint 在线和服务器的 Ident ity Crawler 中，如果要抓取嵌套群组，则必须激活本地和 AD 组抓取。
- 如果你使用的是 SharePoint 联机，并且 Azure 门户中的用户主体名称是大写和小写的组合，那么 SharePoint API 会在内部将其转换为小写。因此， Amazon Kendra SharePoint 连接器以小写形式设置 ACL。

Microsoft SQL Server

Microsoft SQL Server 是 Microsoft 开发的一个关系数据库管理系统（RDBMS）。如果您是Microsoft SQL Server用户，则可以使用索 Amazon Kendra 引您的Microsoft SQL Server数据源。 Amazon Kendra Microsoft SQL Server数据源连接器支持 MS SQL Server 2019。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的Microsoft SQL Server数据源。

要对 Amazon Kendra Microsoft SQL Server数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引Microsoft SQL Server数据源之前，请在Microsoft SQL Server和 AWS 帐户中进行这些更改。

在 Microsoft SQL Server 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 作为最佳实践，请提供只读 Amazon Kendra 的数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。
- 在 Microsoft SQL Server 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Microsoft SQL Server 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将Microsoft SQL Server数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的Microsoft SQL Server数据源，您必须提供Microsoft SQL Server凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置，Microsoft SQL Server请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 Microsoft SQL Server

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择Microsoft SQL Server连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的Microsoft SQL Server连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 在源中，输入以下信息：

   b. 主机 - 输入数据库主机名。

   c. 端口 - 输入数据库端口。

   d. 实例 - 输入数据库实例。

   e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

   f. 在身份验证中 - 请输入以下信息：

      - AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的Microsoft SQL Server身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

        A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

           I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Microsoft SQL Server-” 会自动添加到您的密钥名称中。

           II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

        B. 选择保存。

   g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

   h. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。

> **Note**
>
> 如果表名中包含特殊字符（非字母数字），则必须在表名周围使用方括号。例如，`# [my-database-table] ###*`

- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。
- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Microsoft SQL Server

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 `sqlserver`。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。

> **Note**
>
> 如果表名中包含特殊字符（非字母数字），则必须在表名周围使用方括号。例如，*[my-database-table] ###**

- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。Microsoft SQL Server密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用Microsoft SQL Server连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Microsoft SQL Server S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。

- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Microsoft SQL Server 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅Microsoft SQL Server 模板架构。

注意

- Amazon Kendra 检查已更新的内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。
- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 作为最佳实践，请提供只读 Amazon Kendra 的数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

Microsoft Teams

Microsoft Teams 是一款用于消息、会议和文件共享的企业协作工具。如果你是微软 Teams 用户，你可以使用 Amazon Kendra 索引你的 Microsoft Teams 数据源。

你可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到你的 Microsoft Teams 数据源。

要对 Amazon Kendra Microsoft Teams 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引你的 Microsoft Teams 数据源之前，请在你的微软 Teams 和 AWS 账户中进行这些更改。

在 Microsoft Teams 中，请确保：

- 在 Office 365 中创建了一个 Microsoft Teams 账户。
- 记下了您的 Microsoft 365 租户 ID。在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中可以找到您的租户 ID。
- 在 Azure 门户中配置了 OAuth 应用程序，并记下了客户端 ID 和客户端密钥或客户端凭据。有关更多信息，请参阅 Microsoft 教程和注册应用程序示例。

> **Note**
>
> 在 Azure 门户中创建或注册应用程序时，密钥 ID 代表实际的密钥值。在创建密钥和应用程序时，您必须立即记下或保存实际的密钥值。您可以通过在 Azure 门户中选择应用程序的名称，然后导航到证书和密钥上的菜单选项来访问您的密钥。
> 你可以通过在 Azure 门户中选择应用程序的名称，然后导航到概述页面来访问你的客户端 ID。应用程序（客户端）ID 是客户端 ID。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 添加必要的权限 您可以选择添加所有权限，也可以根据要抓取的实体选择更少的权限来限制范围。下表按相应实体列出了应用程序级别的权限：

实体	数据同步所需的权限	身份同步所需的权限
频道帖子	• ChannelMessage.Read.All • Group.Read.All • User.Read • User.Read.All	TeamMember.Read.All
频道附件	• ChannelMessage.Read.All • Group.Read.All • User.Read • User.Read.All	TeamMember.Read.All
频道 Wiki	• Group.Read.All • User.Read • User.Read.All	TeamMember.Read.All
聊天消息	• Chat.Read.All • ChatMessage.Read.All • ChatMember.Read.All • User.Read • User.Read.All • Group.Read.All	TeamMember.Read.All
会议聊天	• Chat.Read.All • ChatMessage。Read	TeamMember.Read.All

实体	数据同步所需的权限	身份同步所需的权限
	• ChatMember.Read.All • User.Read • User.Read.All • Group.Read.All	
聊天附件	• Chat.Read.All • ChatMessage。Read • ChatMember.Read.All • User.Read • User.Read.All • Group.Read.All	TeamMember.Read.All
会议文件	• Chat.Read.All • ChatMessage.Read.All • ChatMember.Read.All • User.Read • User.Read.All • Group.Read.All • Files.Read.All	TeamMember.Read.All
日历会议	• Chat.Read.All • ChatMessage.Read.All • ChatMember.Read.All • User.Read • User.Read.All • Group.Read.All • Files.Read.All	TeamMember.Read.All

实体	数据同步所需的权限	身份同步所需的权限
会议备注	• User.Read • User.Read.All • Group.Read.All • Files.Read.All	TeamMember.Read.All

- 在 Microsoft Teams 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Microsoft Teams 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果你没有现有的 IAM 角色或密钥，则可以在将 Microsoft Teams 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到你的 Microsoft Teams 数据源，你必须提供你的 Microsoft Teams 数据源的必要细节，这样 Amazon Kendra 才能访问你的数据。如果你尚未为其配置 Microsoft Team Amazon Kendra s，请参阅先决条件。

Console

连接微软 Amazon Kendra Teams

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Microsoft Teams 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Microsoft Teams 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 租户 ID-输入你的微软 365 租户 ID。在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中可以找到您的租户 ID。
   b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

c. AWS Secrets Manager 密钥 — 选择现有密钥或创建新 Secrets Manager 密钥来存储你的 Microsoft Teams 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

   i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

      A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Microsoft Teams-” 会自动添加到你的密钥名称中。

      B. 对于客户端 ID 和客户机密钥-在 Azure 门户中输入在 Microsoft Teams 中配置的身份验证凭据。

   ii. 保存并添加您的密钥。

d. 付款模式 - 您可以为您的 Microsoft Teams 账户选择许可和付款模式。A 型支付模式仅限于需要安全合规的许可和支付模式。B 型支付模式适用于不需要安全合规的许可和支付模式。

e. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

f. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

g. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

h. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 同步内容-选择要抓取的内容类型。您可以选择抓取聊天、团队和日历内容。

   b. 其他配置 —指定特定的日历开始和结束日期、用户电子邮件、团队名称和频道名称、附件和 OneNotes。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改后的全新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率-选择同步数据源内容和更新索引的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接微软 Amazon Kendra Teams

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构MSTEAMS时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 租户 ID - 在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中可以找到您的租户 ID。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。

  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含你的 Microsoft Teams 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "clientId": "client ID",
    "clientSecret": "client secret"
}
```

- IAM 角色 —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Microsoft Teams 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Microsoft Teams 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 文档/内容类型-指定是否抓取聊天消息和附件、频道帖子和附件、频道 Wiki、日历内容、会议聊天以及文件和笔记。
- 日历内容-指定开始和结束日期时间以搜寻日历内容。
- 包含和排除筛选条件 - 指定是包含还是排除 Microsoft Teams 中的某些内容。您可以包含或排除团队名称、频道名称、文件名和文件类型、用户电子邮件、 OneNote 分区和 OneNote 页面。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档

使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

- 字段映射 - 选择将 Microsoft Teams 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅 Microsoft Teams 模板架构。

了解更多信息

要详细了解如何 Amazon Kendra 与你的 Microsoft Teams 数据源集成，请参阅：

- 使用微软团队的 Amazon Kendra 连接器智能搜索组织的 Microsoft Teams 数据源

Micoft Yammer

Microsoft Yammer 是一款用于消息、会议和文件共享的企业协作工具。如果你是微软 Yammer 用户，你可以使用 Amazon Kendra 索引你的微软 Yammer 数据源。

你可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到你的 Microsoft Yammer 数据源。

要对 Amazon Kendra Microsoft Yammer 数据源连接器进行故障排除，请参阅数据来源故障排除。

支持的特征

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引微软 Yammer 数据源之前，请先在微软 Yammer 和 AWS 账户中进行这些更改。

在 Microsoft Yammer 中，请确保：

- 在 Office 365 中创建了微软 Yammer 管理账户。
- 记下您的 Microsoft Yammer 用户名和密码。
- 记下了您的 Microsoft 365 租户 ID。在 Azure Active Directory 门户的“属性”或 OAuth 应用程序中可以找到您的租户 ID。
- 在 Azure 门户中配置了 OAuth 应用程序，并记下了客户端 ID 和客户端密钥或客户端凭据。有关更多信息，请参阅 Microsoft 教程和注册应用程序示例。

> **Note**
>
> 在 Azure 门户中创建或注册应用程序时，密钥 ID 代表实际的密钥值。在创建密钥和应用程序时，您必须立即记下或保存实际的密钥值。你可以通过在 Azure 门户中选择应用程序的名称，然后导航到证书和密钥上的菜单选项来访问你的密钥。
> 你可以通过在 Azure 门户中选择应用程序的名称，然后导航到概述页面来访问你的客户端 ID。应用程序（客户端）ID 是客户端 ID。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 在 Microsoft Yammer 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Microsoft Yammer 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果你没有现有的 IAM 角色或密钥，则可以在将 Microsoft Yammer 数据源连接到 Amazon Kendra时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到你的 Microsoft Yammer 数据源，你必须提供微软 Yammer 数据源的必要细节，这样 Amazon Kendra 才能访问你的数据。如果你尚未为其配置 Microsoft Yammer Amazon Kendra，请参阅先决条件。

Console

连接 Amazon Kendra 微软 Yammer

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。

4. 在添加数据源页面上，选择 Microsoft Yammer 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Microsoft Yammer 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

   b. AWS Secrets Manager 密钥 — 选择现有密钥或创建新 Secrets Manager 密钥来存储你的 Microsoft Yammer 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

      i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

         A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Microsoft Yammer-” 会自动添加到你的密钥名称中。

         B. 对于用户名、密码-输入你的 Microsoft Yammer 用户名和密码。

         C. 对于客户端 ID，客户端密钥-在 Azure 门户中输入在 Microsoft Yammer 中配置的身份验证凭据。

      ii. 保存并添加您的密钥。

   c. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

   d. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以

公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

e. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

f. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 起始日期-指定开始在 Microsoft Yammer 中抓取数据的日期。

b. 同步内容-选择要抓取的内容类型。例如，公共消息、私人消息和附件。

c. 其他配置-指定要抓取的某些社区名称，还可以使用正则表达式模式来包含或排除某些内容。

d. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

e. 在同步运行计划中，对于频率-选择同步数据源内容和更新索引的频率。

f. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接 Amazon Kendra 微软 Yammer

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构YAMMER时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供密钥的亚马逊资源名称 (ARN)，该 Secrets Manager 密钥包含你的 Microsoft Yammer 账户的身份验证凭证。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "username": "user name",
    "password": "password",
    "clientId": "client ID",
    "clientSecret": "client secret"
}
```

- IAM 角色 —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Microsoft Yammer 连接器所需的公共 API 的权限，以及。 Amazon Kendra 有关更多信息，请参阅 Microsoft Yammer 数据源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。

- 文档/内容类型-指定是否抓取社区内容、邮件和附件以及私人消息。
- 包含和排除过滤器-指定是包含还是排除某些内容。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。
- 字段映射-选择将 Microsoft Yammer 数据源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅 Microsoft Yammer 模板架构。

了解更多信息

要详细了解如何 Amazon Kendra 与 Microsoft Yammer 数据源集成，请参阅：

- 宣布推出适用于 Yammer 的连接器 Amazon Kendra

MySQL

MySQL 是一个开源的关系数据库管理系统。如果您是MySQL用户，则可以使用索 Amazon Kendra 引您的MySQL数据源。 Amazon Kendra MySQL数据源连接器支持 MySQL 8.0。21.

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的MySQL数据源。

要对 Amazon Kendra MySQL数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引MySQL数据源之前，请在MySQL和 AWS 帐户中进行这些更改。

在 MySQL 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 最佳做法是提供 Amazon Kendra 只读数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。

- 在 MySQL 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 MySQL 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将MySQL数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的MySQL数据源，您必须提供MySQL凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置，MySQL请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 MySQL

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择MySQL连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的MySQL连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 在源中，输入以下信息：

   b. 主机 - 输入数据库主机名。

   c. 端口 - 输入数据库端口。

   d. 实例 - 输入数据库实例。

   e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

   f. 在身份验证中 - 请输入以下信息：

      - AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的MySQL身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

        A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

           I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-MySQL-” 会自动添加到您的密钥名称中。

           II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

        B. 选择保存。

g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

h. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。
- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 MySQL

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 mySql。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。

  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。MySQL密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用MySQL连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 MySQL S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 MySQL 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

注意

- 在 Amazon Kendra 检查更新内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。
- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 最佳做法是提供 Amazon Kendra 只读数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

Oracle Database

Oracle Database 是一个数据库管理系统。如果您是Oracle Database用户，则可以使用索 Amazon Kendra 引您的Oracle Database数据源。 Amazon Kendra Oracle Database数据源连接器支持 Oracle 数据库 18c、19c 和 21c。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的Oracle Database数据源。

要对 Amazon Kendra Oracle Database数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引Oracle Database数据源之前，请在Oracle Database和 AWS 帐户中进行这些更改。

在 Oracle Database 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 最佳做法是提供 Amazon Kendra 只读数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。
- 在 Oracle Database 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Oracle Database 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将Oracle Database数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的Oracle Database数据源，您必须提供Oracle Database凭据的详细信息， Amazon Kendra 以便访问您的数据。如果您尚未进行配置，Oracle Database请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 Oracle Database

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择Oracle Database连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的Oracle Database连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

a. 在源中，输入以下信息：

b. 主机 - 输入数据库主机名。

c. 端口 - 输入数据库端口。

d. 实例 - 输入数据库实例。

e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

f. 在身份验证中 - 请输入以下信息：

- AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的Oracle Database身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

  A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

     I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Oracle Database-” 会自动添加到您的密钥名称中。

     II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

  B. 选择保存。

g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

h. IAM 角色-选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。
- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中选择要映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Oracle Database

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 oracle。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅为新增、修改和删除的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。Oracle Database密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用Oracle Database连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 Oracle Database S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 `VpcConfiguration`，以便调用 `CreateDataSource`。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Oracle Database 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 必须将数据源中的文档正文字段名称映射到索引字段名称`_document_body`。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅Oracle Database 模板架构。

注意

- 在 Amazon Kendra 检查更新内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。

- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 最佳做法是提供 Amazon Kendra 只读数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

PostgreSQL

PostgreSQL 是一个开源数据库管理系统。如果您是PostgreSQL用户，则可以使用索 Amazon Kendra 引您的PostgreSQL数据源。 Amazon Kendra PostgreSQL数据源连接器支持 PostgreSQL 9.6。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的PostgreSQL数据源。

要对 Amazon Kendra PostgreSQL数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

- 字段映射
- 用户上下文筛选
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引PostgreSQL数据源之前，请在PostgreSQL和 AWS 帐户中进行这些更改。

在 PostgreSQL 中，请确保：

- 已记下您的数据库用户名和密码。

> **Important**
>
> 最佳做法是提供 Amazon Kendra 只读数据库凭据。

- 已复制您的数据库主机 URL、端口和实例。
- 在 PostgreSQL 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 PostgreSQL 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将PostgreSQL数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的PostgreSQL数据源，您必须提供PostgreSQL凭据的详细信息，Amazon Kendra 以便访问您的数据。如果您尚未进行配置，PostgreSQL请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 PostgreSQL

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择PostgreSQL连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的PostgreSQL连接器。
5. 在指定数据来源详细信息页面上输入以下信息：
   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。
   b. （可选）说明 - 为数据来源输入说明。
   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。
   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。
   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：
   a. 在源中，输入以下信息：
   b. 主机 - 输入数据库主机名。
   c. 端口 - 输入数据库端口。
   d. 实例 - 输入数据库实例。
   e. 启用 SSL 证书位置-选择输入 SSL 证书文件的 Amazon S3 路径。

f. 在身份验证中 - 请输入以下信息：

- AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的PostgreSQL身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

  A. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

     I. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-PostgreSQL-” 会自动添加到您的密钥名称中。

     II. 对于数据库用户名和密码 - 输入您从数据库中复制的身份验证凭证值。

  B. 选择保存。

g. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

h. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 在同步范围中，从以下选项中进行选择：

- SQL 查询 - 输入 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。 Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 主键列 - 提供数据库表的主键。这将标识数据库中的表。
- 标题列 - 提供数据库表中文档标题列的名称。
- 正文列-提供数据库表中文档正文列的名称。

b. 在其他配置 - 可选中，从以下选项中选择以同步特定内容，而不是同步所有文件：

- 变更检测列-输入 Amazon Kendra 将用于检测内容变化的列的名称。 Amazon Kendra 当其中任何一列发生变化时，将重新索引内容。
- 用户 ID 列 - 输入包含允许访问内容的用户 ID 的列的名称。
- 组列 - 输入包含允许访问内容的群组的列的名称。

- 源 URL 列 - 输入包含要编制索引的源 URL 的列的名称。
- 时间戳列-输入包含时间戳的列的名称。 Amazon Kendra 使用时间戳信息来检测内容的变化并仅同步已更改的内容。
- 时区列 - 输入列的名称，该列包含要搜索的内容的时区。
- 时间戳格式 - 输入列的名称，该列包含用于检测内容更改和重新同步内容的时间戳格式。

c. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

d. 在同步运行计划中，对于频率 - Amazon Kendra 与数据来源同步的频率。

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段（文档 ID、文档标题和来源 URL）中进行选择，以映射到 Amazon Kendra 索引。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 PostgreSQL

您必须使用 TemplateConfigurationAPI 指定以下内容：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构JDBC时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 数据库类型 - 必须将数据库类型指定为 postgresql。
- SQL 查询-指定 SQL 查询语句，例如 SELECT 和 JOIN 操作。SQL 查询必须小于 32KB。Amazon Kendra 将爬取与您的查询相匹配的所有数据库内容。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。PostgreSQL密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "user name": "database user name",
    "password": "password"
}
```

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用PostgreSQL连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 PostgreSQL S3 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。

- 包含和排除过滤器 - 您可以使用用户 ID、组、来源 URL、时间戳和时区来指定是否包含特定内容。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 PostgreSQL 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅PostgreSQL 模板架构。

注意

- 在 Amazon Kendra 检查更新内容时，不会跟踪已删除的数据库行。
- 在数据库的一行中，字段名和值的大小不能超过 400KB。
- 如果您的数据库数据源中有大量数据，并且不 Amazon Kendra 想在第一次同步后将所有数据库内容编入索引，则可以选择仅同步新的、修改过的或已删除的文档。
- 最佳做法是提供 Amazon Kendra 只读数据库凭据。
- 最佳做法是避免添加包含敏感数据或个人身份信息（PII）的表。

Quip

Quip 是一款协作生产力软件，可提供实时文档创作功能。您可以使用 Amazon Kendra 索引 Quip 文件夹、文件、文件评论、聊天室和附件。

您可以使用Amazon Kendra 控制台和 QuipConfigurationAPI 连接 Amazon Kendra 到 Quip 数据源。

要对 Amazon Kendra Quip 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

Amazon Kendra Quip 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Quip 数据源之前，请在您的 Quip 和 AWS 帐户中进行这些更改。

在 Quip 中，请确保：

- 拥有一个具有管理权限的 Quip 账户
- 已创建包含个人访问令牌的 Quip 身份验证凭证。该令牌用作存储在 AWS Secrets Manager 密钥中的身份验证凭证。有关更多信息，请参阅有关身份验证的 Quip 文档。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 已复制 Quip 站点域名。例如，*https://quip-company.quipdomain.com/browse*，其中 *quipdomain* 是域名。
- 在 Quip 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Quip 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Quip 数据源连接到 Amazon Kendra时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要连接 Amazon Kendra 到 Quip 数据源，您必须提供 Quip 数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为 Quip 配置 Amazon Kendra，请参阅先决条件。

Console

连接 Amazon Kendra 到 Quip

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Quip 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Quip 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. Quip 域名 - 输入您从 Quip 账户中复制的域名。

   b. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Quip 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

      i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

         A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Quip-” 会自动添加到您的密钥名称中。

         B. Quip 令牌-输入配置的 Quip 个人访问权限 Quip。

      ii. 添加并保存您的密钥。

   c. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

   d. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

      > **Note**
      >
      > IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

   e. 选择下一步。
7. 在配置同步设置页面上，请输入以下信息：

a. 添加要爬取的 Quip 文件夹 ID - 要爬取的 Quip 文件夹 ID。

> **Note**
>
> 要对根文件夹（包括其中的所有子文件夹和文档）进行爬网，请添加根文件夹 ID。要搜索特定的子文件夹，请添加特定的子文件夹 ID。

b. 其他配置（内容类型） - 输入要爬取的内容类型。

c. 正则表达式模式 - 包含或排除某些文件的正则表达式模式。最多可以添加 100 个模式。

d. 在 “同步” 运行计划中，“频率”-选择同步数据源内容和更新索引的频率

e. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 从生成的默认数据源字段中选择要映射到 Amazon Kendra 索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接 Amazon Kendra 到 Quip

您必须使用 QuipConfiguration API 指定以下内容：

- Quip 网站域名 - 例如，*https://quip-company.quipdomain.com/browse*，其中 *quipdomain* 是域名。
- 亚马逊秘密资源名称 (ARN)-提供包含您的 Quip 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "accessToken": "token"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Quip 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Quip 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 在数据来源配置中指定 VpcConfiguration。请参阅配置 Amazon Kendra 以使用 VPC。
- 包含和排除筛选条件 - 指定是包含还是排除文件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 文件夹-指定要索引的 Quip 文件夹和子文件夹

> **Note**
>
> 要抓取根文件夹，包括其中的所有子文件夹和文档，请输入根文件夹 ID。要搜索特定的子文件夹，请添加特定的子文件夹 ID。

- 附件、聊天室、文件评论-选择是否包括抓取附件、聊天室内容和文件评论。
- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Quip 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

了解更多信息

要了解有关 Amazon Kendra 与 Quip 数据源集成的更多信息，请参阅：

- 使用 Quip 连接器通过智能搜索在 Quip 文档中搜索知识 Amazon Kendra

Salesforce

Salesforce 是一款客户关系管理（CRM）工具，用于管理支持、销售和营销团队。您可以使用 Amazon Kendra 索引您的 Salesforce 标准对象，甚至是自定义对象。

您可以使用Amazon Kendra 控制台、API 或 TemplateConfigurationAPI Amazon Kendra 连接到您的 Salesforce 数据源。SalesforceConfiguration

Amazon Kendra 有两个版本的 Salesforce 连接器。每个版本支持的功能包括：

Salesforce 连接器 V1.0/API SalesforceConfiguration

- 字段映射
- 用户访问控制
- 包含/排除筛选条件

Salesforce 连接器 V2.0/API TemplateConfiguration

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

> **Note**
>
> 对 Salesforce Connector V1.0/ SalesforceConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 Salesforce 连接器 V2.0/ TemplateConfiguration API。

要对您的 Amazon Kendra Salesforce 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- Salesforce 连接器 V1.0
- Salesforce 连接器 V2.0

Salesforce 连接器 V1.0

Salesforce 是一款客户关系管理（CRM）工具，用于管理支持、销售和营销团队。您可以使用 Amazon Kendra 索引您的 Salesforce 标准对象，甚至是自定义对象。

> **Important**
>
> Amazon Kendra 使用 Salesforce API 版本 48。Salesforce API 对每天可以发出的请求数量有限制。如果 Salesforce 超过了这些请求，它会重试，直到能够继续。

> **Note**
>
> 对 Salesforce Connector V1.0/ SalesforceConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 Salesforce 连接器 V2.0/ TemplateConfiguration API。

要对您的 Amazon Kendra Salesforce 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

Amazon Kendra Salesforce 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件

先决条件

在使用 Amazon Kendra 索引您的 Salesforce 数据源之前，请在您的 Salesforce 和 AWS 账户中进行这些更改。

在 Salesforce 中，请确保：

- 已创建一个 Salesforce 账户，并记下了用于连接 Salesforce 的用户名和密码。
- 已创建一个激活 OAuth 的 Salesforce 连接的应用程序账户，并已复制分配给您的 Salesforce 连接的应用程序的使用者键（客户端 ID）和使用者密钥（客户端密钥）。客户端 ID 和客户端密钥用作存储在密 AWS Secrets Manager 钥中的身份验证凭证。有关更多信息，请参阅有关连接的应用程序的 Salesforce 文档。

  > **Note**
  >
  > 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 已复制与用于连接 Salesforce 的账户关联的 Salesforce 安全令牌。
- 要编制索引的 Salesforce 实例的 URL。通常是 *https://<company>.salesforce.com/*。服务器必须运行与 Salesforce 连接的应用程序。
- 通过克隆 ReadOnly 个人资料，然后添加 “查看所有数据” 和 “管理文章” 权限，为拥有 Salesforce 只读访问权限的用户添加了凭据。这些凭据可以识别建立连接的用户以及连接到的 Salesforce Amazon Kendra 连接的应用程序。
- 在 Salesforce 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

  > **Note**
  >
  > 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Salesforce 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Salesforce 数据源连接到 Amazon Kendra时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到您的 Salesforce 数据源，您必须提供您的 Salesforce 数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为 Salesforce 进行配置，请参 Amazon Kendra 阅先决条件。

Console

连接到 Sales Amazon Kendra force

1. 登录到 AWS 管理控制台并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

   > **Note**
   >
   > 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据来源页面上，选择 Salesforce 连接器 V1.0，然后选择添加连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 数据源名称 - 输入您的数据源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 默认语言 - 用于筛选文档以编制索引的语言。除非另行指定，否则语言默认为英语。在元数据中指定的语言会覆盖所选语言。

   d. 添加新标签 - 用于搜索和筛选资源或跟踪分摊费用的标签。

e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. Salesforce URL - 输入要编制索引的 Salesforce 站点的实例 URL。

   b. 对于身份验证类型，请选择现有或新建，以便存储您的 Salesforce 身份验证凭证。如果您选择创建新密钥，则会打开一个 AWS Secrets Manager 秘密窗口。

      - 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

        A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Salesforce-” 会自动添加到您的密钥名称中。

        B. 对于用户名、密码、安全令牌、使用者键、使用者密钥和身份验证 URL，请输入您在 Salesforce 账户中创建的身份验证凭证值。

        C. 选择保存身份验证。

   c. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

   > **Note**
   >
   > IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

   d. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 对于爬取附件 - 选择爬取所有附加对象、文章和源。

   b. 对于标准对象、知识文章和聊天源，请选择要爬取的 Salesforce 实体或内容类型。

   > **Note**
   >
   > 您必须提供配置信息，以便为至少一个标准对象、知识文章或聊天源编制索引。如果您选择爬取知识文章，则必须指定要编制索引的知识文章的类型、文章的名称，以及是为所有知识文章的标准字段编制索引，还是仅为自定义文章类型的字段编制索引。如果您选择为自定义文章编制索引，则必须指定文章类型的内部名称。最多可以指定 10 种文章类型。

   c. 频率-与您的数据源同步的频率。 Amazon Kendra

   d. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

   a. 对于标准知识文章、标准对象附件和其他建议的字段映射-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

   > **Note**
   >
   > 映射到 _document_body 的索引是必需的。您无法更改 Salesforce ID 字段与 Amazon Kendra _document_id 字段之间的映射。

   b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

   c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接到 Sales Amazon Kendra force

您必须指定以下 SalesforceConfigurationAPI：

- 服务器 URL - 要编制索引的 Salesforce 站点的实例 URL。
- 亚马逊秘密资源名称 (ARN)-提供包含您的 Salesforce 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "authenticationUrl": "OAUTH endpoint that Amazon Kendra connects to get an
 OAUTH token",
    "consumerKey": "Application public key generated when you created your
 Salesforce application",
    "consumerSecret": "Application private key generated when you created your
 Salesforce application.",
    "password": "Password associated with the user logging in to the Salesforce
 instance",
    "securityToken": "Token associated with the user account logging in to the
 Salesforce instance",
    "username": "User name of the user logging in to the Salesforce instance"
}
```

- IAM 角色-指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Salesforce 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Salesforce 数据来源的IAM 角色。
- 您必须提供配置信息，以便为至少一个标准对象、知识文章或聊天源编制索引。
  - 标准对象 - 如果选择爬取标准对象，则必须指定标准对象的名称以及包含文档内容的标准对象表中字段的名称。
  - 知识文章 - 如果您选择爬取知识文章，则必须指定要编制索引的知识文章的类型、要编制索引的知识文章的状态，以及是为所有知识文章的标准字段编制索引，还是仅为自定义文章类型的字段编制索引。
  - Chatter 提要-如果您选择抓取 Chatter 提要，则必须在 Salesforce FeedItem 表中指定包含要索引的内容的列的名称。

您还可以添加以下可选功能：

- 包含和排除筛选条件 - 指定是包含还是排除某些文件附件。

  > **Note**
  >
  > 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 字段映射 - 选择将 Salesforce 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

  > **Note**
  >
  > 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称`_document_body`。其他所有字段均为可选字段。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

Salesforce 连接器 V2.0

Salesforce 是一款客户关系管理（CRM）工具，用于管理支持、销售和营销团队。您可以使用 Amazon Kendra 索引您的 Salesforce 标准对象，甚至是自定义对象。

Amazon Kendra Salesforce 数据源连接器支持以下 Salesforce 版本：开发者版和企业版。

> **Note**
>
> 对 Salesforce Connector V1.0/ SalesforceConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 Salesforce 连接器 V2.0/ TemplateConfiguration API。

要对您的 Amazon Kendra Salesforce 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

Amazon Kendra Salesforce 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引您的 Salesforce 数据源之前，请在您的 Salesforce 和 AWS 账户中进行这些更改。

在 Salesforce 中，请确保：

- 已创建一个 Salesforce 管理账户，并记下了用于连接 Salesforce 的用户名和密码。
- 已复制与用于连接 Salesforce 的账户关联的 Salesforce 安全令牌。
- 已创建一个激活 OAuth 的 Salesforce 连接的应用程序账户，并已复制分配给您的 Salesforce 连接的应用程序的使用者键（客户端 ID）和使用者密钥（客户端密钥）。客户端 ID 和客户端密钥用作存储在密 AWS Secrets Manager 钥中的身份验证凭证。有关更多信息，请参阅有关连接的应用程序的 Salesforce 文档。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 要编制索引的 Salesforce 实例的 URL。通常是 *https://<company>.salesforce.com/*。服务器必须运行与 Salesforce 连接的应用程序。
- 通过克隆 ReadOnly 个人资料，然后添加 “查看所有数据” 和 “管理文章” 权限，为拥有 Salesforce 只读访问权限的用户添加了凭据。这些凭据可以识别建立连接的用户以及连接到的 Salesforce Amazon Kendra 连接的应用程序。
- 在 Salesforce 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Salesforce 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Salesforce 数据源连接到 Amazon Kendra时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

Amazon Kendra 要连接到您的 Salesforce 数据源，您必须提供您的 Salesforce 数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为 Salesforce 进行配置，请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 Salesforce，请：

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

   > **Note**
   >
   > 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Salesforce 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Salesforce 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

a. Salesforce URL - 输入要编制索引的 Salesforce 站点的实例 URL。

b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

c. 输入现有密钥，或者如果您创建了新密钥，则会打开 AWS Secrets Manager 密钥窗口。

- 身份验证-在 “创建 AWS Secrets Manager 密钥” 窗口中输入以下信息：

A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Salesforce-” 会自动添加到您的密钥名称中。

B. 对于用户名、密码、安全令牌、使用者键、使用者密钥和身份验证 URL - 请输入您在 Salesforce 账户中生成和下载的身份验证凭证值。

> **Note**
>
> 如果您使用 Salesforce 开发者版，请使用**`https://login.salesforce.com/services/oauth2/token`**或 “我的域名” 登录网址（例如 ***`https://MyCompany.my.salesforce.com`***）作为身份验证网址。如果你使用 Salesforce 沙盒版，请使用**`https://test.salesforce.com/services/oauth2/token`** 或 “我的域名” 登录网址（例如， ***`MyDomainName-SandboxName .sandbox.my.s`***alesforce.com）作为身份验证网址。

C. 选择保存身份验证。

d. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

e. l Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

f. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

g. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 对于爬取附件 - 选择爬取所有附加 Salesforce 对象。

   b. 对于标准对象、带附件的标准对象以及不带附件的标准对象和知识文章 - 请选择要爬取的 Salesforce 实体或内容类型。

   c. 您必须提供配置信息，以便为至少一个标准对象、知识文章或聊天源编制索引。如果您选择爬取知识文章，则必须指定要编入索引的知识文章的类型。您可以选择已发布、已存档、草稿和附件。

   正则表达式筛选条件 - 指定包含特定目录项的正则表达式模式。

8. 对于附加配置：

   - ACL 信息 - 默认情况下包括所有访问控制列表。取消选择访问控制列表将公开该类别中的所有文件。
   - 正则表达式模式 - 添加包含或排除某些文件的正则表达式模式。最多可以添加 100 个模式。

   同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时， Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

   - 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
   - 修改过的新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
   - 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

9. 选择下一步。

10. 在设置字段映射页面上，请输入以下信息：

a. 对于标准知识文章、标准对象附件和其他建议的字段映射-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

> **Note**
>
> 映射到 _document_body 的索引是必需的。您无法更改 Salesforce ID 字段与 Amazon Kendra _document_id 字段之间的映射。您可以将任何 Salesforce 字段映射到文档标题或文档正文 Amazon Kendra 保留/默认索引字段。
> 如果您将任何 Salesforce 字段映射到 Amazon Kendra 文档标题和文档正文字段，Amazon Kendra 将在搜索响应中使用文档标题和正文字段中的数据。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

11. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接到 Sales Amazon Kendra force

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构SALESFORCEV2时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 主机 URL - 指定 Salesforce 实例的主机 URL。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您的 Salesforce 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "authenticationUrl": "OAUTH endpoint that Amazon Kendra connects to get an
 OAUTH token",
    "consumerKey": "Application public key generated when you created your
 Salesforce application",
    "consumerSecret": "Application private key generated when you created your
 Salesforce application",
    "password": "Password associated with the user logging in to the Salesforce
 instance",
    "securityToken": "Token associated with the user account logging in to the
 Salesforce instance",
    "username": "User name of the user logging in to the Salesforce instance"
}
```

- IAM 角色-指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Salesforce 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Salesforce 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除筛选条件 - 您可以指定是包含还是排除某些文档、账户、活动、案例、联系人、潜在客户、机会、解决方案、任务、群组、聊天者和自定义实体文件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

- 字段映射 - 选择将 Salesforce 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

> **Note**
>
> 映射到 _document_body 的索引是必需的。您无法更改 Salesforce ID 字段与 Amazon Kendra _document_id 字段之间的映射。您可以将任何 Salesforce 字段映射到文档标题或文档正文 Amazon Kendra 保留/默认索引字段。
> 如果您将任何 Salesforce 字段映射到 Amazon Kendra 文档标题和文档正文字段，Amazon Kendra 将在搜索响应中使用文档标题和正文字段中的数据。

有关要配置的其他重要 JSON 密钥的列表，请参阅 S alesforce 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与您的 Salesforce 数据源集成的更多信息，请参阅：

- 宣布更新了 Salesforce 连接器 (V2) Amazon Kendra

ServiceNow

ServiceNow 提供基于云的服务管理系统，用于创建和管理组织级工作流程，例如 IT 服务、票务系统和支持。您可以使用索 Amazon Kendra 引您的 ServiceNow 目录、知识文章、事件及其附件。

您可以使用Amazon Kendra 控制台、API 或 TemplateConfigurationAPI 连接到 Amazon Kendra 您的 ServiceNow 数据源。ServiceNowConfiguration

Amazon Kendra 有两个版本的 ServiceNow 连接器。每个版本支持的功能包括：

ServiceNow 连接器 V1.0/API ServiceNowConfiguration

- 字段映射
- ServiceNow 实例版本：伦敦，其他
- 包含/排除筛选条件

ServiceNow 连接器 V2.0/API TemplateConfiguration

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- ServiceNow 实例版本：罗马、圣地亚哥、东京、其他
- Virtual Private Cloud (VPC)

> **Note**
>
> 对 ServiceNow 连接器 V1.0/ ServiceNowConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 ServiceNow 连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra ServiceNow 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

ServiceNow 连接器 V1.0

ServiceNow 提供基于云的服务管理系统，用于创建和管理组织级工作流程，例如 IT 服务、票务系统和支持。您可以使用索 Amazon Kendra 引您的 ServiceNow 目录、知识文章及其附件。

> **Note**
>
> 对 ServiceNow 连接器 V1.0/ ServiceNowConfiguration API 的支持计划于 2023 年结束。我们建议迁移到或使用 ServiceNow 连接器 V2.0/ TemplateConfiguration API。

要对 Amazon Kendra ServiceNow 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

Amazon Kendra ServiceNow 数据源连接器支持以下功能：

- ServiceNow 实例版本：伦敦，其他
- 包含/排除模式：服务目录、知识文章及其附件

先决条件

在使用索 Amazon Kendra 引 ServiceNow 数据源之前，请在 ServiceNow 和 AWS 帐户中进行这些更改。

在中 ServiceNow，请确保你有：

- 已创建 ServiceNow 管理员账户并创建 ServiceNow实例。
- 已复制您的 ServiceNow 实例 URL 的主机。例如，如果实例的 URL 是 *https://your-domain.service-now.com*，则您输入的主机 URL 的格式为 *your-domain.service-now.com*。
- 已记下您的基本身份验证凭证，其中包含允许 Amazon Kendra 连接到您的 ServiceNow 实例的用户名和密码。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 可选：配置了 OAuth 2.0 凭证令牌，该令牌可以识别 Amazon Kendra 和生成用户名、密码、客户端 ID 和客户端密钥。用户名和密码必须提供对 ServiceNow 知识库和服务目录的访问权限。有关更多信息，请参阅有关 OAuth 2.0 身份验证的ServiceNow 文档。
- 添加了以下权限：
  - kb_category
  - kb_knowledge
  - kb_knowledge_base
  - kb_uc_cannot_read_mtom
  - kb_uc_can_read_mtom
  - sc_catalog
  - sc_category
  - sc_cat_item
  - sys_attachment
  - sys_attachment_doc
  - sys_user_role
- 已选中每个文档在您计划用于同一索引的其他数据源中 ServiceNow 以及其他数据源中都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 ServiceNow 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下该密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 ServiceNow 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 ServiceNow 数据源，您必须提供 ServiceNow 数据源的必要详细信息，以便 Amazon Kendra 能够访问您的数据。如果您尚未进行配置， ServiceNow 请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 ServiceNow

1. 登录 AWS 管理控制台并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择ServiceNow连接器 V1.0，然后选择添加数据源。

5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. ServiceNow 主机 —输入 ServiceNow主机 URL。

   b. ServiceNow 版本-选择您的 ServiceNow 版本。

   c. 根据您的使用案例，选择基本身份验证或 Oauth 2.0 身份验证。

   d. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 ServiceNow 身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

      i. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-ServiceNow-” 会自动添加到您的密钥名称中。

      ii. 如果使用基本身份验证，请输入您帐户的密码名称、用户名和密码。 ServiceNow

      如果使用 OAuth2 身份验证，请输入您在账户中创建的密钥名称、用户名、密码、客户端 ID 和客户端密钥。 ServiceNow

      iii. 选择保存和添加密钥。

   e. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

      > **Note**
      >
      > IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

   f. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

   a. 包含知识文章 - 选择为知识文章编制索引。

b. 知识文章的类型-根据您的用例，在 “仅包含公开文章” 和 “基于 ServiceNow 筛选查询包含文章” 之间进行选择。如果您选择 “包含基于 ServiceNow 筛选查询的文章”，则必须输入从您的 ServiceNow账户中复制的筛选查询。

c. 包含知识文章附件 - 选择为知识文章附件编制索引。您也可以选择要编制索引的特定文件类型。

d. 包含目录项 - 选择为目录项编制索引。

e. 包含目录项附件 - 选择为目录项附件编制索引。您也可以选择要编制索引的特定文件类型。

f. 频率-与您的数据源同步的频率。 Amazon Kendra

g. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 知识文章和服务目录-从 Amazon Kendra 生成的默认数据源字段以及要映射到索引的其他建议字段映射中进行选择。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 ServiceNow

您必须使用 ServiceNowConfiguration API 指定以下内容：

- 数据源 URL-指定 ServiceNow URL。主机端点应如下所示：*your-domain.service-now.com*。
- 数据源主机实例-将 ServiceNow 主机实例版本指定为LONDON或OTHERS。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。 ServiceNow

如果您使用基本身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "username": "user name",
    "password": "password"
```

```
}
```

如果您使用 OAuth2 身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "username": "user name",
    "password": "password",
    "clientId": "client id",
    "clientSecret": "client secret"
}
```

- IAM ro le —指定`RoleArn`何时调用`CreateDataSource`以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 ServiceNow 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 ServiceNow 数据源的IAM 角色。

您还可以添加以下可选功能：

- 字段映射-选择将 ServiceNow 数据源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称`_document_body`。其他所有字段均为可选字段。

- 包含和排除筛选条件 - 指定是包含还是排除某些类别和知识文章的附件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 索引参数 - 您也可以选择指定是否：
  - 为知识文章或服务目录编制索引，或者为两者同时编制索引。如果您选择为知识文章和服务目录项编制索引，则必须提供映射到索引中索引文档内容字段的 Amazon Kendra 字段名称。ServiceNow

- 为知识文章和目录项的附件编制索引。
- 使用从一个或多个知识库中选择文档的 ServiceNow 查询。知识库可以是公有的，也可以是私有的。有关更多信息，请参阅使用查询指定要编制索引的文档。

了解更多信息

要了解有关 Amazon Kendra 与 ServiceNow 数据源集成的更多信息，请参阅：

- Amazon Kendra ServiceNow 在线连接器入门

ServiceNow 连接器 V2.0

ServiceNow 提供基于云的服务管理系统，用于创建和管理组织级工作流程，例如 IT 服务、票务系统和支持。您可以使用索 Amazon Kendra 引您的 ServiceNow 目录、知识文章、事件及其附件。

要对 Amazon Kendra ServiceNow 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

Amazon Kendra ServiceNow 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- ServiceNow 实例版本：罗马、圣地亚哥、东京、其他
- Virtual Private Cloud (VPC)

先决条件

在使用索 Amazon Kendra 引 ServiceNow 数据源之前，请在 ServiceNow 和 AWS 帐户中进行这些更改。

在中 ServiceNow，请确保你有：

- 已创建个人或企业开发者实例，并拥有一个具有管理角色的 ServiceNow实例。
- 已复制您的 ServiceNow 实例 URL 的主机。您输入的主机 URL 的格式是 *your-domain.service-now.com*。您需要您的 ServiceNow 实例 URL 才能连接 Amazon Kendra。
- 记下了您的基本身份验证凭证，包括用户名和密码， Amazon Kendra 以允许您连接到您的 ServiceNow 实例。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 可选：已配置的 OAuth 2.0 客户端凭据，这些凭据可以使用用户名、密码和生成的客户端 ID 以及客户端密钥进行识别 Amazon Kendra 。有关更多信息，请参阅有关 OAuth 2.0 身份验证的 ServiceNow 文档。
- 已选中每个文档在您计划用于同一索引的其他数据源中 ServiceNow 以及其他数据源中都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 ServiceNow 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下该密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 ServiceNow 数据源连接到时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥 Amazon Kendra。如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 ServiceNow 数据源，您必须提供 ServiceNow 数据源的必要详细信息，以便 Amazon Kendra 能够访问您的数据。如果您尚未进行配置， ServiceNow 请参 Amazon Kendra 阅先决条件。

Console

要连接 Amazon Kendra 到 ServiceNow

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

   > **Note**
   >
   > 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择ServiceNow 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的ServiceNow 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

a. ServiceNow 主机 —输入 ServiceNow主机 URL。您输入的主机 URL 的格式是 *your-domain.service-now.com*。

b. ServiceNow 版本-选择您的 ServiceNow 实例版本。您可以选择 “罗马”、“圣地亚哥”、“东京” 或 “其他”。

c. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

d. 身份验证-在基本身份验证和 Oauth 2.0 身份验证之间进行选择。

e. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 ServiceNow 身份验证凭证。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。在窗口中输入以下信息：

i. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-ServiceNow-” 会自动添加到您的密钥名称中。

ii. 如果使用基本身份验证，请输入您帐户的密码名称、用户名和密码。 ServiceNow

如果使用 OAuth2.0 身份验证，请输入您在账户中创建的密钥名称、用户名、密码、客户端 ID 和客户端密钥。 ServiceNow

iii. 保存并添加您的密钥。

f. 虚拟私有云（ VPC ）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

g. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

h. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

i. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 对于知识文章，请选择以下选项：

- 知识文章 - 选择为知识文章编制索引。
- 知识文章附件 - 选择为知识文章附件编制索引。
- 知识文章的类型-根据您的用例，根据 ServiceNow 筛选查询，在 “仅限公开文章” 和 “知识文章” 之间进行选择。如果您选择 “包含基于 ServiceNow 筛选查询的文章”，则必须输入从您的 ServiceNow账户中复制的筛选查询。筛选查询示例包括：*workflow_state=draft^EQ*、*kb_knowledge_base=dfc19531bf2021003f07e2c*
ISNOTEMPTY^EQ、*article_type=text^active=true^EQ*。

> **Important**
>
> 如果您选择仅抓取公开文章，则仅 Amazon Kendra 抓取中分配了公共访问角色的知识文章。 ServiceNow

- 基于简短描述筛选条件包含文章 - 指定正则表达式模式以包含或排除特定文章。

b. 对于服务目录项：

- 服务目录项 - 选择为目录项编制索引。
- 服务目录项附件 - 选择为服务目录项附件编制索引。
- 活动服务目录项 - 选择为活动目录项编制索引。
- 非活动服务目录项 - 选择为非活动目录项编制索引。
- 筛选查询-根据您的 ServiceNow 实例中定义的筛选器选择包含服务目录项目。筛选条件查询示例包
括：*short_descriptionLIKEAccess^category=2809952237b1300054b6a3549dbe5*
- 根据简短描述筛选条件包含服务目录项 - 指定正则表达式模式以包括特定的目录项。

c. 对于事件：

- 事件 - 选择为服务事件编制索引。
- 事件附件 - 选择为事件附件编制索引。
- 活动事件 - 选择为活动事件编制索引。
- 非活动事件 - 选择为非活动事件编制索引。
- 活动事件类型 -根据您的使用案例，选择所有事件、未解决的事件、未解决 - 未分配的事件或已解决的事件。
- 筛选查询-根据您的 ServiceNow 实例中定义的筛选器选择包含事件。筛选条件查询示例包括：*short_descriptionLIKETest^urgency=3^state=1^EQ*、*priority=2^categor*
- 根据简短描述筛选条件包含事件 - 指定包含特定事件的正则表达式模式。

d. 对于其他配置：

- ACL 信息 - 默认情况下，包含所选实体的访问控制列表。取消选择访问控制列表将公开该类别中的所有文件。对于未选择的实体，ACL 选项会自动停用。对于公开文章，ACL 不适用。
- 对于最大文件大小 — 指定 Amazon Kendra 将抓取的文件大小限制（以 MB 为单位）。Amazon Kendra 只会抓取您定义的大小限制范围内的文件。默认文件大小为 50MB。最大文件大小应大于 0MB 且小于或等于 50MB。
- 附件正则表达式模式 - 添加正则表达式模式以包含或排除目录、知识文章和事件的某些附加文件。最多可以添加 100 个模式。

e. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

f. 在 “同步” 运行计划中，“频率”-选择同步数据源内容和更新索引的频率。

g. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认字段映射-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 ServiceNow

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构SERVICENOWV2时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 主机 URL-指定 ServiceNow 主机实例版本。例如，*your-domain.service-now.com*。
- 身份验证类型-指定您使用的身份验证类型，无论是OAuth2针对您的 ServiceNow实例basicAuth还是针对您的实例。
- ServiceNow 实例版本-指定您使用的 ServiceNow 实例，是Tokyo、SandiegoRome、或Others。
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 亚马逊秘密资源名称 (ARN)-提供包含您在账户中创建的身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。 ServiceNow

  如果您使用基本身份验证，密钥将存储在 JSON 结构中，其中包含以下键：

```
{
    "username": "user name",
    "password": "password"
}
```

- 如果您使用 OAuth2 客户端证书，则密钥将存储在 JSON 结构中，其中包含以下密钥：

```
{
```

```
    "username": "user name",
    "password": "password",
    "clientId": "client id",
    "clientSecret": "client secret"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 ServiceNow 连接器所需的公共 API 的权限，以及 Amazon Kendra。有关更多信息，请参阅 ServiceNow 数据源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 包含和排除筛选条件 - 您可以使用知识文章、服务目录和事件的文件名和文件类型来指定是包含还是排除某些附件。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 要索引的特定文档-您可以使用 ServiceNow查询从一个或多个知识库（包括私有知识库）中指定所需的文档。对知识库的访问权限由您用来连接 ServiceNow 实例的用户决定。有关更多信息，请参阅使用查询指定要编制索引的文档。
- 索引参数 - 您也可以选择指定是否：
  - 为知识文章、服务目录或事件编制索引，或者为所有这些内容编制索引。如果您选择为知识文章、服务目录项目和事件编制索引，则必须提供映射到索引中索引文档内容字段的 Amazon Kendra 字段名称。 ServiceNow
  - 为知识文章、服务目录项和事件的附件编制索引。
  - 根据 short description 筛选模式包含知识文章、服务目录项目和事件。
  - 选择筛选活动和非活动服务目录项目和事件。
  - 选择根据事件类型筛选事件。
  - 选择爬取对哪些实体的 ACL。

- 您可以使用 ServiceNow 查询从一个或多个知识库（包括私有知识库）中指定所需的文档。对知识库的访问权限由您用来连接 ServiceNow实例的用户决定。有关更多信息，请参阅使用查询指定要编制索引的文档。
- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra 的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。
- 字段映射-选择将 ServiceNow 数据源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅ServiceNow 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与 ServiceNow 数据源集成的更多信息，请参阅：

- 开始 Amazon Kendra发布更新的 ServiceNow 连接器 (V2) Amazon Kendra

使用查询指定要编制索引的文档

您可以使用 ServiceNow 查询来指定要包含在 Amazon Kendra 索引中的文档。使用查询时，您可以指定多个知识库，包括私有知识库。对知识库的访问权限由您用来连接 ServiceNow 实例的用户决定。

要生成查询，请使用 ServiceNow 查询生成器。您可以使用构建器来创建查询并测试查询是否可以返回正确的文档列表。

使用 ServiceNow 控制台创建查询

1. 登录 ServiceNow 控制台。

2. 从左侧菜单中选择知识，再选择文章，然后选择全部。
3. 在页面顶部选择筛选条件图标。
4. 使用查询构建器创建查询。
5. 查询完成后，右键单击该查询，然后选择复制查询，以便从查询构建器中复制查询。保存此查询以在中使用 Amazon Kendra。

在复制查询时，确保不要更改任何查询参数。如果无法识别任何查询参数，则 ServiceNow 会将该参数视为空且不使用它来筛选结果。

Slack

Slack 是一款企业通信应用程序，允许用户通过各种公共和私人频道发送消息和附件。您可以使用 Amazon Kendra 索引您的 Slack 公共和私人频道、机器人和存档消息、文件和附件、直接消息和群组消息。还可以选择要筛选的特定内容。

> **Note**
>
> Amazon Kendra 现在支持升级后的 Slack 连接器。
> 控制台已自动为您升级。您在控制台中创建的任何新连接器都将使用升级后的架构。如果您使用 API，则现在必须使用TemplateConfiguration对象而不是`SlackConfiguration`对象来配置您的连接器。
> 使用较旧的控制台和 API 架构配置的连接器将继续按配置运行。但是，您将无法对其进行编辑或更新。如果要编辑或更新连接器配置，则必须创建新的连接器。
> 我们建议将您的连接器工作流程迁移到升级版本。对使用旧架构配置的连接器的支持计划于 2024 年 6 月结束。

您可以使用Amazon Kendra 控制台或 TemplateConfigurationAPI Amazon Kendra 连接到您的 Slack 数据源。

要对 Amazon Kendra Slack 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

支持的特征

Amazon Kendra Slack 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Slack 数据源之前，请在您的 Slack 和 AWS 帐户中进行这些更改。

在 Slack 中，请确保：

- 已配置 Slack Bot 用户 OAuth 令牌或 Slack 用户 OAuth 令牌。您可以选择任一令牌 Amazon Kendra 来连接您的 Slack 数据源。需要使用令牌作为您的身份验证凭证。有关更多信息，请参阅有关访问令牌的 Slack 文档。

> **Note**
>
> 如果您使用机器人令牌作为 Slack 凭证的一部分，则无法将私信和群组消息编入索引，必须将机器人令牌添加到要编制索引的频道中。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 在 Slack 工作区主页 URL 中记下您的 Slack 工作区团队 ID。例如，*https://app.slack.com/client/T0123456789/...*，其中 *T0123456789* 是团队 ID。
- 添加了以下 Oauth 范围/权限：

用户令牌范围	机器人令牌范围
• channels:history • channels:read • emoji:read • files:read • groups:history • groups:read • im:history • im:read • mpim:history • mpim:read • team:read • users.profile:read • users:read • 用户:阅读.email	• channels:history • 频道:管理 • channels:read • 对话. 连接:管理 • 对话. connections: 阅读 • files:read • groups:history • groups:read • im:history • im:read • mpim:history • mpim:read • 反应:读取 • team:read • usergroups:read • users.profile:read • users:read • 用户:阅读.email

- 在 Slack 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Slack 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Slack 数据源连接至时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。 Amazon Kendra如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 Slack 数据源，您必须提供 Slack 数据源的必要详细信息， Amazon Kendra 以便访问您的数据。如果您尚未为 Slack 配置 Amazon Kendra，请参阅先决条件。

Console

要连接 Amazon Kendra 到 Slack

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Slack 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Slack 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。
6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. 对于 Slack 工作区团队 ID-您的 Slack 工作空间的团队 ID。你可以在你的 Slack 工作区主页网址中找到你的团队 ID。例如，*https://app.slack.com/client/T0123456789/...*，其中 *T0123456789* 是团队 ID。

   b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

   c. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Slack 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

      i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

         A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Slack-” 会自动添加到您的密钥名称中。

         B. 对于 Slack 令牌-输入您配置的 Slack 的身份验证凭据值。

      ii. 保存并添加您的密钥。

d. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

e. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

f. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

g. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 选择内容类型-选择要抓取的 Slack 实体或内容类型。您可以从所有频道、公共频道、私人频道、群组消息和私人消息中进行选择。

b. 选择抓取开始日期-输入您想要开始抓取内容的日期。

c. 对于其他配置-选择包含机器人和存档消息，并使用正则表达式模式来包含或排除某些内容。

> **Note**
>
> 如果您选择同时包含频道 ID 和频道名称，则 Amazon Kendra Slack 连接器将优先考虑频道 ID 而不是频道名称。
> 如果您选择包含某些私人和群组消息， Amazon Kendra Slack 连接器将忽略所有私人和群组消息，只抓取您指定的私人和群组消息。

d. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。

- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

e. 在 “同步” 运行计划中，“频率”-选择同步数据源内容和更新索引的频率。

f. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

要连接 Amazon Kendra 到 Slack

您必须使用 TemplateConfiguration API 指定数据来源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构SLACK时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- Slack 工作区团队 ID - 您从 Slack 主页 URL 中复制的 Slack 团队 ID。
- 起始日期-开始从 Slack 工作空间团队抓取数据的日期。日期必须遵循此格式: yyyy-mm-dd.
- 同步模式-指定数据源内容发生变化时 Amazon Kendra 应如何更新索引。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。您可以选择：
  - FORCED_FULL_CRAWL对所有内容进行全新索引，每次数据源与索引同步时都要替换现有内容。
  - FULL_CRAWL每次数据源与索引同步时，仅对新的、修改过的和已删除的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
  - CHANGE_LOG每次数据源与索引同步时，仅索引新的和修改过的内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra

的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用 PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

- 亚马逊秘密资源名称 (ARN)-提供包含您的 Slack 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "slackToken": "token"
}
```

- IAM ro le —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Slack 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Slack 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 特定频道-按公共或私人频道筛选，并按其 ID 指定某些频道。
- 频道和消息的类型-是否 Amazon Kendra 应将您的公共和私人频道、群组和私信以及机器人和存档消息编入索引。如果您使用机器人令牌作为 Slack 身份验证凭证的一部分，则必须将机器人令牌添加到要编制索引的频道。您无法使用机器人令牌将私信和群组消息编入索引。
- 回顾-您可以选择配置lookBack参数，以便 Slack 连接器在上次连接器同步之前的指定小时数内抓取已更新或已删除的内容。
- 包含和排除过滤器-指定是包含还是排除某些 Slack 内容。如果您使用机器人令牌作为 Slack 身份验证凭证的一部分，则必须将机器人令牌添加到要编制索引的频道。您无法使用机器人令牌将私信和群组消息编入索引。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 字段映射 - 选择将 Slack 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 键的列表，请参阅 Slack 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与 Slack 数据源集成的更多信息，请参阅：

- 使用 Amazon Kendra Slack 连接器进行智能搜索，揭开 Slack 工作区的面纱

Zendesk

Zendesk 是一个客户关系管理系统，可帮助企业自动化和增强客户支持互动。您可以使用 Amazon Kendra 索引您的 Zendesk 支持工单、工单评论、工单附件、帮助中心文章、文章评论、文章评论附件、指南社区主题、社区帖子和社区帖子评论。

如果要仅对位于特定组织内的工单编制索引，则可以按组织名称进行筛选。您也可以选择为您想要开始从 Zendesk 爬取数据的时间设置一个爬取日期。

您可以使用Amazon Kendra 控制台和 TemplateConfigurationAPI Amazon Kendra 连接到您的 Zendesk 数据源。

要对 Amazon Kendra Zendesk 数据源连接器进行故障排除，请参阅数据来源故障排除。

主题

- 支持的特征
- 先决条件
- 连接说明
- 了解更多信息

支持的特征

Amazon Kendra Zendesk 数据源连接器支持以下功能：

- 字段映射
- 用户访问控制
- 包含/排除筛选条件
- 更改日志、完整内容和增量内容同步
- Virtual Private Cloud (VPC)

先决条件

在使用 Amazon Kendra 索引 Zendesk 数据源之前，请在您的 Zendesk 和 AWS 账户中进行这些更改。

在 Zendesk 中，请确保：

- 已创建 Zendesk 套件（专业版/企业版）管理账户。
- 已记下您的 Zendesk 主机 URL。例如，*https://{sub-domain}.zendesk.com/*。

> **Note**
>
> （本地/服务器）Amazon Kendra 会检查中 AWS Secrets Manager 包含的端点信息是否与数据源配置详细信息中指定的端点信息相同。这有助于防止出现混淆代理人问题，这是一个安全问题，即用户无权执行操作，但可以将 Amazon Kendra 作为代理来访问配置的密钥和执行操作。如果以后更改端点信息，则必须创建一个新密钥来同步此信息。

- 配置了包含客户端 ID、客户端密钥、用户名和密码的 OAuth 2.0 令牌。需要使用 OAuth 2.0 令牌作为您的身份验证凭证。有关更多信息，请参阅有关配置 OAuth 2.0 令牌的 Zendesk 文档。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

- 已添加以下 OAuth 2.0 范围：
  - read

- 可选：已安装允许 Amazon Kendra 进行连接的 SSL 证书。
- 在 Zendesk 以及计划用于编制同一索引的其他数据来源中，已检查每个文档都是唯一的。您要用于编制索引的每个数据来源在所有数据来源中都不能包含相同的文档。文档 ID 对索引来说是全局性的，并且每个索引都必须是唯一的。

在你的 AWS 账户，请确保你有：

- 已创建 Amazon Kendra 索引，如果使用 API，则记下索引 ID。
- 为您的数据源@@ 创建了一个 IAM 角色，如果使用 API，请记下该角色的 ARN。 IAM

> **Note**
>
> 如果您更改了身份验证类型和证书，则必须更新您的 IAM 角色才能访问正确的 AWS Secrets Manager 密钥 ID。

- 将您的 Zendesk 身份验证凭证存储在 AWS Secrets Manager 密钥中，如果使用 API，请记下密钥的 ARN。

> **Note**
>
> 我们建议您定期刷新或轮换您的凭证和密码。为了安全起见，请仅提供必要的访问权限级别。我们建议不要跨数据来源以及连接器版本 1.0 和 2.0（如果适用）重复使用凭证和密钥。

如果您没有现有的 IAM 角色或密钥，则可以在将 Zendesk 数据源连接至时使用控制台创建新的 IAM 角色和 Secrets Manager 密钥。 Amazon Kendra如果您使用的是 API，则必须提供现有 IAM 角色和 Secrets Manager 密钥的 ARN 以及索引 ID。

连接说明

要 Amazon Kendra 连接到您的 Zendesk 数据源，您必须提供 Zendesk 数据源的必要详细信息，Amazon Kendra 以便访问您的数据。如果您尚未为 Zendesk 配置 Amazon Kendra，请参阅先决条件。

Console

连接到 Zen Amazon Kendra desk

1. 登录 AWS Management Console 并打开Amazon Kendra 控制台。
2. 在左侧导航窗格中，选择索引，然后从索引列表中选择要使用的索引。

> **Note**
>
> 您可以选择在索引设置下配置或编辑您的用户访问控制设置。

3. 在入门页面上，选择添加数据来源。
4. 在添加数据源页面上，选择 Zendesk 连接器，然后选择添加连接器。如果使用版本 2（如果适用），请选择带有 “V2.0” 标签的 Zendesk 连接器。
5. 在指定数据来源详细信息页面上输入以下信息：

   a. 在名称和描述中，在数据来源名称中输入您的数据来源的名称。可以包含连字符，但不能包含空格。

   b. （可选）说明 - 为数据来源输入说明。

   c. 使用默认语言-选择一种语言来筛选文档中的索引。除非另行指定，否则语言默认为英语。在文档元数据中指定的语言会覆盖所选语言。

   d. 在标签中，用于添加新标签-包括可选标签以搜索和筛选您的资源或跟踪您的 AWS 成本。

   e. 选择下一步。

6. 在定义访问权限和安全性页面上，请输入以下信息：

   a. Zendesk URL - 输入 Zendesk URL。例如，*https://{sub-domain}.zendesk.com/*。

   b. 授权-如果您有 ACL 并想将其用于访问控制，请打开或关闭文档的访问控制列表 (ACL) 信息。ACL 指定了用户和群组可以访问哪些文档。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。

   c. AWS Secrets Manager s@@ ecret —选择现有密钥或创建新 Secrets Manager 密钥来存储您的 Zendesk 身份验证凭据。如果您选择创建新密钥，则会打开 AWS Secrets Manager 密钥窗口。

      i. 在创建 AWS Secrets Manager 密钥窗口中输入以下信息：

         A. 密钥名称 - 密钥的名称。前缀 “AmazonKendra-Zendesk-” 会自动添加到您的密钥名称中。

         B. 对于 “客户端 ID”、“客户机密”、“用户名”、“密码”-输入在 Zendesk 中配置的身份验证凭据值。

ii. 保存并添加您的密钥。

d. 虚拟私有云（VPC）- 您可以选择使用 VPC。如果是这样，则必须添加子网和 VPC 安全组。

e. I Amazon Kendra dent@@ ity Crawler-指定是否开启身份搜寻器。Identity Crawler 使用文档的访问控制列表 (ACL) 信息，根据用户或其群组对文档的访问权限筛选搜索结果。如果您的文档有 ACL 并选择使用您的 ACL，则也可以选择开启身份爬网程序来配置搜索结果 Amazon Kendra的用户上下文筛选。否则，如果身份搜寻器已关闭，则可以公开搜索所有文档。如果您想对文档使用访问控制并且身份搜寻器已关闭，则可以使用PutPrincipalMappingAPI 上传用户和群组访问信息以进行用户上下文筛选。

f. IAM ro le —选择现有 IAM 角色或创建新 IAM 角色来访问您的存储库凭据和索引内容。

> **Note**
>
> IAM 用于索引的角色不能用于数据源。如果您不确定是否将现有角色用于编制索引或常见问题解答，为了避免出错，请选择创建新角色。

g. 选择下一步。

7. 在配置同步设置页面上，请输入以下信息：

a. 选择内容-选择要从工单中抓取的内容类型，以获取帮助中心文章、社区主题等。

b. 组织名称-输入 Zendesk 组织名称以筛选内容。

c. 同步开始日期-输入您想要开始抓取内容的日期。

d. 正则表达式模式 - 添加包含或排除某些文件的正则表达式模式。最多可以添加 100 个模式。

e. 同步模式 - 选择在数据来源内容发生变化时更新索引的方式。首次与同步数据源时，Amazon Kendra 默认情况下会对所有内容进行抓取和索引。如果初始同步失败，即使您没有选择完全同步作为同步模式选项，也必须对数据进行完全同步。

- 完全同步：对所有内容进行新索引，每次数据源与索引同步时都会替换现有内容。
- 修改后的全新同步：每次数据源与索引同步时，仅为新增和修改过的内容编制索引。Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。
- 新增、已修改、已删除的同步：每次数据源与索引同步时，仅索引新内容、修改内容和已删除内容。 Amazon Kendra 可以使用数据源的机制来跟踪内容更改并索引自上次同步以来更改的内容。

f. 在 F requency 的同步运行计划中-选择同步数据源内容和更新索引的频率。

g. 选择下一步。

8. 在设置字段映射页面上，请输入以下信息：

a. 默认数据源字段-从 Amazon Kendra 生成的默认数据源字段中选择要映射到索引的字段。

b. 添加字段 - 添加自定义数据来源字段以创建要映射到的索引字段名称和字段数据类型。

c. 选择下一步。

9. 在查看和创建页面上，请检查输入的信息是否正确，然后选择添加数据来源。您也可以选择在此页面上编辑信息。成功添加数据来源后，您的数据来源将显示在数据来源页面上。

API

连接到 Zen Amazon Kendra desk

您必须使用 TemplateConfigurationAPI 指定数据源架构的 JSON。您必须提供以下信息：

- 数据源-将数据源类型指定为使用 TemplateConfigurationJSON 架构ZENDESK时的类型。还要像调用 CreateDataSourceAPI TEMPLATE 时一样指定数据源。
- 主机 URL - 在连接配置或存储库端点详细信息中提供您的 Zendesk 主机 URL。例如，*https://yoursubdomain.zendesk.com*。
- 更改日志-是否 Amazon Kendra 应使用 Zendesk 数据源更改日志机制来确定是否必须更新索引中的文档。

> **Note**
>
> 如果您不想让 Amazon Kendra 扫描所有文档，请使用更改日志。如果您的更改日志很大，则扫描 Zendesk 数据源中的文档所花费的时间可能比处理变更日志所需的时间 Amazon Kendra 少。如果您是首次将 Zendesk 数据来源与索引同步，则会扫描所有文档。

- 亚马逊秘密资源名称 (ARN)-提供包含您的 Zendesk 账户身份验证凭证的 Secrets Manager 密钥的亚马逊资源名称 (ARN)。密钥必须使用具有以下键的 JSON 结构存储：

```
{
    "hostUrl": "https://yoursubdomain.zendesk.com",
    "clientId": "client ID",
    "clientSecret": "Zendesk client secret",
```

```
    "userName": "Zendesk user name",
    "password": "Zendesk password"
}
```

- IAM 角色 —指定RoleArn何时调用CreateDataSource以向 IAM 角色提供访问您的 Secrets Manager 密钥和调用 Zendesk 连接器所需的公共 API 的权限，以及。 Amazon Kendra有关更多信息，请参阅 Zendesk 数据来源的IAM 角色。

您还可以添加以下可选功能：

- 虚拟私有云（VPC）- 指定 VpcConfiguration，以便调用 CreateDataSource。有关更多信息，请参阅 配置 Amazon Kendra 为使用 Amazon VPC。
- 文档/内容类型-指定是否进行爬网：
  - 支持工单、工单备注和/或工单备注附件
  - 帮助中心文章、文章附件和文章备注
  - 指南社区主题、帖子或帖子评论
- 包含和排除过滤器-指定是包含还是排除某些 Slack 内容。如果您使用机器人令牌作为 Slack 身份验证凭证的一部分，则必须将机器人令牌添加到要编制索引的频道。您无法使用机器人令牌将私信和群组消息编入索引。

> **Note**
>
> 大多数数据来源使用正则表达式模式，即称为筛选条件的包含或排除模式。如果您指定包含筛选条件，则只会为与包含筛选条件匹配的内容编制索引。不会为任何与包含筛选条件不匹配的文档编制索引。如果您指定包含和排除筛选条件，则不会为与排除筛选条件匹配的文档编制索引，即使它们与包含筛选条件相匹配。

- 用户上下文筛选和访问控制-如果您的文档有 ACL，则会Amazon Kendra 搜索文档的访问控制列表 (ACL)。ACL 信息用于根据用户或用户组对文档的访问权限来筛选搜索结果。有关更多信息，请参阅用户上下文筛选。
- 字段映射 - 选择将 Zendesk 数据来源字段映射到 Amazon Kendra 索引字段。有关更多信息，请参阅映射数据来源字段。

> **Note**
>
> 要搜索您的文档，必须输入文档正文字段或文档正文等效字段。 Amazon Kendra 您必须将数据源中的文档正文字段名称映射到索引字段名称_document_body。其他所有字段均为可选字段。

有关要配置的其他重要 JSON 密钥的列表，请参阅 Zendesk 模板架构。

了解更多信息

要了解有关 Amazon Kendra 与 Zendesk 数据源集成的更多信息，请参阅：

- 通过 Amazon Kendra 智能搜索发现来自 Zendesk 的见解

映射数据来源字段

Amazon Kendra 数据源连接器可以将数据源的文档或内容字段映射到 Amazon Kendra 索引中的字段。默认情况下，每个连接器都设计为爬取特定的数据来源字段。默认数据来源字段及其属性无法更改或自定义。在 Amazon Kendra 控制台上，无法编辑的默认字段和默认字段属性显示为灰色。

Amazon Kendra 连接器还允许您将数据源的自定义文档或内容字段映射到索引中的自定义字段。例如，如果您的数据来源中有一个名为“dept”的字段，其中包含文档的部门信息，则可以将其映射到名为“Department”的索引字段。这样，您就可以在查询文档时使用该字段。

您还可以映射 Amazon Kendra 保留字段或常用字段，例如_created_at。如果您的数据源有一个名为 “creation_date” 的字段，则可以将其映射到名为的等效 Amazon Kendra 保留字段。_created_at有关 Amazon Kendra 保留字段的更多信息，请参阅文档属性或字段。

您可以映射大多数数据来源的字段。您可以为以下数据来源创建字段映射：

- Adobe Experience Manager
- Alfresco
- Aurora (MySQL)
- Aurora (PostgreSQL)
- Amazon FSx （视窗）

- Amazon FSx（NetApp ONTAP）
- Amazon RDS/Aurora
- Amazon RDS (Microsoft SQL Server)
- Amazon RDS (MySQL)
- Amazon RDS (Oracle)
- Amazon RDS (PostgreSQL)
- Amazon Kendra 网络爬虫
- Amazon WorkDocs
- Box
- Confluence
- Dropbox
- Drupal
- GitHub
- Google Workspace Drives
- Gmail
- IBM DB2
- Jira
- Microsoft Exchange
- 微软 OneDrive
- 微软 SharePoint
- Microsoft Teams
- Microsoft SQL Server
- Microsoft Yammer
- MySQL
- Oracle 数据库
- PostgreSQL
- Quip
- Salesforce

- ServiceNow
- Slack
- Zendesk

如果您将文档存储在 S3 存储桶或 S3 数据来源中，则可以使用 JSON 元数据文件指定字段。有关更多信息，请参阅 S3 数据来源连接器。

将数据来源字段映射到索引字段分为三个步骤：

1. 创建索引。有关更多信息，请参阅创建索引。
2. 更新索引以添加字段。
3. 创建数据源并添加字段映射以将保留字段和任何自定义字段映射到 Amazon Kendra 索引字段。

要更新索引以添加自定义字段，请使用控制台编辑数据源字段映射并添加自定义字段或使用 UpdateIndexAPI。您一共可以向索引添加 500 个自定义字段。

对于数据库数据来源，如果数据库列的名称与保留字段的名称匹配，则会自动映射该字段和列。

使用 UpdateIndexAPI，您可以使用添加保留字段和自定义字段`DocumentMetadataConfigurationUpdates`。

以下 JSON 示例使用 `DocumentMetadataConfigurationUpdates` 来向索引添加名为“Department”的字段。

```
"DocumentmetadataConfigurationUpdates": [
   {
       "Name": "Department",
       "Type": "STRING_VALUE"
   }
]
```

创建字段时，您可以选择设置字段的搜索方式。可从以下选项中进行选择：

- 可显示 - 确定是否在查询响应中返回字段。默认值为 `true`。
- 可分面 - 指示字段可用于创建分面。默认值为 `false`。
- 可搜索 - 确定是否在搜索中使用该字段。对于字符串字段，默认值为 `true`；对于数字和日期字段，默认值为 `false`。

- 可排序 - 指示可使用该字段对搜索结果进行排序。只能对日期、数字和字符串字段进行设置。无法为字符串列表字段进行设置。

以下 JSON 示例使用 DocumentMetadataConfigurationUpdates 来向索引添加名为“Department”的字段并将其标记为可分面。

```
"DocumentMetadataConfigurationUpdates": [
   {
       "Name": "Department",
       "Type": "STRING_VALUE",
       "Search": {
           "Facetable": true
       }
   }
]
```

使用 Amazon Kendra 保留或常用文档字段

借助 UpdateIndex API，您可以使用DocumentMetadataConfigurationUpdates并指定要映射到等效文档属性/字段名称的 Amazon Kendra 保留索引字段名称来创建保留字段或常用字段。您还可以创建自定义字段。如果您使用数据源连接器，则大多数连接器都包含将数据源文档字段映射到 Amazon Kendra 索引字段的字段映射。如果您使用控制台，则要更新字段，方法是选择数据来源，选择编辑操作，然后在“字段映射”部分旁边继续配置数据来源。

您可以将 Search 对象配置为将字段设置为可显示、可分面、可搜索和可排序。您可以将 Relevance 对象配置为设置字段的排名顺序、提升持续时间或时间段，以应用于映射到特定字段值的提升、新鲜度、重要性值和重要性值。如果您使用控制台，则可以通过在导航菜单中选择 facet 选项来设置字段的搜索设置。要设置相关性调整，请在导航菜单中选择搜索索引的选项，输入查询，然后使用侧面板选项调整搜索相关性。创建字段后无法更改字段类型。

Amazon Kendra 有以下可供您使用的保留或常用文档字段：

- _authors - 负责文档内容的一位或多位作者名单。
- _category - 将文档置于特定组中的类别。
- _created_at - 以 ISO 8601 格式创建文档的日期和时间。例如，2012-03-25T12:30:10+01:00 是中部欧洲时间 2012 年 3 月 25 日中午 12:30（10 秒）的 ISO 8601 日期-时间格式。
- _data_source_id - 包含文档数据来源的标识符。
- _document_body - 文档的内容。

- _document_id - 文档的唯一标识符。
- _document_title - 文档标题。
- _excerpt_page_number - PDF 文件中显示文档摘录的页码。如果您的索引是在 2020 年 9 月 8 日之前创建的，则必须重新编制文档索引才能使用此属性。
- _faq_id - 如果这是问答类型文档 (FAQ)，则为常见问题解答的唯一标识符。
- _file_type - 文档的文件类型，例如 pdf 或 doc。
- _last_updated_at - 上次更新端点的日期和时间，采用 ISO 8601 格式。例如，2012-03-25T12:30:10+01:00 是中部欧洲时间 2012 年 3 月 25 日中午 12:30（10 秒）的 ISO 8601 日期-时间格式。
- _source_uri - 文档可用的 URI。例如，公司网站上的文档的 URI。
- _version - 文档特定版本的标识符。
- _view_count - 查看文档的次数。
- _language_code（字符串）-适用于文档的语言的代码。如果您未指定语言，默认为英语。有关支持的语言（包括其代码）的更多信息，请参阅添加非英语语言文档。

对于自定义字段，您可以将 DocumentMetadataConfigurationUpdates 与 UpdateIndex API 配合使用来创建这些字段，就像创建保留字段或公用字段时一样。您必须为自定义字段设置相应的数据类型。如果您使用控制台，则要更新字段，方法是选择数据来源，选择编辑操作，然后在“字段映射”部分旁边继续配置数据来源。某些数据来源不支持添加新字段或自定义字段。创建字段后无法更改字段类型。

以下是您可以为自定义字段设置的类型：

- Date
- 数字
- 字符串
- 字符串列表

如果您使用 BatchPutDocumentAPI 将文档添加到索引，则会Attributes列出文档的字段/属性，然后使用该DocumentAttribute对象创建字段。

对于从 Amazon S3 数据源编制索引的文档，您可以使用包含字段信息的 JSON 元数据文件创建字段。

如果您使用支持的数据库作为数据来源，则可以使用字段映射选项配置字段。

添加非英语语言的文档

您可以为多种语言的文档编制索引。如果未指定语言，则在默认情况下， Amazon Kendra 会使用英语为文档编制索引。您可以将文档的语言代码作为字段包含在文档元数据中。有关文档的 字段的更多信息，请参阅_language_code字段映射和自定义属性。

在调用时，您可以为数据源中的所有文档指定语言代码CreateDataSource。如果文档没有在元数据字段中指定语言代码，则使用为数据来源级别的所有文档指定的语言代码为该文档编制索引。在控制台中，您只能在数据来源级别使用支持的语言对文档进行索引。转到数据来源，然后进入指定数据来源详细信息页面，然后从语言下拉列表中选择一种语言。

您也可以用支持的语言搜索或查询文档。有关更多信息，请参阅搜索语言。

支持以下语言及其代码（如果不指定语言，则默认支持英语或 en）。此表包括 Amazon Kendra 支持完整语义搜索的语言，以及仅支持简单关键字匹配的语言。在下表中，支持完全语义搜索的语言用星号标记，并以粗体文本显示。完全语义搜索还支持英语（默认语言）。

语言名称	语言代码
阿拉伯语	ar
亚美尼亚语	hy
巴斯克语	eu
孟加拉语	bn
保加利亚语	bg
加泰罗尼亚语	ca
中文 - 简体和繁体*	zh
捷克语	cs
丹麦语	da
荷兰语	nl
芬兰语	fi

语言名称	语言代码
法语 - 包括法语（加拿大）*	fr
加利西亚语	gl
德语*	de
希腊语	el
印地语	hi
匈牙利语	hu
印度尼西亚语	id
爱尔兰语	ga
意大利语	it
日语*	ja
韩语*	ko
拉脱维亚语	lv
立陶宛语	lt
挪威语	no
波斯语	fa
葡萄牙语	pt
葡萄牙语（巴西）*	pt-BR
罗马尼亚语	ro
俄语	ru
中库尔德语	ckb

语言名称	语言代码
西班牙语 - 包括西班牙语（墨西哥）*	es
瑞典语	sv
土耳其语	tr

*该语言支持语义搜索。

对于支持语义搜索的语言，支持以下功能。

- 文档相关性不仅仅是简单的关键字匹配。
- 除了简单的关键字匹配之外的常见问题解答。
- 根据阅读理解从文档中提取答案。 Amazon Kendra
- 搜索结果的置信分段（非常高、高、中和低）。

对于不支持语义搜索的语言，文档相关性和常见问题解答支持简单的关键字匹配。

仅英语（默认语言）支持同义词（包括自定义同义词）、增量学习和反馈以及查询建议。

配置 Amazon Kendra 为使用 Amazon VPC

Amazon Kendra 可以连接到您创建的虚拟私有云 (VPC)， Amazon Virtual Private Cloud 以索引存储在私有云中运行的数据源中的内容。创建数据源连接器时，您可以为包含您的数据源的子网提供安全组和子网标识符。利用这些信息， Amazon Kendra 创建一个 elastic network 接口，用于与您的 VPC 内的数据源进行安全通信。

要使用设置 Amazon Kendra 数据源连接器 Amazon VPC，您可以使用 AWS Management Console 或 CreateDataSourceAPI 操作。如果您使用控制台，则可以在连接器配置过程中连接 VPC。

> **Note**
>
> 设置 Amazon Kendra 数据源连接器时，该 Amazon VPC 功能是可选的。如果您的数据源可以从公共互联网访问，则无需启用该 Amazon VPC 功能。并非所有 Amazon Kendra 数据源连接器都支持 Amazon VPC。

如果您的数据源未运行 Amazon VPC 且无法通过公共 Internet 进行访问，则首先使用虚拟专用网络 (VPN) 将您的数据源连接到 VPC。然后，您可以使用 Amazon VPC 和的组合将您的数据源连接到 Amazon Kendra AWS Virtual Private Network。有关设置 VPN 的信息，请参阅AWS VPN 文档。

主题

配置对 Amazon Kendra 连接器的 Amazon VPC 支持

要配置 Amazon VPC 为与 Amazon Kendra 连接器配合使用，请执行以下步骤。

步骤

第 1 步。为创建 Amazon VPC 子网 Amazon Kendra

创建或选择 Amazon Kendra 可用于访问您的数据源的现有 Amazon VPC 子网。准备好的子网必须位于以下 AWS 区域 可用区之一：

- 美国西部（俄勒冈州）/us-west-2 - usw2-azz1、usw2-az2、usw2-az3
- 美国东部（弗吉尼亚州北部）/us-eaz-1 - useaz1、use1-az2、use1-az4
- 美国东部（俄亥俄州）/us-eaz-eaz-2 - useaz1、use2-azz2、use2-az3
- 亚太地区（东京）/ap-northeast-1 - apne1-azz1、apne1-azz2、apne1-azz4
- 亚太地区（孟买）/ap-south-1 - aps1-azz1、aps1-azz2、aps1-az3
- 亚太地区（新加坡）/ap-southeasteast-1 - apse1-azz1、apse1-azz2、apse1-az3
- 亚太地区（悉尼）/ap-southeasteast-2 - apse2-azz1、apse2-azz2、apse2-azz3
- 加拿大（中部）/ca-central-1 - cacz1、cac1-azz2、cac1-azz4
- 欧洲（爱尔兰）/euwest-1 - euww1-azz1、uew1-azz2、euw1-az3

- 欧洲（伦敦）/eu-west-2 - usw2-azz1、usw2-az2、usw2-az3

必须能够从您提供给 conn Amazon Kendra ector 的子网访问您的数据源。

有关如何配置 Amazon VPC 子网的更多信息，请参阅 Amazon VPC 用户指南 Amazon VPC中的适合您的子网。

如果 Amazon Kendra 必须在两个或多个子网之间路由连接，则可以准备多个子网。例如，包含您的数据源的子网没有 IP 地址。在这种情况下，您可以提供 Amazon Kendra 一个额外的子网，该子网具有足够的 IP 地址并连接到第一个子网。如果您列出多个子网，则子网必须能够相互通信。

第 2 步。为以下 Amazon VPC 对象创建安全组 Amazon Kendra

要将您的 Amazon Kendra 数据源连接器连接到 Amazon VPC，您必须从 VPC 准备一个或多个要分配给的安全组 Amazon Kendra。安全组将与创建的 elastic network 接口相关联 Amazon Kendra。此网络接口控制访问 Amazon VPC 子网 Amazon Kendra 时进出的入站和出站流量。

确保您的安全组的出站规则允许来自 Amazon Kendra 数据源连接器的流量访问子网和要与之同步的数据源。例如，您可以使用MySQL连接器从MySQL数据库进行同步。如果您使用的是默认端口，则安全组必须 Amazon Kendra 允许访问运行数据库的主机上的端口 3306。

我们建议您使用以下值配置默认安全组 Amazon Kendra 以供使用：

- 入站规则-如果您选择将其留空，则所有入站流量都将被阻止。
- 出站规则-添加一条规则以允许所有出站流量，这样 Amazon Kendra 就可以从您的数据源发起同步请求。
  - IP 版本 — IPv4
  - 类型-所有流量
  - 协议-所有流量
  - 端口范围-全部
  - 目的地 — 0.0.0.0/0

有关如何配置 Amazon VPC 安全组的更多信息，请参阅 Amazon VPC 用户指南中的安全组规则。

第 3 步。配置您的外部数据源和 Amazon VPC

确保您的外部数据源具有正确的访问权限配置和网络设置。 Amazon Kendra 您可以在每个连接器页面的先决条件部分中找到有关如何配置数据源的详细说明。

此外，请检查您的 Amazon VPC 设置，并确保可以从您要分配的子网访问您的外部数据源 Amazon Kendra。为此，我们建议您在同一子网中创建具有相同安全组的 Amazon EC2 实例，并测试该 Amazon EC2 实例对数据源的访问权限。有关更多信息，请参阅Amazon VPC 连接疑难解答。

设置要连接 Amazon Kendra 的数据源 Amazon VPC

在中添加新数据源时 Amazon Kendra，如果所选数据源连接器支持此 Amazon VPC 功能，则可以使用该功能。

您可以使用 AWS Management Console 或 Amazon Kendra API 在 Amazon VPC 启用状态下设置新的 Amazon Kendra 数据源。具体而言，使用 CreateDataSourceAPI 操作，然后使用`VpcConfiguration`参数提供以下信息：

- `SubnetIds`— Amazon VPC 子网标识符列表
- `SecurityGroupIds`— Amazon VPC 安全组标识符列表

如果您使用控制台，则需要在连接器配置期间提供所需的 Amazon VPC 信息。要使用控制台为连接器启用 Amazon VPC 功能，您需要先选择一个 Amazon VPC。然后，您提供任何 Amazon VPC 子网的标识符和任何 Amazon VPC 安全组的标识符。您可以选择在配置 Amazon VPC 中创建的 Amazon VPC 子网和 Amazon VPC 安全组，也可以使用任何现有的子网和 Amazon VPC 安全组。

主题

- 查看 Amazon VPC 标识符
- 检查您的数据源 IAM 角色

查看 Amazon VPC 标识符

子网和安全组的标识符是在 Amazon VPC 控制台中配置的。要查看标识符，请按以下步骤操作。

查看子网标识符

1. 登录 AWS Management Console 并打开亚马逊 VPC 控制台，网址为 https://console.aws.amazon.com/vpc/。
2. 在导航窗格中，选择子网。
3. 从子网列表中，选择包含您的数据库服务器的子网。
4. 在详细信息选项卡中，记下子网 ID 字段中的标识符。

查看安全组标识符

1. 登录 AWS Management Console 并打开亚马逊 VPC 控制台，网址为 https://console.aws.amazon.com/vpc/。
2. 从导航窗格中选择 “安全组”。
3. 从安全组列表中，选择要为其提供标识的组。
4. 在详细信息选项卡中，记下安全组 ID 字段中的标识符。

检查您的数据源 IAM 角色

确保您的数据源 AWS Identity and Access Management IAM（连接器）角色包含访问您的权限 Amazon VPC。

如果您使用控制台为角色创建新 IAM 角色，则 Amazon Kendra 会自动代表您为 IAM 角色添加正确的权限。如果您使用 API 或使用现有 IAM 角色，请检查您的角色是否包含访问权限 Amazon VPC。要验证您是否拥有正确的权限，请参阅 VPC 的IAM 角色。

您可以修改现有数据源以使用不同的子 Amazon VPC 网。但是，请检查数据源的 IAM 角色，并在必要时对其进行修改以反映 Amazon Kendra 数据源连接器正常工作的更改。

在 VPC 中连接到数据库

以下示例说明如何连接在虚拟私有云 (VPC) 中运行MySQL的数据库。该示例假设您从默认 VPC 开始，并且需要创建MySQL数据库。如果您已有 VPC，请确保其配置如图所示。如果您有MySQL数据库，则可以使用它来代替创建新数据库。

步骤

- 步骤 1：配置 VPC
- 步骤 2：创建和配置安全组
- 步骤 3：创建数据库
- 步骤 4：创建数据源连接器

步骤 1：配置 VPC

配置您的 VPC，以便您拥有私有子网和安全组 Amazon Kendra ，用于访问子网中运行MySQL的数据库。VPC 配置中提供的子网必须位于美国西部（俄勒冈）区域、美国东部（弗吉尼亚北部）区域或欧洲（爱尔兰）区域。

使用配置 VPC Amazon VPC

1. 登录 AWS Management Console 并打开亚马逊 VPC 控制台，网址为 https://console.aws.amazon.com/vpc/。
2. 在导航窗格中，选择路由表，然后选择创建路由表。
3. 在“名称”字段中输入**`Private subnet route table`**。从 VPC 下拉列表中，选择您的 VPC，然后选择创建路由表。选择关闭以返回路由表列表。
4. 从导航窗格中选择 NAT 网关，然后选择创建 NAT 网关。
5. 从“子网”下拉列表中，选择作为公有子网的子网。记下子网 ID。
6. 如果您没有弹性 IP 地址，请选择创建新 EIP，选择创建 NAT 网关，然后选择关闭。
7. 在导航窗格中，选择路由表。
8. 从路由表列表中，选择您在步骤 3 中创建的私有子网路由表。在操作中，选择编辑路线。
9. 选择 Add route（添加路由）。对于目的地，输入**`0.0.0.0/0`**以允许所有传出流量进入互联网。对于目标，选择 NAT 网关，然后选择您在步骤 4 中创建的网关。选择“保存更改”，然后选择“关闭”。
10. 从操作菜单中选择编辑子网关联。
11. 选择要设为私有的子网。不要选择之前记下的 NAT 网关的子网。完成后，选择“保存关联”。

步骤 2：创建和配置安全组

接下来，为您的数据库配置安全组。

创建和配置安全组

1. 登录 AWS Management Console 并打开亚马逊 VPC 控制台，网址为 https://console.aws.amazon.com/vpc/。
2. 在您的 VPC 描述中，记下 IPv4 CIDR。
3. 在导航窗格中，选择安全组，然后选择创建安全组。
4. 对于安全组名称，输入 **`DataSourceInboundSecurityGroup`**。提供描述，然后从列表中选择 VPC。选择“创建安全组”，然后选择“关闭”。
5. 选择入站规则选项卡。
6. 选择“编辑入站规则”，然后选择“添加规则”
7. 对于数据库，请输入端口范围的端口号。例如，对于MySQL它是**`3306`**，对于 HTTPS 来说，则是**`443`**。对于源，键入您的 VPC 的无类别域间路由 (CIDR)。选择保存规则，然后选择关闭。

安全组允许 VPC 中的任何用户连接到数据库，并允许出站连接到 Internet。

步骤 3：创建数据库

创建一个数据库来保存您的文档，或者您也可以使用现有的数据库。

有关如何创建MySQL数据库的说明，请参见MySQL。

步骤 4：创建数据源连接器

配置 VPC 并创建数据库后，您可以为数据库创建数据源连接器。有关支持的数据库连接器的信息，请参阅支持的连接器。 Amazon Kendra

对于您的数据库，请务必配置您的 VPC、您在 VPC 中创建的私有子网以及在 VPC 中创建的安全组。

排除 VPC 连接问题

如果您的虚拟私有云 (VPC) 连接遇到任何问题，请检查您的 IAM 权限、安全组设置和子网的路由表是否配置正确。

数据源连接器同步失败的一个潜在原因是，可能无法从您分配的子网访问数据源。 Amazon Kendra 要解决此问题，我们建议您使用相同的 Amazon VPC 设置创建 Amazon EC2 实例。然后，尝试使用 REST API 调用或其他方法（基于数据源的特定类型）从该 Amazon EC2 实例访问数据源。

如果您成功地从您创建的 Amazon EC2 实例访问数据源，则意味着可以从该子网访问您的数据源。因此，您的同步问题与无法访问您的数据源无关。 Amazon VPC

如果您无法从 VPC 配置访问您的 Amazon EC2 实例，也无法使用您创建的 Amazon EC2 实例对其进行验证，则需要进一步排除故障。例如，如果您的连接器由于 Amazon S3 连接问题错误而同步失败，则可以使用分配给 Amazon S3 连接器的相同 Amazon VPC 配置来设置 Amazon EC2 实例。然后，使用此 Amazon EC2 实例来测试您的 Amazon VPC 设置是否正确。

以下是设置 Amazon EC2 实例以排除与 Amazon S3 数据源的 Amazon VPC 连接故障的示例。

主题

步骤 1：启动实 Amazon EC2 例

1. 登录 AWS Management Console 并打开亚马逊 EC2 控制台，网址为 https://console.aws.amazon.com/ec2/。
2. 选择启动实例。
3. 选择 “网络设置”，然后选择 “编辑”，然后执行以下操作：

   a. 选择与您分配的相同 VPC 和子网 Amazon Kendra。

   b. 对于防火墙（安全组），选择选择现有安全组。然后，选择您分配给的安全组 Amazon Kendra。

   > **Note**
   >
   > 安全组应允许出站流量进入 Amazon S3。

   c. 将 “自动分配公有 IP” 设置为 “禁用”。

   d. 在 “高级详细信息” 中，执行以下操作：

   - 对于 IAM 实例配置文件，选择创建新 IAM 配置文件以创建 IAM 实例配置文件并将其附加到您的实例。确保该配置文件具有访问权限 Amazon S3。有关更多信息，请参阅如何向我的 Amazon EC2 实例授予对 Amazon S3 存储桶的访问权限？ 在 AWS re:Post。
   - 将所有其他设置保留为默认设置。

   e. 查看并启动实 Amazon EC2 例。

第 2 步：Connect 连接到 Amazon EC2 实例

Amazon EC2 实例运行后，前往您的实例详细信息页面并连接到您的实例。为此，请使用 Linux 实例 Amazon EC2 用户指南中的 EC2 Instance Connect 终端节点无需公有 IPv4 地址即可连接到您的实例。

步骤 3：测试 Amazon S3 访问权限

连接到 Amazon EC2 实例终端后，运行 AWS CLI 命令以测试从该私有子网到您的 Amazon S3 存储桶的连接。

要测试 Amazon S3 访问权限， AWS CLI 请在中键入以下命令 AWS CLI：`aws s3 ls`

AWS CLI 命令运行后，请查看以下内容：

- 如果您已正确设置必要的 IAM 权限并且 Amazon S3 设置正确，则应该会看到 Amazon S3 存储桶列表。
- 如果您看到权限错误`Access Denied`，例如，您的 VPC 配置可能正确，但您的 IAM 权限或 Amazon S3 存储桶策略有问题。

如果命令超时，则可能是因为您的 VPC 设置不正确，并且 Amazon EC2 实例无法从您的子网访问 Amazon S3 而导致连接超时。请重新配置您的 VPC，然后重试。

删除索引、数据来源或批量上传的文档

本节介绍如何删除索引、索引中文档的数据来源存储库或索引中批量上传的文档。

主题

删除索引

不再使用索引时，您可以从 Amazon Kendra 中将其删除。例如，在以下情况下删除索引：

- 不再使用该索引，并希望减少 AWS 账户产生的费用。无论您是否对 Amazon Kendra 索引进行查询，索引在运行时都会产生费用。
- 您想为不同的 Amazon Kendra 版本重新配置索引。删除现有索引，然后使用不同的版本创建新索引。
- 您已达到账户中的最大索引数量，并且不想超过限额。删除现有索引，然后添加新索引。有关可创建的最大索引数量的信息，请参阅限额。

要删除索引，请使用控制台 AWS Command Line Interface、、AWS CloudFormation 脚本或 `DeleteIndex` API。删除索引会移除索引以及所有关联的数据来源和文档数据。删除索引不会从存储中移除原始文档。

删除索引是一项异步操作。开始删除索引时，索引状态将更改为 `DELETING`。在删除所有与索引相关的信息之前，它一直处于 `DELETING` 状态。删除索引后，它不会再出现在调用 ListIndices API 的结果中。如果您使用已删除的索引标识符调用 DescribeIndex API，则会收到 `ResourceNotFound` 异常。

删除索引（控制台）

1. 登录到 AWS Management Console，然后通过以下网址打开 Amazon Kendra 控制台：https://console.aws.amazon.com/kendra/。
2. 在导航窗格中选择索引，然后选择要删除的索引。
3. 选择删除，以便删除所选索引。

删除索引 (CLI)

- 在 AWS CLI 中，使用下列命令。该命令针对 Linux 和 macOS 编排了格式。如果您使用 Windows，请将 Unix 行继续符 (\) 替换为脱字号 (^)。

```
aws kendra delete-index \
   --id index-id
```

删除数据来源

如果要从 Amazon Kendra 索引中移除数据来源中包含的信息，则可以删除该数据来源。例如，在以下情况下删除数据来源：

- 数据来源配置不正确。删除数据来源，等待数据来源完成删除，然后重新创建。
- 您将文档从一个数据来源迁移到另一个数据来源。删除原始数据来源并在新位置重新创建。
- 您已达到索引的数据来源数限制。删除一个现有数据来源并添加一个新数据来源。有关可创建的数据来源数量的更多信息，请参阅配额。

要删除数据来源，请使用控制台、AWS Command Line Interface (AWS CLI)、DeleteDataSource API 或 AWS CloudFormation 脚本。删除数据来源会从索引中移除有关该数据来源的所有信息。如果您只想停止同步数据来源，请将数据来源的同步计划更改为“按需运行”。

删除数据来源是一项异步操作。开始删除数据来源时，数据来源状态将更改为 DELETING。在删除与数据来源相关的信息之前，它一直处于 DELETING 状态。删除数据来源后，该数据来源将不再出现在调用 ListDataSources 的结果中。如果您使用已删除数据来源的标识符调用 DescribeDataSource API，则会收到 ResourceNotFound 异常。

> **Note**
>
> 从数据来源中删除特定文档后，删除整个数据来源或重新同步索引可能需要一小时或更长时间，具体取决于要删除的文档数量。

删除数据来源（控制台）

1. 登录到 AWS Management Console，然后通过以下网址打开 Amazon Kendra 控制台：https://console.aws.amazon.com/kendra/。

2. 在导航窗格中，选择索引，然后选择包含要删除的数据来源的索引。
3. 在导航窗格中，选择 Data sources（数据来源）。
4. 选择要移除的数据来源。
5. 选择删除以删除数据来源。

删除数据来源 (CLI)

- 在 AWS Command Line Interface 中，使用下列命令。该命令针对 Linux 和 macOS 编排了格式。如果您使用 Windows，请将 Unix 行继续符 (\) 替换为脱字号 (^)。

```
aws kendra delete-data-source \
   --id data-source-id \
   --index-id index-id
```

删除数据来源时，Amazon Kendra 会删除有关该数据来源的所有存储信息。Amazon Kendra 会移除索引中存储的所有文档数据，以及与数据来源关联的所有运行历史记录和指标。删除数据来源并不会从存储中移除原始文档。

当 Amazon Kendra 删除数据来源时，数据来源中的文档可能会包含在 `DescribeIndex` API 返回的文档数量中。当 Amazon Kendra 删除数据来源时，数据来源中的文档可能会出现在搜索结果中。

只要在控制台中调用 `DeleteDataSource` API 或选择删除数据来源，Amazon Kendra 就会立即释放该数据来源的资源。如果您要删除数据来源以将数据来源数量减少到限制以下，则可以立即创建新的数据来源。

如果您要删除一个数据来源，然后为该文档数据创建另一个数据来源，请等待第一个数据来源删除完成，然后同步新的数据来源。

您可以删除正在与 Amazon Kendra 同步的数据来源。同步已停止，数据来源已删除。如果在删除数据来源时尝试启动同步，则会出现 `ConflictException` 异常。

如果关联的索引处于 `DELETING` 状态，则无法删除数据来源。如果删除索引，则会删除该索引的所有数据来源。当索引的数据来源处于 `DELETING` 状态时，您可以开始删除该索引。

如果您有两个数据来源指向相同的文档，例如，两个数据来源指向同一个 Amazon S3 存储桶，则删除其中一个数据来源时，索引中的文档可能会出现不一致。当两个数据来源引用相同的文档时，索引中仅存储文档数据的一个副本。移除一个数据来源会移除文档的索引数据。另一个数据来源不知道这些文档

已删除，因此，在下次同步时，Amazon Kendra 无法正确地重新为文档编制索引。当有两个数据来源指向同一个文档位置时，您应删除这两个数据来源，然后重新创建一个数据来源。

删除批量上传的文档

您可以使用 BatchDeleteDocument API 直接从索引中删除文档。您无法使用控制台直接删除文档。如果您使用控制台，则可以从数据来源存储库中删除特定文档并与索引重新同步，也可以删除整个数据来源连接器。

使用 BatchDeleteDocument 删除索引中的文档是一项异步操作。调用 BatchDeleteDocument API 后，您可以使用 BatchGetDocumentStatus API 来监控删除文档的进度。从索引中删除文档时，Amazon Kendra 会返回 NOT_FOUND 作为状态。

> **Note**
>
> 使用 BatchDeleteDocument 从索引中删除文档可能需要一个小时或更长时间，具体取决于要删除的文档数量。

从索引中删除批量上传的文档 (CLI)

- 在 AWS Command Line Interface 中，使用下列命令。该命令针对 Linux 和 macOS 编排了格式。如果您使用 Windows，请将 Unix 行继续符 (\) 替换为脱字号 (^)。

```
aws kendra batch-delete-document \
   --index-id index-id \
   --document-id-list 'doc-id-1' 'doc-id-2'
```

在提取过程中丰富您的文档

在文档提取过程中，您可以更改内容和文档元数据字段或属性。借 Amazon Kendra 的自定义文档富集功能，当您将文档收录到 Amazon Kendra 中时，您可以创建、修改或删除文档属性和内容。这意味着您可以根据需要操作和提取数据。

此功能使您可以控制文档的处理和提取到 Amazon Kendra 的方式。例如，在将文档提取到 Amazon Kendra 中时，您可以清除文档元数据中的个人身份信息。

使用此功能的另一种方法是在 AWS Lambda 中调用 Lambda 函数，对图像运行光学字符识别 (OCR)、对文本进行翻译以及执行其他任务，以准备用于搜索或分析的数据。例如，您可以调用函数对图像运行 OCR。该函数可以解释图像中的文本，并将每张图像视为文本文档。接收邮寄的客户调查并将这些调查作为图像存储的公司可以将这些图像作为文本文档提取到 Amazon Kendra 中。然后，公司可以在 Amazon Kendra 中搜索有价值的客户调查信息。

您可以使用基本操作作为数据的首次解析，然后使用 Lambda 函数对数据应用更复杂的操作。例如，您可以使用基本操作简单地删除文档元数据字段“Customer_ID”中的所有值，然后应用 Lambda 函数从文档中的文本图像中提取文本。

自定义文档富集功能的工作原理

自定义文档富集的总体过程如下：

1. 在创建或更新数据来源时，或者直接将文档编入 Amazon Kendra 索引时，可以配置自定义文档富集。
2. Amazon Kendra 应用内联配置或基本逻辑来更改您的数据。有关更多信息，请参阅the section called “更改元数据的基本操作”。
3. 如果您选择配置高级数据操作，则 Amazon Kendra 可以将其应用于原始原始文档或结构化的、已解析的文档。有关更多信息，请参阅the section called “Lambda 函数：提取和更改元数据或内容”。
4. 您修改过的文档会被提取到 Amazon Kendra 中。

在此过程的任何时候，如果您的配置无效，Amazon Kendra 都会引发错误。

当您调用 CreateDataSource、UpdateDataSource 或 BatchPutDocument API 时，需要提供自定义文档富集配置。如果您调用 `BatchPutDocument`，则必须为每个请求配置自定义文档富集功能。如果您使用控制台，则可以选择您的索引，然后选择文档丰富来配置 “自定义文档富集”。

如果您在控制台中使用文档富集功能，则可以选择仅配置基本操作或仅配置 Lambda 函数，或者同时配置两者，就像使用 API 一样。您可以在控制台步骤中选择下一步，选择不配置基本操作，只配置 Lambda 函数，包括应用于原始（提取前）数据还是结构化（提取后）数据。您只能通过完成控制台中的所有步骤来保存配置。如果您未完成所有步骤，则不会保存您的文档配置。

更改元数据的基本操作

您可以使用基本逻辑操作文档字段和内容。这包括移除字段中的值、使用条件修改字段中的值或创建字段。要进行超出基本逻辑操作范围的高级操作，请调用 Lambda 函数。有关更多信息，请参阅the section called “Lambda 函数：提取和更改元数据或内容”。

要应用基本逻辑，请使用 DocumentAttributeTarget 对象来指定要操作的目标字段。您提供属性键。例如，“Department”键是一个字段或属性，其中包含与文档关联的所有部门名称。如果满足特定条件，您还可以指定要在目标字段中使用的值。您可以使用 DocumentAttributeCondition 对象设置条件。例如，如果“source_URI”字段的 URI 值包含“financial”，则使用文档的“Finance”目标值预填写“Department”目标字段。您也可以删除目标文档属性值。

要使用控制台应用基本逻辑，请选择您的索引，然后在导航菜单中选择文档富集。转至配置基本操作，将基本操作应用于您的文档字段和内容。

以下是使用基本逻辑删除名为“Customer_ID”的文档字段中所有客户标识号的示例。

示例 1：删除与文件关联的客户识别码

应用基本操作之前的数据。

Document_ID	Body_Text	Customer_ID
1	Lorem Ipsum。	CID1234
2	Lorem Ipsum。	CID1235
3	Lorem Ipsum。	CID1236

应用基本操作后的数据。

Document_ID	Body_Text	Customer_ID
1	Lorem Ipsum。	

Document_ID	Body_Text	Customer_ID
2	Lorem Ipsum。	
3	Lorem Ipsum。	

以下是使用基本逻辑创建名为“Department”的字段，并根据来自“Source_URI”字段的信息在该字段中预填部门名称的示例。如果“source_URI”字段的 URI 值包含“financial”，则使用文档的“Finance”目标值预填写“Department”目标字段。

示例 2：创建“Department”字段，并使用条件在其中预填与文档关联的部门名称。

应用基本操作之前的数据。

Document_ID	Body_Text	Source_URI
1	Lorem Ipsum。	financial/1
2	Lorem Ipsum。	financial/2
3	Lorem Ipsum。	financial/3

应用基本操作后的数据。

Document_ID	Body_Text	Source_URI	Department
1	Lorem Ipsum。	financial/1	财务
2	Lorem Ipsum。	financial/2	财务
3	Lorem Ipsum。	financial/3	财务

> **Note**
>
> 如果目标文档字段尚未创建并作为索引字段，则 Amazon Kendra 无法创建该字段。创建索引字段后，您可以使用 DocumentAttributeTarget 创建文档字段。然后，Amazon Kendra 会将您新创建的文档元数据字段映射到您的索引字段。

以下代码是配置基本数据操作以删除与文档关联的客户识别码的示例。

Console

配置基本数据操作以删除客户识别码

1. 在左侧导航窗格的索引下，选择文档富集，然后选择添加文档富集。
2. 在配置基本操作页面上，从下拉列表中选择要更改文档字段和内容的数据来源。然后从下拉列表中选择文档字段名称“Customer_ID”，从下拉列表中选择索引字段名称“Customer_ID”，然后从下拉列表中选择目标操作删除。然后选择添加基本操作。

CLI

配置基本数据操作以删除客户识别码

```
aws kendra create-data-source \
 --name data-source-name \
 --index-id index-id \
 --role-arn arn:aws:iam::account-id:role/role-name \
 --type S3 \
 --configuration '{"S3Configuration":{"BucketName":"S3-bucket-name"}}' \
 --custom-document-enrichment-configuration '{"InlineConfigurations":[{"Target":
{"TargetDocumentAttributeKey":"Customer_ID", "TargetDocumentAttributeValueDeletion":
 true}}]}'
```

Python

配置基本数据操作以删除客户识别码

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time
```

```
kendra = boto3.client("kendra")

print("Create a data source with customizations")

# Provide the name of the data source
name = "data-source-name"
# Provide the index ID for the data source
index_id = "index-id"
# Provide the IAM role ARN required for data sources
role_arn = "arn:aws:iam::${account-id}:role/${role-name}"
# Provide the data source connection information
data_source_type = "S3"
S3_bucket_name = "S3-bucket-name"
# Configure the data source with Custom Document Enrichment
configuration = {"S3Configuration":
        {
            "BucketName": S3_bucket_name
        }
    }
custom_document_enrichment_configuration = {"InlineConfigurations":[
        {
            "Target":{"TargetDocumentAttributeKey":"Customer_ID",
                      "TargetDocumentAttributeValueDeletion": True}
        }]
    }

try:
    data_source_response = kendra.create_data_source(
        Name = name,
        IndexId = index_id,
        RoleArn = role_arn,
        Type = data_source_type
        Configuration = configuration
        CustomDocumentEnrichmentConfiguration =
 custom_document_enrichment_configuration
    )

    pprint.pprint(data_source_response)

    data_source_id = data_source_response["Id"]

    print("Wait for Amazon Kendra to create the data source with your
 customizations.")
```

```
    while True:
        # Get the details of the data source, such as the status
        data_source_description = kendra.describe_data_source(
            Id = data_source_id,
            IndexId = index_id
        )
        status = data_source_description["Status"]
        print(" Creating data source. Status: "+status)
        time.sleep(60)
        if status != "CREATING":
            break

    print("Synchronize the data source.")

    sync_response = kendra.start_data_source_sync_job(
        Id = data_source_id,
        IndexId = index_id
    )

    pprint.pprint(sync_response)

    print("Wait for the data source to sync with the index.")

    while True:

        jobs = kendra.list_data_source_sync_jobs(
            Id= data_source_id,
            IndexId= index_id
        )

        # For this example, there should be one job
        status = jobs["History"][0]["Status"]

        print(" Syncing data source. Status: "+status)
        time.sleep(60)
        if status != "SYNCING":
            break

except  ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

配置基本数据操作以删除客户识别码

```
package com.amazonaws.kendra;

import java.util.concurrent.TimeUnit;
import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.CreateDataSourceRequest;
import software.amazon.awssdk.services.kendra.model.CreateDataSourceResponse;
import software.amazon.awssdk.services.kendra.model.CreateIndexRequest;
import software.amazon.awssdk.services.kendra.model.CreateIndexResponse;
import software.amazon.awssdk.services.kendra.model.DataSourceConfiguration;
import software.amazon.awssdk.services.kendra.model.DataSourceStatus;
import software.amazon.awssdk.services.kendra.model.DataSourceSyncJob;
import software.amazon.awssdk.services.kendra.model.DataSourceSyncJobStatus;
import software.amazon.awssdk.services.kendra.model.DataSourceType;
import software.amazon.awssdk.services.kendra.model.DescribeDataSourceRequest;
import software.amazon.awssdk.services.kendra.model.DescribeDataSourceResponse;
import software.amazon.awssdk.services.kendra.model.DescribeIndexRequest;
import software.amazon.awssdk.services.kendra.model.DescribeIndexResponse;
import software.amazon.awssdk.services.kendra.model.IndexStatus;
import software.amazon.awssdk.services.kendra.model.ListDataSourceSyncJobsRequest;
import software.amazon.awssdk.services.kendra.model.ListDataSourceSyncJobsResponse;
import software.amazon.awssdk.services.kendra.model.S3DataSourceConfiguration;
import software.amazon.awssdk.services.kendra.model.StartDataSourceSyncJobRequest;
import software.amazon.awssdk.services.kendra.model.StartDataSourceSyncJobResponse;

public class CreateDataSourceWithCustomizationsExample {

    public static void main(String[] args) throws InterruptedException {
        System.out.println("Create a data source with customizations");

        String dataSourceName = "data-source-name";
        String indexId = "index-id";
        String dataSourceRoleArn = "arn:aws:iam::account-id:role/role-name";
        String s3BucketName = "S3-bucket-name"

        KendraClient kendra = KendraClient.builder().build();

        CreateDataSourceRequest createDataSourceRequest = CreateDataSourceRequest
            .builder()
            .name(dataSourceName)
            .description(experienceDescription)
```

```
            .roleArn(experienceRoleArn)
            .type(DataSourceType.S3)
            .configuration(
                DataSourceConfiguration
                    .builder()
                    .s3Configuration(
                        S3DataSourceConfiguration
                            .builder()
                            .bucketName(s3BucketName)
                            .build()
                    ).build()
            )
            .customDocumentEnrichmentConfiguration(
                CustomDocumentEnrichmentConfiguration
                    .builder()
                    .inlineConfigurations(Arrays.asList(
                        InlineCustomDocumentEnrichmentConfiguration
                            .builder()
                            .target(
                                DocumentAttributeTarget
                                    .builder()
                                    .targetDocumentAttributeKey("Customer_ID")
                                    .targetDocumentAttributeValueDeletion(true)
                                    .build())
                            .build()
                    )).build();

        CreateDataSourceResponse createDataSourceResponse =
kendra.createDataSource(createDataSourceRequest);
        System.out.println(String.format("Response of creating data source: %s",
createDataSourceResponse));

        String dataSourceId = createDataSourceResponse.id();
        System.out.println(String.format("Waiting for Kendra to create the data
source %s", dataSourceId));
        DescribeDataSourceRequest describeDataSourceRequest =
DescribeDataSourceRequest
            .builder()
            .indexId(indexId)
            .id(dataSourceId)
            .build();

        while (true) {
```

```
            DescribeDataSourceResponse describeDataSourceResponse =
kendra.describeDataSource(describeDataSourceRequest);

            DataSourceStatus status = describeDataSourceResponse.status();
            System.out.println(String.format("Creating data source. Status: %s",
status));
            TimeUnit.SECONDS.sleep(60);
            if (status != DataSourceStatus.CREATING) {
                break;
            }
        }

        System.out.println(String.format("Synchronize the data source %s",
dataSourceId));
        StartDataSourceSyncJobRequest startDataSourceSyncJobRequest =
StartDataSourceSyncJobRequest
            .builder()
            .indexId(indexId)
            .id(dataSourceId)
            .build();
        StartDataSourceSyncJobResponse startDataSourceSyncJobResponse =
kendra.startDataSourceSyncJob(startDataSourceSyncJobRequest);
        System.out.println(String.format("Waiting for the data
source to sync with the index %s for execution ID %s", indexId,
startDataSourceSyncJobResponse.executionId()));

        // For this example, there should be one job
        ListDataSourceSyncJobsRequest listDataSourceSyncJobsRequest =
ListDataSourceSyncJobsRequest
            .builder()
            .indexId(indexId)
            .id(dataSourceId)
            .build();

        while (true) {
            ListDataSourceSyncJobsResponse listDataSourceSyncJobsResponse =
kendra.listDataSourceSyncJobs(listDataSourceSyncJobsRequest);
            DataSourceSyncJob job = listDataSourceSyncJobsResponse.history().get(0);
            System.out.println(String.format("Syncing data source. Status: %s",
job.status()));

            TimeUnit.SECONDS.sleep(60);
            if (job.status() != DataSourceSyncJobStatus.SYNCING) {
                break;
```

```
            }

        }

        System.out.println("Data source creation with customizations is complete");
    }
}
```

Lambda 函数：提取和更改元数据或内容

您可以使用 Lambda 函数操作您的文档字段和内容。如果您想超越基本逻辑并应用高级数据操作，这很有用。例如，使用光学字符识别 (OCR)，它可以解释图像中的文本，并将每张图像视为文本文档。或者，检索特定时区的当前日期时间，然后在日期字段的空值处插入日期时间。

您可以先应用基本逻辑，然后使用 Lambda 函数进一步操作数据，反之亦然。您也可以选择仅应用 Lambda 函数。

Amazon Kendra 可以调用 Lambda 函数在提取过程中应用高级数据操作，这是 CustomDocumentEnrichmentConfiguration 的一部分。您指定的角色包括执行 Lambda 函数和访问 Amazon S3存储桶以存储数据操作IAM输出的权限，请参阅 访问角色。

Amazon Kendra 可以对原来的原始文档或结构化的解析文档应用 Lambda 函数。您可以使用 PreExtractionHookConfiguration 配置来配置一个 Lambda 函数，该函数会获取您的原始数据或原始数据，然后应用您的数据操作。您还可以使用 PostExtractionHookConfiguration 配置来配置一个 Lambda 函数，该函数获取您的结构化文档并应用您的数据操作。Amazon Kendra 提取文档元数据和文本以构建您的文档。您的 Lambda 函数必须遵循必需的请求和响应结构。有关更多信息，请参阅the section called “Lambda 函数的数据合约”。

要在控制台中配置 Lambda 函数，请选择您的索引，然后在导航菜单中选择文档富集。前往配置 Lambda 函数来配置 Lambda 函数。

您只能为 `PreExtractionHookConfiguration` 配置一个 Lambda 函数，只能为 `PostExtractionHookConfiguration` 配置一个 Lambda 函数。然而，Lambda 函数可在需要时调用其他函数。您可以同时配置 `PreExtractionHookConfiguration` 和/或 `PostExtractionHookConfiguration`。适用于 `PreExtractionHookConfiguration` 的 Lambda 函数的运行时间不得超过 5 分钟，适用于 `PostExtractionHookConfiguration` 的 Lambda 函数的运行时间不得超过 1 分钟。配置自定义文档扩充功能自然要比不配置时更长的时间才能将您的文档提取到 Amazon Kendra 中。

您可以配置 Amazon Kendra，以便仅在满足条件时调用 Lambda 函数。例如，如果日期-时间值为空，您可以指定一项条件，然后 Amazon Kendra 应调用函数以插入当前的日期-时间。

以下是使用 Lambda 函数运行 OCR 来解释图像中的文本并将此文本存储在名为“Document_Image_Text”的字段中的示例。

示例 1：从图像中提取文本以创建文本文档

应用高级操作之前的数据。

Document_ID	Document_Image
1	image_1.png
2	image_2.png
3	image_3.png

应用高级操作后的数据。

Document_ID	Document_Image	Document_Image_Text
1	image_1.png	邮寄的调查回复
2	image_2.png	邮寄的调查回复
3	image_3.png	邮寄的调查回复

以下是使用 Lambda 函数为空日期值插入当前日期时间的示例。这使用的条件是，如果日期字段值为“null”，则将其替换为当前日期时间。

示例 2：将 Last_Updated 字段中的空值替换为当前日期时间。

应用高级操作之前的数据。

Document_ID	Body_Text	Last_Updated
1	Lorem Ipsum。	2020 年 1 月 1 日

Document_ID	Body_Text	Last_Updated
2	Lorem Ipsum。	
3	Lorem Ipsum。	2020 年 7 月 1 日

应用高级操作后的数据。

Document_ID	Body_Text	Last_Updated
1	Lorem Ipsum。	2020 年 1 月 1 日
2	Lorem Ipsum。	2021 年 12 月 1 日
3	Lorem Ipsum。	2020 年 7 月 1 日

以下代码是配置 Lambda 函数以对原始原始数据进行高级数据操作的示例。

Console

配置 Lambda 函数，以便对原始原始数据进行高级数据操作

1. 在左侧导航窗格的索引下，选择文档富集，然后选择添加文档富集。
2. 在配置 Lambda 函数页面的用于预提取的 Lambda 部分，从下拉列表中选择您的 Lambda 函数 ARN 和您的 Amazon S3 存储桶。通过从下拉列表中选择创建新角色的选项来添加您的 IAM 访问角色。这将创建创建文档富集功能所需的 Amazon Kendra 权限。

CLI

配置 Lambda 函数，以便对原始原始数据进行高级数据操作

```
aws kendra create-data-source \
 --name data-source-name \
 --index-id index-id \
 --role-arn arn:aws:iam::account-id:role/role-name \
 --type S3 \
 --configuration '{"S3Configuration":{"BucketName":"S3-bucket-name"}}' \
```

```
 --custom-document-enrichment-configuration '{"PreExtractionHookConfiguration":
{"LambdaArn":"arn:aws:iam::account-id:function/function-name", "S3Bucket":"S3-
bucket-name"}, "RoleArn": "arn:aws:iam:account-id:role/cde-role-name"}'
```

Python

配置 Lambda 函数，以便对原始原始数据进行高级数据操作

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Create a data source with customizations.")

# Provide the name of the data source
name = "data-source-name"
# Provide the index ID for the data source
index_id = "index-id"
# Provide the IAM role ARN required for data sources
role_arn = "arn:aws:iam::${account-id}:role/${role-name}"
# Provide the data source connection information
data_source_type = "S3"
S3_bucket_name = "S3-bucket-name"
# Configure the data source with Custom Document Enrichment
configuration = {"S3Configuration":
        {
            "BucketName": S3_bucket_name
        }
    }
custom_document_enrichment_configuration = {"PreExtractionHookConfiguration":
        {
            "LambdaArn":"arn:aws:iam::account-id:function/function-name",
            "S3Bucket":"S3-bucket-name"
        }
    "RoleArn":"arn:aws:iam::account-id:role/cde-role-name"
    }

try:
    data_source_response = kendra.create_data_source(
        Name = name,
        IndexId = index_id,
```

```
        RoleArn = role_arn,
        Type = data_source_type
        Configuration = configuration
        CustomDocumentEnrichmentConfiguration =
custom_document_enrichment_configuration
    )

    pprint.pprint(data_source_response)

    data_source_id = data_source_response["Id"]

    print("Wait for Amazon Kendra to create the data source with your
customizations.")

    while True:
        # Get the details of the data source, such as the status
        data_source_description = kendra.describe_data_source(
            Id = data_source_id,
            IndexId = index_id
        )
        status = data_source_description["Status"]
        print(" Creating data source. Status: "+status)
        time.sleep(60)
        if status != "CREATING":
            break

    print("Synchronize the data source.")

    sync_response = kendra.start_data_source_sync_job(
        Id = data_source_id,
        IndexId = index_id
    )

    pprint.pprint(sync_response)

    print("Wait for the data source to sync with the index.")

    while True:

        jobs = kendra.list_data_source_sync_jobs(
            Id = data_source_id,
            IndexId = index_id
        )
```

```
        # For this example, there should be one job
        status = jobs["History"][0]["Status"]

        print(" Syncing data source. Status: "+status)
        time.sleep(60)
        if status != "SYNCING":
            break

except  ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

配置 Lambda 函数，以便对原始原始数据进行高级数据操作

```
package com.amazonaws.kendra;

import java.util.concurrent.TimeUnit;
import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.CreateDataSourceRequest;
import software.amazon.awssdk.services.kendra.model.CreateDataSourceResponse;
import software.amazon.awssdk.services.kendra.model.CreateIndexRequest;
import software.amazon.awssdk.services.kendra.model.CreateIndexResponse;
import software.amazon.awssdk.services.kendra.model.DataSourceConfiguration;
import software.amazon.awssdk.services.kendra.model.DataSourceStatus;
import software.amazon.awssdk.services.kendra.model.DataSourceSyncJob;
import software.amazon.awssdk.services.kendra.model.DataSourceSyncJobStatus;
import software.amazon.awssdk.services.kendra.model.DataSourceType;
import software.amazon.awssdk.services.kendra.model.DescribeDataSourceRequest;
import software.amazon.awssdk.services.kendra.model.DescribeDataSourceResponse;
import software.amazon.awssdk.services.kendra.model.DescribeIndexRequest;
import software.amazon.awssdk.services.kendra.model.DescribeIndexResponse;
import software.amazon.awssdk.services.kendra.model.IndexStatus;
import software.amazon.awssdk.services.kendra.model.ListDataSourceSyncJobsRequest;
import software.amazon.awssdk.services.kendra.model.ListDataSourceSyncJobsResponse;
import software.amazon.awssdk.services.kendra.model.S3DataSourceConfiguration;
import software.amazon.awssdk.services.kendra.model.StartDataSourceSyncJobRequest;
import software.amazon.awssdk.services.kendra.model.StartDataSourceSyncJobResponse;


public class CreateDataSourceWithCustomizationsExample {
```

```
    public static void main(String[] args) throws InterruptedException {
        System.out.println("Create a data source with customizations");

        String dataSourceName = "data-source-name";
        String indexId = "index-id";
        String dataSourceRoleArn = "arn:aws:iam::account-id:role/role-name";
        String s3BucketName = "S3-bucket-name"

        KendraClient kendra = KendraClient.builder().build();

        CreateDataSourceRequest createDataSourceRequest = CreateDataSourceRequest
            .builder()
            .name(dataSourceName)
            .description(experienceDescription)
            .roleArn(experienceRoleArn)
            .type(DataSourceType.S3)
            .configuration(
                DataSourceConfiguration
                    .builder()
                    .s3Configuration(
                        S3DataSourceConfiguration
                            .builder()
                            .bucketName(s3BucketName)
                            .build()
                    ).build()
            )
            .customDocumentEnrichmentConfiguration(
                CustomDocumentEnrichmentConfiguration
                    .builder()
                    .preExtractionHookConfiguration(
                        HookConfiguration
                            .builder()
                            .lambdaArn("arn:aws:iam::account-id:function/function-
name")
                            .s3Bucket("S3-bucket-name")
                            .build())
                    .roleArn("arn:aws:iam::account-id:role/cde-role-name")
                    .build();

        CreateDataSourceResponse createDataSourceResponse =
 kendra.createDataSource(createDataSourceRequest);
        System.out.println(String.format("Response of creating data source: %s",
 createDataSourceResponse));
```

```
        String dataSourceId = createDataSourceResponse.id();
        System.out.println(String.format("Waiting for Kendra to create the data
source %s", dataSourceId));
        DescribeDataSourceRequest describeDataSourceRequest =
DescribeDataSourceRequest
            .builder()
            .indexId(indexId)
            .id(dataSourceId)
            .build();

        while (true) {
            DescribeDataSourceResponse describeDataSourceResponse =
kendra.describeDataSource(describeDataSourceRequest);

            DataSourceStatus status = describeDataSourceResponse.status();
            System.out.println(String.format("Creating data source. Status: %s",
status));
            TimeUnit.SECONDS.sleep(60);
            if (status != DataSourceStatus.CREATING) {
                break;
            }
        }

        System.out.println(String.format("Synchronize the data source %s",
dataSourceId));
        StartDataSourceSyncJobRequest startDataSourceSyncJobRequest =
StartDataSourceSyncJobRequest
            .builder()
            .indexId(indexId)
            .id(dataSourceId)
            .build();
        StartDataSourceSyncJobResponse startDataSourceSyncJobResponse =
kendra.startDataSourceSyncJob(startDataSourceSyncJobRequest);
        System.out.println(String.format("Waiting for the data
source to sync with the index %s for execution ID %s", indexId,
startDataSourceSyncJobResponse.executionId()));

        // For this example, there should be one job
        ListDataSourceSyncJobsRequest listDataSourceSyncJobsRequest =
ListDataSourceSyncJobsRequest
            .builder()
            .indexId(indexId)
            .id(dataSourceId)
            .build();
```

```
        while (true) {
            ListDataSourceSyncJobsResponse listDataSourceSyncJobsResponse =
 kendra.listDataSourceSyncJobs(listDataSourceSyncJobsRequest);
            DataSourceSyncJob job = listDataSourceSyncJobsResponse.history().get(0);
            System.out.println(String.format("Syncing data source. Status: %s",
 job.status()));

            TimeUnit.SECONDS.sleep(60);
            if (job.status() != DataSourceSyncJobStatus.SYNCING) {
                break;
            }

        }

        System.out.println("Data source creation with customizations is complete");
    }
}
```

Lambda 函数的数据合约

用于高级数据操作的 Lambda 函数与 Amazon Kendra 数据合约交互。合约是您的 Lambda 函数的必备请求和响应结构。如果您的 Lambda 函数不遵循这些结构，则 Amazon Kendra 会引发错误。

适用于 `PreExtractionHookConfiguration` 的 Lambda 函数应采用以下请求结构：

```
{
    "version": <str>,
    "dataBlobStringEncodedInBase64": <str>, //In the case of a data blob
    "s3Bucket": <str>, //In the case of an S3 bucket
    "s3ObjectKey": <str>, //In the case of an S3 bucket
    "metadata": <Metadata>
}
```

包括 `CustomDocumentAttribute` 结构在内的 `metadata` 结构如下所示：

```
{
    "attributes": [<CustomDocumentAttribute<]
}

CustomDocumentAttribute
```

```
{
    "name": <str>,
    "value": <CustomDocumentAttributeValue>
}

CustomDocumentAttributeValue
{
    "stringValue": <str>,
    "integerValue": <int>,
    "longValue": <long>,
    "stringListValue": list<str>,
    "dateValue": <str>
}
```

PreExtractionHookConfiguration 的 Lambda 函数必须符合以下响应结构：

```
{
    "version": <str>,
    "dataBlobStringEncodedInBase64": <str>, //In the case of a data blob
    "s3ObjectKey": <str>, //In the case of an S3 bucket
    "metadataUpdates": [<CustomDocumentAttribute>]
}
```

适用于 PostExtractionHookConfiguration 的 Lambda 函数应采用以下请求结构：

```
{
    "version": <str>,
    "s3Bucket": <str>,
    "s3ObjectKey": <str>,
    "metadata": <Metadata>
}
```

PostExtractionHookConfiguration 的 Lambda 函数必须符合以下响应结构：

```
PostExtractionHookConfiguration Lambda Response
{
    "version": <str>,
    "s3ObjectKey": <str>,
    "metadataUpdates": [<CustomDocumentAttribute>]
}
```

您修改过的文档已上传到您的 Amazon S3 存储桶。修改后的文档必须遵循 the section called “结构化的文档格式” 中显示的格式。

结构化的文档格式

Amazon Kendra 将您的结构化文档上传到给定的 Amazon S3 存储桶。结构化文档遵循以下格式：

```
Kendra document

{
   "textContent": <TextContent>
}

TextContent
{
  "documentBodyText": <str>
}
```

遵守数据合同的 Lambda 函数示例

以下 Python 代码是 Lambda 函数的示例，该函数对元数据字段 _authors、_document_title 以及原始文档或原始文档的正文内容进行高级操作。

如果正文内容存放在 Amazon S3 存储桶中

```
import json
import boto3

s3 = boto3.client("s3")

# Lambda function for advanced data manipulation
def lambda_handler(event, context):
    # Get the value of "S3Bucket" key name or item from the given event input
    s3_bucket = event.get("s3Bucket")
    # Get the value of "S3ObjectKey" key name or item from the given event input
    s3_object_key = event.get("s3ObjectKey")

    content_object_before_CDE = s3.get_object(Bucket = s3_bucket, Key = s3_object_key)
    content_before_CDE = content_object_before_CDE["Body"].read().decode("utf-8");
    content_after_CDE = "CDEInvolved " + content_before_CDE

    # Get the value of "metadata" key name or item from the given event input
```

```
    metadata = event.get("metadata")
    # Get the document "attributes" from the metadata
    document_attributes = metadata.get("attributes")

    s3.put_object(Bucket = s3_bucket, Key = "dummy_updated_kendra_document",
 Body=json.dumps(content_after_CDE))
    return {
        "version": "v0",
        "s3ObjectKey": "dummy_updated_kendra_document",
        "metadataUpdates": [
            {"name":"_document_title", "value":
{"stringValue":"title_from_pre_extraction_lambda"}},
            {"name":"_authors", "value":{"stringListValue":["author1", "author2"]}}
        ]
    }
```

如果正文内容驻留在数据块中

```
import json
import boto3
import base64

# Lambda function for advanced data manipulation
def lambda_handler(event, context):

    # Get the value of "dataBlobStringEncodedInBase64" key name or item from the given
 event input
    data_blob_string_encoded_in_base64 = event.get("dataBlobStringEncodedInBase64")
    # Decode the data blob string in UTF-8
    data_blob_string =
 base64.b64decode(data_blob_string_encoded_in_base64).decode("utf-8")
    # Get the value of "metadata" key name or item from the given event input
    metadata = event.get("metadata")
    # Get the document "attributes" from the metadata
    document_attributes = metadata.get("attributes")

    new_data_blob = "This should be the modified data in the document by pre processing
 lambda ".encode("utf-8")
    return {
        "version": "v0",
        "dataBlobStringEncodedInBase64":
 base64.b64encode(new_data_blob).decode("utf-8"),
        "metadataUpdates": [
```

```
            {"name":"_document_title", "value":
{"stringValue":"title_from_pre_extraction_lambda"}},
            {"name":"_authors", "value":{"stringListValue":["author1", "author2"]}}
        ]
    }
```

以下 Python 代码是 Lambda 函数的示例，该函数应用了对元数据字段 _authors、_document_title 的高级操作，以及结构化或已解析文档的正文内容。

```
import json
import boto3
import time

s3 = boto3.client("s3")

# Lambda function for advanced data manipulation
def lambda_handler(event, context):

    # Get the value of "S3Bucket" key name or item from the given event input
    s3_bucket = event.get("s3Bucket")
    # Get the value of "S3ObjectKey" key name or item from the given event input
    s3_key = event.get("s3ObjectKey")
    # Get the value of "metadata" key name or item from the given event input
    metadata = event.get("metadata")
    # Get the document "attributes" from the metadata
    document_attributes = metadata.get("attributes")

    kendra_document_object = s3.get_object(Bucket = s3_bucket, Key = s3_key)
    kendra_document_string = kendra_document_object['Body'].read().decode('utf-8')
    kendra_document = json.loads(kendra_document_string)
    kendra_document["textContent"]["documentBodyText"] = "Changing document body to a
 short sentence."

    s3.put_object(Bucket = s3_bucket, Key = "dummy_updated_kendra_document",
 Body=json.dumps(kendra_document))

    return {
        "version" : "v0",
        "s3ObjectKey": "dummy_updated_kendra_document",
        "metadataUpdates": [
            {"name": "_document_title", "value":{"stringValue":
 "title_from_post_extraction_lambda"}},
            {"name": "_authors", "value":{"stringListValue":["author1", "author2"]}}
```

```
        ]
    }
```

搜索索引

要搜索 Amazon Kendra 索引，您可以使用查询 API。Query API 会返回有关您在应用程序中使用的已编入索引的文档的信息。本节介绍如何进行查询、执行筛选以及如何解释从 Query API 获得的响应。

要搜索已编入索引的 Amazon Kendra 文档 Amazon Lex，请使用亚马逊。 KendraSearchIntent。有关使用进行配置的示例 Amazon Kendra Amazon Lex，请参阅为 Amazon Kendra 索引创建常见问题解答机器人。

主题

查询索引

搜索索引时， Amazon Kendra 使用您提供的有关文档的所有信息来确定与输入的搜索词最相关的文档。需要 Amazon Kendra 考虑的一些项目是：

- 文档的文本或正文。
- 文档标题。
- 您已标记为可搜索的自定义文本字段。
- 应使用您指定的日期字段来确定文档的“新鲜度”。
- 可以提供相关信息的任何其他字段。

Amazon Kendra 还可以根据您可能为搜索设置的任何字段/属性过滤器来筛选响应。例如，如果您有一个名为“部门”的自定义字段，则可以筛选响应以仅返回名为“法律”的部门的文档。有关更多信息，请参阅自定义字段和属性。

返回的搜索结果按每个文档所确 Amazon Kendra 定的相关性进行排序。结果是分页的，这样您就可以一次向用户显示一个页面。

要搜索已编入索引的 Amazon Kendra 文档 Amazon Lex，请使用亚马逊。 KendraSearchIntent。有关使用进行配置的示例 Amazon Kendra Amazon Lex，请参阅为 Amazon Kendra 索引创建常见问题解答机器人。

以下示例说明如何搜索索引。 Amazon Kendra 确定最适合查询的搜索结果类型（答案、文档、问题答案）。您无法配置 Amazon Kendra 为向查询返回特定类型的搜索响应（答案、文档、问题答案）。

有关查询响应的信息，请参阅查询响应和响应类型。

先决条件

在使用查询 API 查询索引之前：

- 为索引设置所需的权限并连接到您的数据来源或批量上传文档。有关更多信息，请参阅 IAM 角色。在调用 API 创建索引和数据来源连接器或批量上传文档时，您可以使用角色的 Amazon 资源名称。
- 设置 AWS Command Line Interface、SDK 或前往 Amazon Kendra 控制台。有关更多信息，请参阅设置 Amazon Kendra。
- 创建索引并连接到文档的数据来源或批量上传文档。有关更多信息，请参阅创建索引和创建数据来源连接器。

搜索索引（控制台）

您可以使用 Amazon Kendra 控制台搜索和测试您的索引。您可以进行查询并查看结果。

使用控制台搜索索引

1. 登录 AWS Management Console 并打开 Amazon Kendra 控制台，网址为 http://console.aws.amazon.com/kendra/。
2. 在导航窗格上，选择索引。
3. 选择您的索引。
4. 在导航菜单中，选择搜索索引的选项。

5. 在文本框中输入查询，然后按 Enter。
6. Amazon Kendra 返回搜索结果。

您还可以通过选择侧面板中的灯泡图标来获取搜索的查询 ID。

搜索索引（SDK）

使用 Python 或 Java 搜索索引

- 以下示例搜索索引。将 `query` 的值更改为搜索查询，将 `index_id` 或 `indexId` 值更改为要搜索的索引的索引标识符。

  您还可以在调用 Query API 时获取作为响应元素一部分的搜索查询 ID。

  Python

```
import boto3
import pprint

kendra = boto3.client("kendra")

# Provide the index ID
index_id = "index-id"
# Provide the query text
query = "query text"

response = kendra.query(
        QueryText = query,
        IndexId = index_id)

print("\nSearch results for query: " + query + "\n")

for query_result in response["ResultItems"]:

    print("--------------------")
    print("Type: " + str(query_result["Type"]))

    if query_result["Type"]=="ANSWER" or
 query_result["Type"]=="QUESTION_ANSWER":
        answer_text = query_result["DocumentExcerpt"]["Text"]
        print(answer_text)
```

```
    if query_result["Type"]=="DOCUMENT":
        if "DocumentTitle" in query_result:
            document_title = query_result["DocumentTitle"]["Text"]
            print("Title: " + document_title)
        document_text = query_result["DocumentExcerpt"]["Text"]
        print(document_text)

    print("------------------\n\n")
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.QueryRequest;
import software.amazon.awssdk.services.kendra.model.QueryResponse;
import software.amazon.awssdk.services.kendra.model.QueryResultItem;

public class SearchIndexExample {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();

        String query = "query text";
        String indexId = "index-id";

        QueryRequest queryRequest = QueryRequest
            .builder()
            .queryText(query)
            .indexId(indexId)
            .build();

        QueryResponse queryResponse = kendra.query(queryRequest);

        System.out.println(String.format("\nSearch results for query: %s",
 query));
        for(QueryResultItem item: queryResponse.resultItems()) {
            System.out.println("----------------------");
            System.out.println(String.format("Type: %s", item.type()));

            switch(item.type()) {
                case QUESTION_ANSWER:
                case ANSWER:
```

```
                    String answerText = item.documentExcerpt().text();
                    System.out.println(answerText);
                    break;
                case DOCUMENT:
                    String documentTitle = item.documentTitle().text();
                    System.out.println(String.format("Title: %s",
 documentTitle));
                    String documentExcerpt = item.documentExcerpt().text();
                    System.out.println(String.format("Excerpt: %s",
 documentExcerpt));
                    break;
                default:
                    System.out.println(String.format("Unknown query result type:
 %s", item.type()));

            }

            System.out.println("----------------------\n");
        }
    }
}
```

搜索索引（Postman）

你可以使用 Postman 来查询和测试你的 Amazon Kendra 索引。

使用 Postman 搜索索引

1. 在 Postman 中创建一个新集合，并将请求类型设置为 POST。
2. 输入端点 URL。例如，https://kendra.<region>.amazonaws.com。
3. 选择身份验证选项卡并输入以下信息。
   - 类型 - 选择 AWS 签名。
   - AccessKey— 输入创建 IAM 用户时生成的访问密钥。
   - SecretKey— 输入创建 IAM 用户时生成的密钥。
   - AWS 区域-输入索引区域。例如，us-west-2。
   - 服务名称 - 输入 kendra。这区分大小写，因此必须是小写字母。

> **Warning**
>
> 如果您输入的服务名称不正确或未使用小写字母，则在选择发送请求后会引发错误：“凭证的范围应限定为正确的服务‘kendra’。”
> 您还必须检查您输入的访问密钥和密钥钥是否正确。

4. 选择标头选项卡，然后输入以下键和值信息。

   - 键：X-Amz-Target

     价值：com.amazonaws.kendra。 AWSKendraFrontendService.Query

   - 键：Content-Encoding

     值：amz-1.0

5. 选择正文选项卡并执行以下操作。

   - 为请求正文选择原始 JSON 类型。
   - 输入包含您的索引 ID 和查询文本的 JSON。

```
{
    "IndexId": "index-id",
    "QueryText": "enter a query here"
}
```

> **Warning**
>
> 如果你的 JSON 没有使用正确的缩进，则会抛出一个错误：“” SerializationException。
> 检查您的 JSON 中的缩进。

6. 选择发送（靠近右上角）。

使用高级查询语法进行搜索

您可以使用高级查询语法或运算符创建比简单的关键字或自然语言查询更具体的查询。这包括范围、布尔值、通配符等。通过使用运算符，您可以为查询提供更多背景信息，并进一步优化搜索结果。

Amazon Kendra 支持以下运算符。

- 布尔值：限制或扩大搜索范围的逻辑。例如，将搜索 amazon AND sports 限制为仅搜索包含两个术语的文档。
- 圆括号：按优先顺序读取嵌套的查询词。例如，(amazon AND sports) NOT rainforest读取 NOT rainforest 之前的 (amazon AND sports)。
- 范围：日期或数字范围值。范围可以是包容性的、排他的，也可以是无界的。例如，您可以搜索上次更新于 2020 年 1 月 1 日至 2020 年 12 月 31 日之间的文档，包括这些日期。
- 字段：使用特定字段限制搜索范围。例如，您可以搜索在“位置”字段中带有“美国”字样的文档。
- 通配符：部分匹配一串文本。例如，Cloud*可以匹配 CloudFormation。 Amazon Kendra 目前仅支持尾随通配符。
- 精确引号：精确匹配一串文本。例如，包含 "Amazon Kendra" "pricing" 的文档。

您可以使用上述任何运算符的组合。

请注意，过度使用运算符或查询高度复杂可能会影响查询延迟。就延迟而言，通配符是最昂贵的运算符之一。一般规则是，使用的术语和运算符越多，对延迟的潜在影响就越大。影响延迟的其他因素包括已编入索引的文档的平均大小、索引的大小、对搜索结果的任何筛选以及 Amazon Kendra 索引的总体负载。

布尔值

您可以使用布尔运算符 AND、OR、NOT 来组合或排除单词。

以下是使用布尔运算符的示例。

amazon AND sports

返回文本中同时包含术语“amazon”和“sports”的搜索结果，例如 Amazon Prime 视频体育或其他类似内容。

sports OR recreation

返回文本中包含“sports”或“recreation”或两者兼而有之的搜索结果。

amazon NOT rainforest

返回文本中包含“amazon”一词但不包含“rainforest”一词的搜索结果。这是为了搜索有关 Amazon 公司的文档，而不是 Amazon 雨林的文档。

圆括号

您可以使用圆括号按优先顺序查询嵌套单词。圆括号表示应 Amazon Kendra 如何读取查询。

以下是使用圆括号运算符的示例。

`(amazon AND sports) NOT rainforest`

返回文本中同时包含“amazon”和“sports”，但不包含“rainforest”一词的文档。这是为了搜索 Amazon Prime 视频体育或其他类似内容，而不是 Amazon 雨林中的冒险运动。括号有助于表明 `amazon AND sports` 应该在 `NOT rainforest` 之前读取。不应将该查询理解为 `amazon AND (sports NOT rainforest)`。

`(amazon AND (sports OR recreation)) NOT rainforest`

返回包含“sports”或“recreation”（或两者兼而有之）以及“amazon”一词的文档。但它不包括“rainforest”一词。这是为了搜索 Amazon Prime 视频体育或娱乐，而不是 Amazon 雨林中的冒险运动。括号有助于指明 `sports OR recreation` 应在与“amazon”合并之前阅读，即在 `NOT rainforest` 之前阅读。不应将该查询理解为 `amazon AND (sports OR (recreation NOT rainforest))`。

范围

您可以使用一系列值来筛选搜索结果。您必须为指定属性和范围值。这可以是日期或数字类型。

日期范围必须为以下格式：

- Epoch
- YYYY
- YYYY-mm
- YYYY-mm-dd
- YYYY-mm-dd'T'HH

您还可以指定是包括还是排除范围的较低值和较高值。

以下是使用范围运算符的示例。

`_processed_date:>2019-12-31 AND _processed_date:<2021-01-01`

退回在 2020 年处理的文件，大于 2019 年 12 月 31 日且少于 2021 年 1 月 1 日。

_processed_date:>=2020-01-01 AND _processed_date:<=2020-12-31

退回在 2020 年处理的文件，大于或等于 2020 年 1 月 1 日且小于或等于 2020 年 12 月 31 日。

_document_likes:<1

返回点赞次数为零或没有用户反馈的文档，点赞次数少于 1。

您可以指定应将范围视为包含给定范围值还是不包含给定范围值。

包含

_last_updated_at:[2020-01-01 TO 2020-12-31]

返回文档最后更新时间为 2020 年，包括 2020 年 12 月 1 日和 2020 年 12 月 31 日。

排除

_last_updated_at:{2019-12-31 TO 2021-01-01}

返回文档最后更新时间为 2020 年，不包括 2019 年 12 月 31 日和 2021 年 1 月 1 日。

对于既不包含也不排除的无界定范围，只需使用 < and > 运算符即可。例如，`_last_updated_at:>2019-12-31 AND _last_updated_at:<2021-01-01`

字段

您可以将搜索限制为仅返回符合特定字段值的文档。该字段可以是任意类型。

以下是使用字段级上下文运算符的示例。

status:"Incomplete" AND financial_year:2021

返回状态为“未完成”的 2021 财务年度文件。

(sports OR recreation) AND country:"United States" AND level:"professional"

返回讨论美国职业体育或娱乐活动的文档。

通配符

您可以使用通配符运算符扩大搜索范围，以考虑单词和短语的变体。这在搜索名称变体时很有用。Amazon Kendra 目前仅支持尾随通配符。尾随通配符的前缀字符数必须大于二。

以下是使用通配符运算符的示例。

Cloud*

返回包含 CloudFormation 和等变体的文档 CloudWatch。

kendra*aws

返回包含 kendra.amazonaws 等变体的文档。

kendra*aws*

返回包含 kendra.amazonaws.com 等变体的文档

确切限额

您可以使用引号来搜索一段文本的精确匹配项。

以下是使用引号的示例。

"Amazon Kendra" "pricing"

返回同时包含短语“Amazon Kendra”和“定价”一词的文档。文档必须同时包含“Amazon Kendra”和“定价”才能返回结果。

"Amazon Kendra" "pricing" cost

返回同时包含短语“Amazon Kendra”和“定价”一词以及可选的“成本”一词的文档。文档必须同时包含“Amazon Kendra”和“定价”才能返回结果，但不一定包含“成本”。

查询语法无效

Amazon Kendra 如果您的查询语法存在问题或当前不支持您的查询，则会发出警告 Amazon Kendra。有关更多信息，请参阅有关查询警告的 API 文档。

以下查询是无效查询语法的示例。

_last_updated_at:<2021-12-32

日期无效。公历没有 32 日，只有 Amazon Kendra才会使用 32 日。

_view_count:ten

数值无效。必须使用数字来表示数值。

nonExistentField:123

字段搜索无效。该字段必须存在才能使用字段搜索。

Product:[A TO D]

范围无效。范围必须使用数值或日期。

`OR Hello`

布尔值无效。运算符必须与术语一起使用并置于术语之间。

搜索语言

您可以使用支持的语言搜索文档。您可以在中传递语言代码，AttributeFilter返回所选语言的已筛选文档。您可以使用支持的语言输入查询。

如果未指定语言，则默认 Amazon Kendra 查询英文文档。有关支持的语言（包括其代码）的更多信息，请参阅添加非英语语言文档。

要在控制台中以支持的语言搜索文档，请选择您的索引，然后从导航菜单中选择搜索索引的选项。选择搜索设置，然后从语言下拉列表中选择一种语言，即可选择要返回文档的语言。

以下示例显示了如何搜索西班牙语文档。

在控制台中用西班牙语搜索索引

1. 登录 AWS Management Console 并打开 Amazon Kendra 控制台，网址为 http://console.aws.amazon.com/kendra/。
2. 在导航菜单中选择索引，然后选择索引。
3. 在导航菜单中，选择搜索索引的选项。
4. 在搜索设置中选择语言下拉列表并选择西班牙语。
5. 在文本框中输入查询，然后按 Enter。
6. Amazon Kendra 以西班牙语返回搜索结果。

使用 CLI、Python 或 Java 搜索西班牙语索引

- 以下示例使用西班牙语搜索索引。将值 `searchString` 更改为搜索查询，并将值 `indexID` 更改为要搜索的索引的标识符。西班牙语的语言代码是 `es`。您可以将其替换为自己的语言代码。

  CLI

```
{
  "EqualsTo":{
    "Key": "_language_code",
    "Value": {
```

```
    "StringValue": "es"
    }
  }
}
```

Python

```
import boto3
import pprint

kendra = boto3.client("kendra")

# Provide the index ID
index_id = "index-id"
# Provide the query text
query = "search-string"

# Includes the index ID, query text, and language attribute filter
response = kendra.query(
        QueryText = query,
        IndexId = index_id,
        AttributeFilter = {
            "EqualsTo": {
                "Key": "_language_code",
                "Value": {
                    "StringValue": "es"
                    }
                }
            })

print ("\nSearch results|Resultados de la búsqueda: " + query + "\n")

for query_result in response["ResultItems"]:

    print("--------------------")
    print("Type: " + str(query_result["Type"]))

    if query_result["Type"]=="ANSWER" or
 query_result["Type"]=="QUESTION_ANSWER":
        answer_text = query_result["DocumentExcerpt"]["Text"]
        print(answer_text)

    if query_result["Type"]=="DOCUMENT":
```

```
        if "DocumentTitle" in query_result:
            document_title = query_result["DocumentTitle"]["Text"]
            print("Title: " + document_title)
        document_text = query_result["DocumentExcerpt"]["Text"]
        print(document_text)

    print("------------------\n\n")
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.QueryRequest;
import software.amazon.awssdk.services.kendra.model.QueryResponse;
import software.amazon.awssdk.services.kendra.model.QueryResultItem;

public class SearchIndexExample {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();

        String query = "searchString";
        String indexId = "indexID";

        QueryRequest queryRequest = QueryRequest.builder()
            .queryText(query)
            .indexId(indexId)
            .attributeFilter(
                 AttributeFilter.builder()
                     .withEqualsTo(
                         DocumentAttribute.builder()
                             .withKey("_language_code")
                             .withValue("es")
                             .build())
                     .build())
            .build();

        QueryResponse queryResponse = kendra.query(queryRequest);

        System.out.println(String.format("\nSearch results|
```

```
                                            Resultados de la búsqueda: %s",
 query));
        for(QueryResultItem item: queryResponse.resultItems()) {
            System.out.println("----------------------");
            System.out.println(String.format("Type: %s", item.type()));

            switch(item.type()) {
                case QUESTION_ANSWER:
                case ANSWER:
                    String answerText = item.documentExcerpt().text();
                    System.out.println(answerText);
                    break;
                case DOCUMENT:
                    String documentTitle = item.documentTitle().text();
                    System.out.println(String.format("Title: %s",
 documentTitle));
                    String documentExcerpt = item.documentExcerpt().text();
                    System.out.println(String.format("Excerpt: %s",
 documentExcerpt));
                    break;
                default:
                    System.out.println(String.format("Unknown query result type:
 %s", item.type()));

            }

            System.out.println("----------------------\n");
        }
    }
}
```

检索段落

您可以将 Retrieve API 用作检索增强生成（RAG）系统的检索器。

RAG 系统使用生成式人工智能来构建问答应用程序。RAG 系统由检索器和大型语言模型（LLM）组成。给定一个查询，检索器会从文档语料库中识别出最相关的文本块，并将其提供给 LLM 以提供最有用的答案。然后，LLM 会分析相关的文本块或段落，并为查询生成全面的响应。

RetrieveAPI 会查看被称为段落的大块文本或摘录，并返回与查询最相关的热门段落。

与 Query API 一样，Retrieve API 也使用语义搜索来搜索相关信息。语义搜索会考虑搜索查询的上下文，以及索引文档中的所有可用信息。但是，默认情况下，Query API 仅返回最多 100 个标记词的摘录段落。使用 Retrieve API，您可以检索最多 200 个标记词和多达 100 个语义相关段落的较长段落。这不包括索引中的问题答案或常见问题解答类型的回复。这些段落是文本摘录，可以在语义上从多个文档和同一文档的多个部分中提取出来。如果在极端情况下，您的文档使用 Retrieve API 生成零段落，则可以选择使用 Query API 及其响应类型。

您可以使用 Retrieve API 执行以下操作：

- 覆盖指数级别的提升
- 根据文档字段或属性进行筛选
- 根据用户或其群组对文档的访问权限进行筛选
- 查看置信度分数区以获取检索到的通过结果。 Amazon Kendra 的置信度分区提供相对排名，表示响应与查询相关的信心程度。

> **Note**
>
> 置信度分数桶目前仅适用于英语。

您还可以在响应中加入某些字段，这些字段可能会提供有用的其他信息。

Retrieve API 目前不支持该 Query API 支持的所有功能。不支持以下功能：使用高级查询语法进行查询，使用建议的拼写更正进行查询，使用分面进行查询，自动完成搜索查询的查询建议以及增量学习。请注意，并非所有功能都适用于 Retrieve API。Retrieve该 API 的任何未来版本都将记录在本指南中。

Retrieve API 共享您为索引设置的查询容量单位数。有关单个容量单位中包含的内容以及索引的默认基本容量的更多信息，请参阅调整容量。

> **Note**
>
> 如果您使用的是 Amazon Kendra 开发者版，则无法添加容量；只能在使用 Amazon Kendra 企业版时添加容量。有关开发人员版和企业版中包含的内容的更多信息，请参阅 Amazon Kendra 版本。

以下是使用 Retrieve API 从 "how does amazon kendra work?" 查询索引中的文档中检索前 100 条最相关的段落的示例

Python

```
import boto3
import pprint

kendra = boto3.client("kendra")

# Provide the index ID
index_id = "index-id"
# Provide the query text
query = "how does amazon kendra work?"
# You can retrieve up to 100 relevant passages
# You can paginate 100 passages across 10 pages, for example
page_size = 10
page_number = 10

result = kendra.retrieve(
        IndexId = index_id,
        QueryText = query,
        PageSize = page_size,
        PageNumber = page_number)

print("\nRetrieved passage results for query: " + query + "\n")

for retrieve_result in result["ResultItems"]:

    print("-------------------")
    print("Title: " + str(retrieve_result["DocumentTitle"]))
    print("URI: " + str(retrieve_result["DocumentURI"]))
    print("Passage content: " + str(retrieve_result["Content"]))
    print("------------------\n\n")
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.RetrieveRequest;
import software.amazon.awssdk.services.kendra.model.RetrieveResult;
import software.amazon.awssdk.services.kendra.model.RetrieveResultItem;
```

```
public class RetrievePassageExample {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();

        String indxId = "index-id";
        String query = "how does amazon kendra work?";
        Integer pgSize = 10;
        Integer pgNumber = 10;

        RetrieveRequest retrieveRequest = retrieveRequest
            .builder()
            .indexId(indxId)
            .queryText(query)
            .pageSize(pgSize)
            .pageNumber(pgNumber)
            .build();

        RetrieveResult retrieveResult = kendra.retrieve(retrieveRequest);

        System.out.println(String.format("\nRetrieved passage results for query:
 %s", query));
        for(RetrieveResultItem item: retrieveResult.resultItems()) {
            System.out.println("----------------------");
            System.out.println(String.format("Title: %s", documentTitle));
            System.out.println(String.format("URI: %s", documentURI));
            System.out.println(String.format("Passage content: %s", content));
            System.out.println("----------------------\n");
        }
    }
}
```

浏览索引

您无需键入搜索查询即可按其属性或分面浏览文档。 Amazon Kendra 浏览索引可以帮助您的用户通过自由浏览索引来发现文档，而无需考虑特定的查询。这还可以帮助您的用户广泛浏览索引，将其作为搜索的起点。

浏览索引只能用于按文档属性或具有排序类型的分面进行搜索。不能使用“浏览索引”搜索整个索引。如果缺少查询文本，则 Amazon Kendra 要求提供文档属性筛选器或分面以及排序类型。

要允许使用查询 API 浏览索引，必须包含AttributeFilter或 Facet 和。SortingConfiguration要允许在控制台中浏览索引，请在导航菜单的索引下选择您的索引，然后选择搜索索引的选项。在搜索框中，按两次 Enter 键。选择下拉列表筛选搜索结果以选择筛选条件，然后选择下拉列表排序以选择排序类型。

以下是按文档创建日期降序浏览西班牙语文档索引的示例。

CLI

```
aws kendra query \
--index-id "index-id" \
--attribute-filter '{
    "EqualsTo":{
      "Key": "_language_code",
      "Value": {
        "StringValue": "es"
      }
    }
  }' \
--sorting-configuration '{
    "DocumentAttributeKey": "_created_at",
    "SortOrder": "DESC"
  }'
```

Python

```
import boto3

kendra = boto3.client("kendra")

# Must include the index ID, the attribute filter, and sorting configuration
response = kendra.query(
        IndexId = "index-id",
        AttributeFilter = {
            "EqualsTo": {
                "Key": "_language_code",
                "Value": {
                    "StringValue": "es"
                    }
                }
            },
        SortingConfiguration = {
            "DocumentAttributeKey": "_created_at",
            "SortOrder": "DESC"})
```

```
print("\nSearch results|Resultados de la búsqueda: \n")

for query_result in response["ResultItems"]:

    print("--------------------")
    print("Type: " + str(query_result["Type"]))

    if query_result["Type"]=="ANSWER" or query_result["Type"]=="QUESTION_ANSWER":
        answer_text = query_result["DocumentExcerpt"]["Text"]
        print(answer_text)

    if query_result["Type"]=="DOCUMENT":
        if "DocumentTitle" in query_result:
            document_title = query_result["DocumentTitle"]["Text"]
            print("Title: " + document_title)
        document_text = query_result["DocumentExcerpt"]["Text"]
        print(document_text)

    print("------------------\n\n")
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.QueryRequest;
import software.amazon.awssdk.services.kendra.model.QueryResult;
import software.amazon.awssdk.services.kendra.model.QueryResultItem;

public class SearchIndexExample {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();
        QueryRequest queryRequest = QueryRequest.builder()
            .withIndexId("index-id")
            .withAttributeFilter(AttributeFilter.builder()
                .withEqualsTo(DocumentAttribute.builder()
                    .withKey("_language_code")
                    .withValue(DocumentAttributeValue.builder()
                        .withStringValue("es")
                        .build())
                    .build())
                .build())
```

```
            .withSortingConfiguration(SortingConfiguration.builder()
                .withDocumentAttributeKey("_created_at")
                .withSortOrder("DESC")
                .build())
            .build());

        QueryResult queryResult = kendra.query(queryRequest);
        for (QueryResultItem item : queryResult.getResultItems()) {
            System.out.println("----------------------");
            System.out.println(String.format("Type: %s", item.getType()));

            switch (item.getType()) {
                case QueryResultType.QUESTION_ANSWER:
                case QueryResultType.ANSWER:
                    String answerText = item.getDocumentExcerpt().getText();
                    System.out.println(answerText);
                    break;
                case QueryResultType.DOCUMENT:
                    String documentTitle = item.getDocumentTitle().getText();
                    System.out.println(String.format("Title: %s", documentTitle));
                    String documentExcerpt = item.getDocumentExcerpt().getText();
                    System.out.println(String.format("Excerpt: %s",
 documentExcerpt));
                    break;
                default:
                    System.out.println(String.format("Unknown query result type:
 %s", item.getType()));
            }
            System.out.println("----------------------\n");
        }
    }
}
```

精选搜索结果

当您的用户发出某些查询时，您可以在搜索结果中显示某些文档。这有助于让用户更清楚地看到和突出显示结果。精选结果与通常的结果列表分开，并显示在搜索页面的顶部。您可以尝试为不同的查询提供不同的文档，或者确保某些文档获得应有的知名度。

您可以将特定查询映射到特定的文档，以便出现在结果中。如果查询包含完全匹配项，则搜索结果中会显示一个或多个特定文档。

例如，您可以指定，如果您的用户发布查询“2023 年新产品”，则选择标题为“最新动态”和“即将推出”的文档以显示在搜索结果页面的顶部。这有助于确保这些关于新产品的文档获得应有的知名度。

Amazon Kendra 如果搜索结果已被选中显示在搜索结果页面的顶部，则不会重复搜索结果。如果精选结果已经高于所有其他结果，则该结果不会再次被列为第一个结果。

为了显示某些结果，您必须使用查询中包含的关键字或短语来指定全文查询的精确匹配，而不是指定查询的部分匹配。例如，如果您仅在精选结果集中指定查询“Kendra”，则诸如“Kendra 在语义上如何对结果进行排名？”之类的查询 不会呈现精选结果。精选结果专为特定的查询而设计，而不是范围过于宽泛的查询。 Amazon Kendra 自然会处理关键字类型查询，以在搜索结果中对最有用的文档进行排名，从而避免根据简单关键字对结果进行过度精选。

如果您的用户经常使用某些查询，则可以为精选结果指定这些查询。例如，如果您使用 Amazon Kendra Analytics 查看热门查询，然后发现特定的查询，例如“kendra 在语义上如何对结果进行排名？” 和 “kendra 语义搜索” 经常被使用，那么这些查询对于指定标题为 “search 101” 的文档可能很有用。Amazon Kendra

Amazon Kendra 将对精选结果的查询视为不区分大小写。 Amazon Kendra 将查询转换为小写，并将尾随的空格字符替换为单个空格。 Amazon Kendra 匹配所有其他字符，就像您为精选结果指定查询时一样。

您可以使用 CreateFeaturedResultsSetAPI 创建一组精选结果，将其映射到某些查询。如果您使用控制台，则可以选择您的索引，然后在导航菜单中选择精选结果来创建精选结果集。每个索引最多可以创建 50 组精选结果，每组最多可以创建 4 个精选文档，每个精选结果集最多可以创建 49 个查询文本。您可以联系 AWS 支持部门以提高这些限制。

您可以在多组精选结果中选择同一个文档。但是，不得在多个集合中使用相同的完全匹配查询文本。对于每个索引的每个精选结果集，您为精选结果指定的查询必须是唯一的。

在最多选择四个精选文档时，您可以排列文档的顺序。如果您使用 API，则列出精选文档的顺序与精选结果中显示的顺序相同。如果您使用控制台，则在选择要在结果中显示的文档时，只需拖放文档顺序即可。

配置精选结果时，仍然可以进行访问控制，即某些用户和群组可以访问某些文档，而其他用户和群组则不能。对于用户上下文筛选也是如此。例如，用户 A 属于“实习生”公司群组，该群组不应访问有关公司机密的文档。如果用户 A 输入包含公司机密文档的查询，则用户 A 不会在搜索结果中看到该文档。搜索结果页面上的任何其他结果也是如此。您还可以使用标签来控制对精选结果集的访问权限，该结果集是您可以控制访问权限的 Amazon Kendra 资源。

以下是创建一组精选结果的示例，其中查询“2023 年新产品”、“新产品上市”映射到标题为“最新动态”（doc-id-1）和“即将推出”（doc-id-2）的文档。

CLI

```
aws kendra create-featured-results-set \
 --featured-results-set-name 'New product docs to feature' \
 --description "Featuring What's new and Coming soon docs" \
 --index-id index-id \
 --query-texts 'new products 2023' 'new products available' \
 --featured-documents '{"Id":"doc-id-1", "Id":"doc-id-2"}'
```

Python

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Create a featured results set.")

# Provide a name for the featured results set
featured_results_name = "New product docs to feature"
# Provide an optional decription for the featured results set
description = "Featuring What's new and Coming soon docs"
# Provide the index ID for the featured results set
index = "index-id"
# Provide a list of query texts for the featured results set
queries = ['new products 2023', 'new products available']
# Provide a list of document IDs for the featured results set
featured_doc_ids = [{"Id":"doc-id-1"}, {"Id":"doc-id-2"}]

try:
    featured_results_set_response = kendra.create_featured_results_set(
        FeaturedResultsSetName = featured_results_name,
        Decription = description,
        Index = index,
        QueryTexts = queries,
        FeaturedDocuments = featured_doc_ids
    )

    pprint.pprint(featured_results_set_response)

    featured_results_set_id = featured_results_set_response["FeaturedResultsSetId"]
```

```
    while True:
        # Get the details of the featured results set, such as the status
        featured_results_set_description = kendra.describe_featured_results_set(
            Id = featured_results_set_id
        )
        status = featured_results_set_description["Status"]
        print(" Featured results set status: "+status)

except  ClientError as e:
        print("%s" % e)

print("Program ends.")
```

HTML 表格搜索

Amazon Kendra的表格搜索功能可以从 HTML 文档中嵌入的表格中搜索和提取答案。搜索索引时，如果表中的摘录与查询相关，则 Amazon Kendra 包括该摘录并提供有用的信息。

Amazon Kendra 查看文档正文中的所有信息，包括表格中的有用信息。例如，索引包含业务报告，其中包含运营成本、收入和其他财务信息的表格。对于查询，“从2020-2022年起，每年的运营成本是多少？”， Amazon Kendra 可以返回表格的摘录，该表包含相关的表格列 “运营（百万美元）” 和 “财政年度”，以及包含 2020 年、2021 年和 2022 年收入值的表格行。结果中包含表格摘录、文档标题、指向完整文档的链接以及您选择包含的任何其他文档字段。

无论信息是在表格的一个单元格中还是在多个单元格中找到，都可以在搜索结果中显示表格摘录。例如， Amazon Kendra 可以显示针对以下每种查询量身定制的表格摘录：

- “2020 年利率最高的信用卡”
- “2020-2022 年以来利率最高的信用卡”
- “2020-2022 年利率排名前三的信用卡”
- “利率低于 10% 的信用卡”
- “所有可用的低息信用卡”

Amazon Kendra 突出显示与查询最相关的一个或多个表格单元格。搜索结果中会显示最相关的单元格及其对应的行、列和列名。表格摘录最多显示五列三行，具体取决于与查询相关的表格单元格数量以及原始表中有多少列可用。最相关的单元格与次要相关的单元格一起显示在表格摘录中。

响应包括置信度桶（MEDIUM,HIGH,VERY_HIGH），以显示表答案与查询的相关性。如果表格单元格的值是处于可信度的 VERY_HIGH，则该值将变为“最佳答案”并突出显示。对于置信度为 HIGH 的表格单元格值，则会突出显示这些值。对于可信的表格单元格值，则 MEDIUM 不会突出显示这些值。响应中会返回表答案的总体置信度。例如，如果表格主要包含有 HIGH 置信度的表格单元格，则表答案的响应中返回的总体置信度为 HIGH 置信度。

默认情况下，表格的重要性级别或权重不会高于文档的其他组成部分。在文档中，如果表与查询的相关性很小，但有一个高度相关的段落，则 Amazon Kendra 返回该段落的摘录。搜索结果显示同一文档或其他文档中提供最佳答案和最有用信息的内容。如果表格的置信度低于 MEDIUM 置信度，则响应中不会返回表格摘录。

要对现有索引使用表格搜索，必须将内容重新编入索引。

Amazon Kendra 表格搜索支持同义词（包括自定义同义词）。 Amazon Kendra 仅支持带有表格标签内的 HTML 表格的英语文档。

以下示例显示了查询结果中包含的表摘录。要查看包含查询响应（包括表格摘录）的 JSON 示例，请参阅查询响应和类型。

Python

```
import boto3
import pprint

kendra = boto3.client("kendra")

# Provide the index ID
index_id = <index-id>
# Provide the query text
query = "search string"

response = kendra.query(
        QueryText = query,
        IndexId = index_id)

print("\nSearch results for query: " + query + "\n")

for query_result in response["ResultItems"]:

    print("--------------------")
    print("Type: " + str(query_result["Type"]))
    print("Type: " + str(query_result["Format"]))
```

```
    if query_result["Type"]=="ANSWER" and query_result["Format"]=="TABLE":
        answer_table = query_result["TableExcerpt"]
        print(answer_table)

    if query_result["Type"]=="ANSWER" and query_result["Format"]=="TEXT":
        answer_text = query_result["DocumentExcerpt"]
        print(answer_text)

    if query_result["Type"]=="QUESTION_ANSWER":
        question_answer_text = query_result["DocumentExcerpt"]["Text"]
        print(question_answer_text)

    if query_result["Type"]=="DOCUMENT":
        if "DocumentTitle" in query_result:
            document_title = query_result["DocumentTitle"]["Text"]
            print("Title: " + document_title)
        document_text = query_result["DocumentExcerpt"]["Text"]
        print(document_text)

    print("------------------\n\n")
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.QueryRequest;
import software.amazon.awssdk.services.kendra.model.QueryResponse;
import software.amazon.awssdk.services.kendra.model.QueryResultItem;

public class SearchIndexExample {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();

        String query = "search string";
        String indexId = "index-id";

        QueryRequest queryRequest = QueryRequest
            .builder()
            .queryText(query)
            .indexId(indexId)
            .build();
```

```
        QueryResponse queryResponse = kendra.query(queryRequest);

        System.out.println(String.format("\nSearch results for query: %s", query));
        for(QueryResultItem item: queryResponse.resultItems()) {
            System.out.println("----------------------");
            System.out.println(String.format("Type: %s", item.type()));
            System.out.println(String.format("Format: %s", item.format()));

            switch(item.format()) {
                case TABLE:
                    String answerTable = item.TableExcerpt();
                    System.out.println(answerTable);
                    break;
            }

            switch(item.format()) {
                case TEXT:
                    String answerText = item.DocumentExcerpt();
                    System.out.println(answerText);
                    break;
            }

            switch(item.type()) {
                case QUESTION_ANSWER:
                    String questionAnswerText = item.documentExcerpt().text();
                    System.out.println(questionAnswerText);
                    break;
                case DOCUMENT:
                    String documentTitle = item.documentTitle().text();
                    System.out.println(String.format("Title: %s", documentTitle));
                    String documentExcerpt = item.documentExcerpt().text();
                    System.out.println(String.format("Excerpt: %s",
documentExcerpt));
                    break;
                default:
                    System.out.println(String.format("Unknown query result type:
%s", item.type()));

            }

            System.out.println("----------------------\n");
        }
    }
```

```
}
```

查询建议

Amazon Kendra 查询建议可以帮助您的用户更快地键入搜索查询并指导他们的搜索。

Amazon Kendra 根据以下其中一项建议与您的用户相关的查询：

- 查询历史记录或查询日志中的热门查询
- 文档字段/属性的内容

您可以通过将 SuggestionTypes 设置为 QUERY 或 DOCUMENT_ATTRIBUTES 并调用 GetQuerySuggestions 来设置使用查询历史记录或文档字段的首选项。默认情况下， Amazon Kendra 使用查询历史记录作为建议的基础。如果在您致电时查询历史记录和文档字段均已激活，UpdateQuerySuggestionsConfig并且您尚未将SuggestionTypes首选项设置为使用文档字段，则 Amazon Kendra 使用查询历史记录。

如果您使用控制台，则可以根据查询历史记录或文档字段提出查询建议。首先选择索引，然后在导航菜单的富集下选择查询建议。然后选择配置查询建议。配置查询建议后，系统会将您定向到搜索控制台，您可以在其中选择右侧面板中的查询历史记录或文档字段，然后在搜索栏中输入搜索查询。

默认情况下，使用查询历史记录和文档字段的查询建议均已激活，无需支付额外费用。您可以随时使用 UpdateQuerySuggestionsConfig API 停用这些类型的查询建议。要停用基于查询历史记录的查询建议，请在调用 UpdateQuerySuggestionsConfig 时将 Mode 设置为 DISABLED。要停用基于文档字段的查询建议，请在文档字段配置中将 AttributeSuggestionsMode 设置为 INACTIVE，然后调用 UpdateQuerySuggestionsConfig>。如果您使用控制台，则可以在查询建议设置中停用查询建议。

查询建议不区分大小写。 Amazon Kendra 将查询前缀和建议的查询转换为小写，忽略所有单引号和双引号，并将多个空格字符替换为单个空格。 Amazon Kendra 按原样匹配所有其他特殊字符。 Amazon Kendra 如果用户键入的字符少于两个或超过 60 个字符，则不会显示任何建议。

主题

- 使用查询历史记录查询建议
- 使用文档字段查询建议
- 从建议中屏蔽某些查询或文档字段内容

使用查询历史记录查询建议

主题

您可以选择根据查询历史记录或查询日志中的热门查询来建议与用户相关的查询。 Amazon Kendra 使用您的用户搜索并从这些查询中学到的所有查询来向您的用户提出建议。 Amazon Kendra 当用户开始键入查询时，向他们推荐热门查询。 Amazon Kendra 如果查询的前缀或前几个字符与用户开始键入的查询内容相匹配，则建议进行查询。

例如，用户开始键入查询“即将举行的活动”。 Amazon Kendra 从查询历史中了解到，许多用户已经多次搜索“2050 年即将举行的活动”。用户会看到“2050 年即将举行的活动”直接出现在他们的搜索栏下方，从而自动完成搜索查询。用户选择此查询建议，搜索结果中会返回文档“新事件：2050 年发生的事情”。

您可以指定如何 Amazon Kendra 选择符合条件的查询以向用户推荐。例如，您可以指定查询建议必须由至少 10 个独立用户进行搜索（默认值为 3 个），且在过去 30 天内搜索过，并且不包含屏蔽列表中的任何单词或短语。 Amazon Kendra 要求查询至少有一个搜索结果并且包含至少一个超过四个字符的单词。

为建议选择查询的设置

您可以使用 UpdateQuerySuggestionsConfig API 配置以下设置，以便为建议选择查询：

- 模式 - 使用查询历史记录为 ENABLED 或 LEARN_ONLY 的查询建议。 Amazon Kendra 默认情况下会激活查询建议。LEARN_ONLY 会关闭查询建议。如果禁用，则会 Amazon Kendra 继续学习建议，但不会向用户提出查询建议。
- 查询日志时段 - 查询在查询日志时段中的最近查询时间。时段是从当天到过去几天的天数的整数值。
- 不包含用户信息的查询 - 设置为 TRUE 以包含所有查询，或者设置为 FALSE 以仅包含具有用户信息的查询。如果您的搜索应用程序在用户发出查询时包含用户信息（例如用户 ID），则可以使用此设置。默认情况下，如果没有与查询关联的特定用户信息，则此设置不会筛选出查询。但是，您可以使用此设置仅根据包含用户信息的查询提出建议。
- 唯一用户 - 必须搜索查询才有资格向您的用户推荐查询的最少唯一用户数。此数字是一个整数值。
- 查询次数 - 必须搜索查询的最少次数，才有资格向用户推荐该查询。此数字是一个整数值。

这些设置会影响如何选择查询作为向用户推荐的热门查询。如何调整设置将取决于您的特定需求，例如：

- 如果您的用户通常平均每月搜索一次，则可以将查询日志时段中的天数设置为 30 天。通过使用该设置，您可以在用户最近的大多数查询在时段内过时之前捕获这些查询。
- 如果您的查询中只有少量包含用户信息，并且您不想根据较小的样本量建议查询，则可以将查询设置为包括所有用户。
- 如果您将热门查询定义为至少有 10 个独立用户搜索并且搜索了至少 100 次，则可以将唯一用户设置为 10，将查询计数设置为 100。

> **Warning**
>
> 您对设置的更改可能不会立即生效。您可以使用 DescribeQuerySuggestionsConfig API 跟踪设置更改。更新后的设置生效的时间取决于您所做的更新和索引中搜索查询的数量。在您更改设置或应用屏蔽列表后， Amazon Kendra 每 24 小时自动更新一次建议。

CLI

检索查询建议

```
aws kendra get-query-suggestions \
 --index-id index-id \
 --query-text "query-text" \
 --suggestion-types '["QUERY"]' \
 --max-suggestions-count 1 // If you want to limit the number of suggestions
```

更新查询建议

例如，要更改查询日志的时段和必须搜索查询的最小次数，请执行以下操作：

```
aws kendra update-query-suggestions-config \
 --index-id index-id \
 --query-log-look-back-window-in-days 30 \
 --minimum-query-count 100
```

Python

检索查询建议

```
import boto3
from botocore.exceptions import ClientError

kendra = boto3.client("kendra")

print("Get query suggestions.")

# Provide the index ID
index_id = "index-id"

# Provide the query text
query_text = "query"

# Provide the query suggestions type
query_suggestions_type = "QUERY"


# If you want to limit the number of suggestions
num_suggestions = 1

try:
    query_suggestions_response = kendra.get_query_suggestions(
        IndexId = index_id,
        QueryText = query_text,
        SuggestionTypes = query_suggestions_type,
        MaxSuggestionsCount = num_suggestions
    )

    # Print out the suggestions you received
    if ("Suggestions" in query_suggestions_response.keys()) {
        for (suggestion: query_suggestions_response["Suggestions"]) {
            print(suggestion["Value"]["Text"]["Text"]);
        }
    }

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

更新查询建议

例如，要更改查询日志的时段和必须搜索查询的最小次数，请执行以下操作：

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Updating query suggestions settings/configuration for an index.")

# Provide the index ID
index_id = "index-id"

# Configure the settings you want to update
minimum_query_count = 100
query_log_look_back_window_in_days = 30

try:
    kendra.update_query_suggestions_config(
        IndexId = index_id,
        MinimumQueryCount = minimum_query_count,
        QueryLogLookBackWindowInDays = query_log_look_back_window_in_days
    )

    print("Wait for Amazon Kendra to update the query suggestions.")

    while True:
        # Get query suggestions description of settings/configuration
        query_sugg_config_response = kendra.describe_query_suggestions_config(
            IndexId = index_id
        )

        # If status is not UPDATING, then quit
        status = query_sugg_config_response["Status"]
        print(" Updating query suggestions config. Status: " + status)
        if status != "UPDATING":
            break
        time.sleep(60)

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

在保留查询历史记录的同时清除建议

您可以使用 ClearQuerySuggestions API 清除查询建议。清除建议只会删除现有的查询建议，而不会删除查询历史记录中的查询。当您清除建议时， Amazon Kendra 会根据您清除建议后添加到查询日志中的新查询来学习新的建议。

CLI

清除查询建议

```
aws  kendra clear-query-suggestions \
 --index-id index-id
```

Python

清除查询建议

```
import boto3
from botocore.exceptions import ClientError

kendra = boto3.client("kendra")

print("Clearing out query suggestions for an index.")

# Provide the index ID
index_id = "index-id"

try:
    kendra.clear_query_suggestions(
        IndexId = index_id
    )

    # Confirm last cleared date-time and that there are no suggestions
    query_sugg_config_response = kendra.describe_query_suggestions_config(
        IndexId = index_id
    )
    print("Query Suggestions last cleared at: " +
 str(query_sugg_config_response["LastClearTime"]));
    print("Number of suggestions available from the time of clearing: " +
 str(query_sugg_config_response["TotalSuggestionsCount"]));

except ClientError as e:
```

```
        print("%s" % e)

print("Program ends.")
```

没有可用的建议

如果未看到查询建议，可能是出于以下原因之一：

- 您的索引中没有足够的查询 Amazon Kendra 可供借鉴。
- 您的查询建议设置过于严格，导致大多数查询都被从建议中筛选掉。
- 您最近清除了建议， Amazon Kendra 但仍需要时间积累新的查询，以便学习新的建议。

您可以使用 DescribeQuerySuggestionsConfig API 检查当前设置。

使用文档字段查询建议

主题

- 为建议选择字段的设置
- 文档字段中的用户控制

您可以选择根据文档字段的内容建议与用户相关的查询。您可以使用文档字段中包含的有助于自动完成查询的信息，而不是使用查询历史记录来建议其他常见的相关查询。 Amazon Kendra 在设置为Suggestable且与用户查询非常一致的字段中查找相关内容。然后，当您的用户开始键入查询内容时，向他们 Amazon Kendra 推荐这些内容。

例如，如果您指定了作为建议基础的标题字段，然后用户开始键入查询 “How amazon ken... '，可以建议使用最相关的标题 “ Amazon Kendra 工作原理” 来自动完成搜索。用户会看到 “ Amazon Kendra 工作原理” 直接出现在他们的搜索栏下方，自动完成了他们的搜索查询。用户选择此查询建议，搜索结果中将返回 “ Amazon Kendra 工作原理” 文档。

您可以使用 String 和 StringList 类型的任意文档字段的内容来建议查询，方法是将该字段作为查询建议字段配置的一部分设置为 Suggestable。您也可以使用屏蔽列表，这样就不会向用户显示包含某些单词或短语的建议文档字段。您可以使用一个阻止列表。无论您将查询建议设置为使用查询历史记录还是文档字段，屏蔽列表都适用。

为建议选择字段的设置

您可以使用 AttributeSuggestionsConfig 并调用 UpdateQuerySuggestionsConfig API 来更新索引级别的设置来配置以下设置，以便为建议选择文档字段：

- 字段/属性建议模式 - 使用文档字段为 ACTIVE 或 INACTIVE 的查询建议。 Amazon Kendra 在默认情况下会激活查询建议。
- 可建议的字段/属性 - 作为建议基础的字段名称或字段键。作为字段配置的一部分，必须将这些字段设置为 Suggestable 的 TRUE。您可以在查询级别覆盖字段配置，同时保持索引级别的配置。使用 GetQuerySuggestionsAPI 在查询AttributeSuggestionConfig级别进行更改。查询级别的这种配置对于快速尝试使用不同的文档字段非常有用，而不必在索引级别更新配置。
- 其他字段/属性 - 您要在查询建议的响应中包含的其他字段。这些字段用于在回复中提供额外信息；但是，它们不用于作为建议的依据。

> **Warning**
>
> 您对设置的更改可能不会立即生效。您可以使用 DescribeQuerySuggestionsConfig API 跟踪设置更改。更新后的设置生效的时间取决于您所做的更新。 Amazon Kendra 在您更改设置后或应用屏蔽列表后，每 24 小时自动更新一次建议。

CLI

在查询级别检索查询建议并覆盖文档字段配置，而不必在索引级别更改配置。

```
aws kendra get-query-suggestions \
 --index-id index-id \
 --query-text "query-text" \
 --suggestion-types '["DOCUMENT_ATTRIBUTES"]' \
 --attribute-suggestions-config '{"SuggestionAttributes":'["field/attribute key
 1", "field/attribute key 2"]', "AdditionalResponseAttributes":'["response field/
attribute key 1", "response field/attribute key 2"]'}' \
 --max-suggestions-count 1 // If you want to limit the number of suggestions
```

更新查询建议

例如，要更改索引级别的文档字段配置，请执行以下操作：

```
aws kendra update-query-suggestions-config \
```

```
 --index-id index-id \
 --attribute-suggestions-config '{"SuggestableConfigList": '[{"SuggestableConfig":
 "_document_title", "Suggestable": true}]', "AttributeSuggestionsMode": "ACTIVE"}'
```

Python

在查询级别检索查询建议并覆盖文档字段配置，而不必在索引级别更改配置。

```
import boto3
from botocore.exceptions import ClientError

kendra = boto3.client("kendra")

print("Get query suggestions.")

# Provide the index ID
index_id = "index-id"

# Provide the query text
query_text = "query"

# Provide the query suggestions type
query_suggestions_type = "DOCUMENT_ATTRIBUTES"

# Override fields/attributes configuration at query level
configuration = {"SuggestionAttributes":
    '["field/attribute key 1", "field/attribute key 2"]',
      "AdditionalResponseAttributes":
          '["response field/attribute key 1", "response field/attribute key 2"]'
          }

# If you want to limit the number of suggestions
num_suggestions = 1

try:
    query_suggestions_response = kendra.get_query_suggestions(
        IndexId = index_id,
        QueryText = query_text,
        SuggestionTypes = [query_suggestions_type],
        AttributeSuggestionsConfig = configuration,
        MaxSuggestionsCount = num_suggestions
    )

    # Print out the suggestions you received
```

```
    if ("Suggestions" in query_suggestions_response.keys()) {
        for (suggestion: query_suggestions_response["Suggestions"]) {
            print(suggestion["Value"]["Text"]["Text"]);
        }
    }

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

更新查询建议

例如，要更改索引级别的文档字段配置，请执行以下操作：

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Updating query suggestions settings/configuration for an index.")

# Provide the index ID
index_id = "index-id"

# Configure the settings you want to update at the index level
configuration = {"SuggestableConfigList":
    '[{"SuggestableConfig": "_document_title", "Suggestable": true}]',
       "AttributeSuggestionsMode": "ACTIVE"
       }

try:
    kendra.update_query_suggestions_config(
        IndexId = index_id,
        AttributeSuggestionsConfig = configuration
    )

    print("Wait for Amazon Kendra to update the query suggestions.")

    while True:
        # Get query suggestions description of settings/configuration
        query_sugg_config_response = kendra.describe_query_suggestions_config(
```

```
            IndexId = index_id
        )

        # If status is not UPDATING, then quit
        status = query_sugg_config_response["Status"]
        print(" Updating query suggestions config. Status: " + status)
        if status != "UPDATING":
            break
        time.sleep(60)

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

文档字段中的用户控制

您可以将用户上下文筛选应用于要作为查询建议基础的文档字段。这会根据用户或其群组对文档的访问权限来筛选文档字段信息。例如，实习生搜索公司的门户网站，却无权访问公司绝密文档。因此，不会向实习生显示基于绝密文档标题或任何其他可建议字段的建议查询。

您可以使用访问控制列表（ACL）为文档编制索引，该列表定义为哪些用户和组分配了对哪些文档的访问权限。然后，您可以对文档字段应用用户上下文筛选以获得查询建议。当前为您的索引设置的用户上下文筛选与应用于您的文档字段配置的查询建议的用户上下文筛选相同。用户上下文筛选是您的文档字段配置的一部分。您使用 AttributeSuggestionsGetConfig 并调用 GetQuerySuggestions。

从建议中屏蔽某些查询或文档字段内容

屏蔽 Amazon Kendra 列表不会向您的用户建议某些查询。屏蔽列表是您要从查询建议中排除的单词或短语的列表。 Amazon Kendra 不包括包含与屏蔽列表中的单词或短语完全匹配的查询。

您可以使用屏蔽列表来防范经常出现在您的查询历史记录或文档字段中，并且 Amazon Kendra 可能将其选为建议的冒犯性单词或短语。屏蔽列表还可以 Amazon Kendra 防止建议包含尚未准备好公开发布或宣布的信息的查询。例如，您的用户经常询问即将发布的潜在新产品。但是，您不想推荐该产品，因为您尚未准备好发布该产品。您可以屏蔽建议中包含产品名称和产品信息的查询。

您可以使用 CreateQuerySuggestionsBlockList API 为查询创建阻止列表。将每个屏蔽单词或短语放在文本文件中单独一行中。然后，您将文本文件上传到您的 Amazon S3 存储桶，并在中提供该文件的路径或位置 Amazon S3。 Amazon Kendra 目前仅支持创建一个阻止列表。

您可以替换 Amazon S3 存储桶中屏蔽的单词和短语的文本文件。要更新中的屏蔽列表 Amazon Kendra，请使用 UpdateQuerySuggestionsBlockListAPI。

使用 DescribeQuerySuggestionsBlockList API 获取屏蔽列表的状态。DescribeQuerySuggestionsBlockList 还可以为您提供其他有用的信息，例如以下信息：

- 上次更新屏蔽名单的时间
- 您当前的屏蔽列表中有多少个单词或短语
- 创建屏蔽名单时的有用错误消息

您还可以使用 ListQuerySuggestionsBlockLists API 获取索引的屏蔽列表摘要列表。

要删除您的屏蔽名单，请使用 DeleteQuerySuggestionsBlockListAPI。

您对屏蔽列表的更新可能不会立即生效。您可以使用 DescribeQuerySuggestionsBlockList API 跟踪更新。

CLI

创建屏蔽列表

```
aws kendra create-query-suggestions-block-list \
 --index-id index-id \
 --name "block-list-name" \
 --description "block-list-description" \
 --source-s3-path "Bucket=bucket-name,Key=query-suggestions/block_list.txt" \
 --role-arn role-arn
```

更新屏蔽列表

```
aws kendra update-query-suggestions-block-list \
 --index-id index-id \
 --name "new-block-list-name" \
 --description "new-block-list-description" \
 --source-s3-path "Bucket=bucket-name,Key=query-suggestions/new_block_list.txt" \
 --role-arn role-arn
```

删除屏蔽列表

```
aws kendra delete-query-suggestions-block-list \
```

```
 --index-id index-id \
 --id block-list-id
```

Python

创建屏蔽列表

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Create a query suggestions block list.")

# Provide a name for the block list
block_list_name = "block-list-name"
# Provide an optional description for the block list
block_list_description = "block-list-description"
# Provide the IAM role ARN required for query suggestions block lists
block_list_role_arn = "role-arn"

# Provide the index ID
index_id = "index-id"

s3_bucket_name = "bucket-name"
s3_key = "query-suggestions/block_list.txt"
source_s3_path = {
    'Bucket': s3_bucket_name,
    'Key': s3_key
}

try:
    block_list_response = kendra.create_query_suggestions_block_list(
        Description = block_list_description,
        Name = block_list_name,
        RoleArn = block_list_role_arn,
        IndexId = index_id,
        SourceS3Path = source_s3_path
    )

    print(block_list_response)
```

```
    block_list_id = block_list_response["Id"]

    print("Wait for Amazon Kendra to create the block list.")

    while True:
        # Get block list description
        block_list_description = kendra.describe_query_suggestions_block_list(
            Id = block_list_id,
            IndexId = index_id
        )
        # If status is not CREATING, then quit
        status = block_list_description["Status"]
        print("Creating block list. Status: " + status)
        if status != "CREATING":
            break
        time.sleep(60)

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

更新屏蔽列表

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Update a block list for query suggestions.")

# Provide the block list name you want to update
block_list_name = "new-block-list-name"
# Provide the block list description you want to update
block_list_description = "new-block-list-description"
# Provide the IAM role ARN required for query suggestions block lists
block_list_role_arn = "role-arn"

# Provide the block list ID
block_list_id = "block-list-id"
# Provide the index ID
```

```
index_id = "index-id"

s3_bucket_name = "bucket-name"
s3_key = "query-suggestions/new_block_list.txt"
source_s3_path = {
'Bucket': s3_bucket_name,
'Key': s3_key
}

try:
    kendra.update_query_suggestions_block_list(
        Id = block_list_id,
        IndexId = index_id,
        Description = block_list_description,
        Name = block_list_name,
        RoleArn = block_list_role_arn,
        SourceS3Path = source_s3_path
    )

    print("Wait for Amazon Kendra to update the block list.")

    while True:
        # Get block list description
        block_list_description = kendra.describe_query_suggestions_block_list(
            Id = block_list_id,
            IndexId = index_id
        )
        # If status is not UPDATING, then the update has finished
        status = block_list_description["Status"]
        print("Updating block list. Status: " + status)
        if status != "UPDATING":
            break
        time.sleep(60)

except ClientError as e:
print("%s" % e)

print("Program ends.")
```

删除屏蔽列表

```
import boto3
from botocore.exceptions import ClientError
```

```
kendra = boto3.client("kendra")

print("Delete a block list for query suggestions.")

# provide the block list ID
query_suggestions_block_list_id = "query-suggestions-block-list-id"
# Provide the index ID
index_id = "index-id"

try:
    kendra.delete_query_suggestions_block_list(
        Id = query_suggestions_block_list_id,
        IndexId = index_id
    )

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

查询拼写检查程序

Amazon Kendra 拼写检查工具会建议对查询进行拼写更正。这可以帮助您将出现零搜索结果的次数降至最低，并返回相关结果。对于拼写错误的查询，您的用户可能会收到零搜索结果为，即没有匹配结果或未返回文档。或者，您的用户可能会从拼写错误的查询中收到不相关的搜索结果。

拼写检查器旨在根据索引文档中出现的单词以及更正后的单词与拼写错误的单词的匹配程度，对拼写错误的单词提出更正建议。例如，如果您的索引文档中出现“statements”一词，则这可能与查询“年终财务报表”中拼写错误的“statments”一词非常匹配。

拼写检查器返回替换原始查询文本中拼写错误的单词的预期或更正的单词。例如，“depoying kendre search”可能返回“deploying Kendra search”。您还可以使用 API 中提供的偏移位置在前端应用程序的查询中突出显示或斜体显示返回的更正单词。在控制台中，默认情况下，更正后的单词会突出显示或斜体显示。例如，“deploying Kendra search”。

对于索引文档中出现的业务特定术语或专业术语，拼写检查工具不会将这些术语误解为查询中的拼写错误。例如，“amazon macie”不会更正为“amazon mace”。

对于带连字符的单词，例如“year-end”，拼写检查器会将其视为单个单词，以建议对这些单词进行更正。例如，“yaer-end”的建议更正可能是“year-end”。

对于 DOCUMENT 和 QUESTION_ANSWER 查询响应类型，拼写检查器会根据文档正文中的单词建议更正拼写错误的单词。在建议与拼写错误的单词非常匹配的更正时，文档正文比标题更可靠。对于 ANSWER 查询响应类型，拼写检查器会根据索引中默认问答文档中的单词建议更正。

你可以使用该SpellCorrectionConfiguration物体激活拼写检查器。将 IncludeQuerySpellCheckSuggestions 设置为 TRUE。默认情况下，控制台中的拼写检查器处于激活状态。默认情况下，它内置在控制台中。

拼写检查器还可以为多种语言的查询提供拼写更正建议，而不仅仅是英语。有关拼写检查器支持的语言列表，请参阅 Amazon Kendra 支持的语言。

使用带有默认限制的查询拼写检查器

拼写检查器设计有特定的默认值或限制。以下是激活拼写校正建议时适用的当前限制列表。

- 对于长度少于三个字符或大于 30 个字符的单词，不能返回建议的拼写更正。要允许超过 30 个字符或少于三个字符，请联系支持部门。
- 建议的拼写更正不能根据用户访问控制或用户上下文筛选的访问控制列表来限制建议。拼写校正基于已编入索引的文档中的所有单词，无论这些单词是否仅限于某些用户。如果您想避免某些单词出现在建议的查询拼写更正中，请不要激活 SpellCorrectionConfiguration。
- 对于包含数字的单词，无法返回建议的拼写更正。例如，“how 2 not br8k ubun2”。
- 建议的拼写更正不能使用未出现在已编入索引的文档中的单词。
- 建议的拼写更正不能使用索引文档中出现频率低于 0.01% 的单词。要更改 0.01% 的阈值，请联系支持部门。

筛选和分面搜索

您可以使用筛选条件来改善来自 Query API 的搜索结果或响应。筛选条件将响应中的文档限制为直接应用于查询的文档。要创建分面搜索建议，请使用布尔逻辑从响应或不符合特定条件的文档中筛选出特定的文档属性。您可以使用 Query API 中的 Facets 参数指定分面。

要搜索已编入索引的 Amazon Kendra 文档 Amazon Lex，请使用亚马逊。 KendraSearchIntent。有关使用进行配置的示例 Amazon Kendra Amazon Lex，请参阅为 Amazon Kendra 索引创建常见问题解答机器人。您也可以使用为响应提供过滤器AttributeFilter。这是配置 AMAZON.KendraSearchIntent 时采用 JSON 格式的查询筛选条件。要在控制台中配置搜索意图时提供属性筛选条件，请转到意图编辑器并选择 Amazon Kendra 查询以提供 JSON 格式的查询筛选条件。有关 AMAZON.KendraSearchIntent 的更多信息，请参阅 Amazon Lex 文档。

分面

分面是一组搜索结果的限定视图。例如，您可以为世界各地的城市提供搜索结果，其中文档按与其关联的特定城市进行筛选。或者，您可以创建分面来显示特定作者的结果。

您可以将与文档关联的文档属性或元数据字段用作分面，这样您的用户就可以按类别或该分面内的值进行搜索。您还可以在搜索结果中显示嵌套分面，这样您的用户不仅可以按类别或字段进行搜索，还可以按子类别或子字段进行搜索。

以下示例显示了如何获取“City”自定义属性的分面信息。

```
response=kendra.query(
        QueryText = query,
        IndexId = index,
        Facets = [
            {
                "DocumentAttributeKey" : "City"
            }
        ]
        )
```

您可以使用嵌套分面来进一步缩小搜索范围。例如，文档属性或分面“City”包括一个名为“Seattle”的值。此外，文档属性或分面 “CityRegion” 包括分配给 “西雅图” 的文档的 “北” 和 “南” 值。您可以在搜索结果中显示嵌套的刻面及其计数，这样不仅可以按城市搜索文档，还可以按城市内的区域搜索文档。

请注意，嵌套分面可能会影响查询延迟。一般规则是，使用的术语和运算符越多，对延迟的潜在影响就越大。影响延迟的其他因素包括已编入索引的文档的平均大小、索引的大小、高度复杂的查询以及 Amazon Kendra 索引的总体负载。

以下示例说明如何获取 “” 自定义属性的分面信息，作为 “CityCityRegion” 中的嵌套分面。

```
response=kendra.query(
        QueryText = query,
        IndexId = index,
        Facets = [
            {
                "DocumentAttributeKey" : "City",
                "Facets": [
                    {
                        "DocumentAttributeKey" : "CityRegion"
                    }
```

```
                ]
            }
        ]
        )
```

分面信息（例如文档数量）将在 FacetResults 响应数组中返回。您可以使用这些内容在应用程序中显示分面搜索建议。例如，如果文档属性“City”包含可应用搜索的城市，则使用该信息显示城市搜索列表。用户可以选择城市来筛选搜索结果。要进行分面搜索，请调用 Query API 并使用所选文档属性筛选结果。

对于查询，每个分面最多可以显示 10 个分面值，并且在一个分面内只能显示一个嵌套分面。如果要增加这些限制，请联系支持部门。如果要将每个刻面的刻面值数限制为小于 10，则可以在 Facet 对象中进行指定。

以下 JSON 响应示例，显示了限于“City”文档属性的各个方面。响应包括分面值的文档数量。

```
{
    'FacetResults': [
        {
            'DocumentAttributeKey': 'City',
            'DocumentAttributeValueCountPairs': [
                {
                    'Count': 3,
                    'DocumentAttributeValue': {
                        'StringValue': 'Dubai'
                    }
                },
                {
                    'Count': 3,
                    'DocumentAttributeValue': {
                        'StringValue': 'Seattle'
                    }
                },
                {
                    'Count': 1,
                    'DocumentAttributeValue': {
                        'StringValue': 'Paris'
                    }
                }
            ]
        }
    ]
```

您还可以显示嵌套分面（例如城市中的区域）的刻面信息，以进一步筛选搜索结果。

以下 JSON 响应示例，将范围限于 “CityRegion” 文档属性的分面显示为 “城市” 中的嵌套分面。响应包括嵌套分面值的文档数量。

```
{
    'FacetResults': [
        {
            'DocumentAttributeKey': 'City',
            'DocumentAttributeValueCountPairs': [
                {
                    'Count': 3,
                    'DocumentAttributeValue': {
                        'StringValue': 'Dubai'
                    },
                    'FacetResults': [
                        {
                            'DocumentAttributeKey': 'CityRegion',
                            'DocumentAttributeValueCountPairs': [
                                {
                                    'Count': 2,
                                    'DocumentAttributeValue': {
                                        'StringValue': 'Bur Dubai'
                                    }
                                },
                                {
                                    'Count': 1,
                                    'DocumentAttributeValue': {
                                        'StringValue': 'Deira'
                                    }
                                }
                             ]
                        }
                    ]
                },
                {
                    'Count': 3,
                    'DocumentAttributeValue': {
                        'StringValue': 'Seattle'
                    },
                    'FacetResults': [
                        {
                            'DocumentAttributeKey': 'CityRegion',
                            'DocumentAttributeValueCountPairs': [
```

```
                                {
                                    'Count': 1,
                                    'DocumentAttributeValue': {
                                        'StringValue': 'North'
                                    }
                                },
                                {
                                    'Count': 2,
                                    'DocumentAttributeValue': {
                                        'StringValue': 'South'
                                    }
                                }
                            ]
                        }
                    ]
                },
                {
                    'Count': 1,
                    'DocumentAttributeValue': {
                        'StringValue': 'Paris'
                    },
                    'FacetResults': [
                        {
                            'DocumentAttributeKey': 'CityRegion',
                            'DocumentAttributeValueCountPairs': [
                                {
                                    'Count': 1,
                                    'DocumentAttributeValue': {
                                        'StringValue': 'City center'
                                    }
                                }
                            ]
                        }
                    ]
                }
        }
    ]
}
```

使用字符串列表字段创建分面时，返回的分面结果基于字符串列表的内容。例如，如果您有一个包含两个项目的字符串列表字段，一个列表为“dachshund”、“sausage dog”，另一个的值为“husky”，那么 `FacetResults` 就会得到三个分面。

有关更多信息，请参阅 查询响应和响应类型。

使用文档属性筛选搜索结果

默认情况下，Query 会返回所有搜索结果。要筛选响应，可以对文档属性执行逻辑操作。例如，如果您只需要特定城市的文档，则可以筛选“City”和“State”的自定义文档属性。您可以使用AttributeFilter对提供的过滤器创建布尔运算。

大多数属性可用于筛选所有响应类型的响应。但是，_excerpt_page_number 属性仅适用于筛选响应时的 ANSWER 响应类型。

以下示例说明如何通过筛选特定城市（西雅图和华盛顿州）来执行逻辑 AND 运算。

```
response=kendra.query(
        QueryText = query,
        IndexId = index,
        AttributeFilter = {'AndAllFilters':
            [
                {"EqualsTo": {"Key": "City","Value": {"StringValue": "Seattle"}}},
                {"EqualsTo": {"Key": "State","Value": {"StringValue": "Washington"}}}
            ]
            }
        )
```

以下示例说明如何在Fileformat、Author或SourceURI键中任意一个与指定值匹配时执行逻辑 OR 运算。

```
response=kendra.query(
        QueryText = query,
        IndexId = index,
        AttributeFilter = {'OrAllFilters':
            [
                {"EqualsTo": {"Key": "Fileformat","Value": {"StringValue":
 "AUTO_DETECT"}}},
                {"EqualsTo": {"Key": "Author","Value": {"StringValue": "Ana
 Carolina"}}},
                {"EqualsTo": {"Key": "SourceURI","Value": {"StringValue": "https://
aws.amazonaws.com/234234242342"}}}
            ]
            }
        )
```

对于StringList字段，使用ContainsAny或ContainsAll属性筛选条件返回包含指定字符串的文档。以下示例说明如何返回Locations自定义属性中值为“Seattle”或“Portland”的所有文档。

```
response=kendra.query(
        QueryText = query,
        IndexId = index,
        AttributeFilter = {
                "ContainsAny": { "Key": "Locations", "Value": { "StringListValue":
 [ "Seattle", "Portland"] }}
            }
        )
```

筛选搜索结果中每个文档的属性

Amazon Kendra 返回搜索结果中每个文档的文档属性。您可以筛选要作为搜索结果一部分包含在响应中的某些文档属性。默认情况下，分配给文档的所有文档属性都将在响应中返回。

在以下示例中，_author文档的响应中仅包含_source_uri和文档属性。

```
response=kendra.query(
        QueryText = query,
        IndexId = index,
        RequestedDocumentAttributes = ["_source_uri", "_author"]
        )
```

根据用户上下文进行筛选

您可以根据用户或其群组对文档的访问权限来筛选用户的搜索结果。您可以使用用户令牌、用户 ID 或用户属性来筛选文档。 Amazon Kendra 也可以将用户映射到他们的群组。您可以选择 AWS IAM Identity Center 用作您的身份存储/来源。

用户上下文筛选是一种个性化搜索，其优点是控制对文档的访问权限。例如，并非所有在公司门户网站上搜索信息的团队都应该访问绝密的公司文档，这些文档也不应与所有用户相关。只有获得绝密文档访问权限的特定用户或团队组才能在搜索结果中看到这些文档。

将文档编入索引后 Amazon Kendra，会为大多数文档提取相应的访问控制列表 (ACL)。ACL 指定允许或拒绝哪些用户名和组名访问文档。没有 ACL 的文档是公共文档。

Amazon Kendra 可以为大多数数据源提取与每个文档关联的用户或群组信息。例如，Quip 中的文档可以包含有权访问该文档的精选用户的“共享”列表。如果您使用 S3 存储桶作为数据来源，则需要为 ACL

提供一个 JSON 文件，并将该文件的 S3 路径作为数据来源配置的一部分。如果将文档直接添加到索引，则可以在 Princip al 对象中指定 ACL 作为 BatchPutDocumentAPI 中文档对象的一部分。

您可以使用 CreateAccessControlConfigurationAPI 重新配置现有的文档级别访问控制，而无需再次为所有文档编制索引。例如，您的索引包含只有特定员工或用户才能访问的绝密公司文档。其中一位用户离开公司或转到应被禁止访问绝密文档的团队。用户仍然可以访问绝密文档，因为在您的文档之前被编入索引时，该用户拥有访问权限。您可以为具有拒绝访问权限的用户创建特定的访问控制配置。您可以稍后更新访问控制配置，以便在用户返回公司并重新加入“绝密”团队时允许访问。随着情况的变化，您可以重新配置文档的访问控制。

要将您的访问控制配置应用于某些文档，请使用 D oc ument 对象中`AccessControlConfigurationId`包含的 BatchPutDocumentAPI。如果您使用 S3 存储桶作为数据源，则`.metadata.json`使用更新`AccessControlConfigurationId`并同步您的数据源。Amazon Kendra 目前仅支持对使用 `BatchPutDocument` API 编制索引的 S3 数据源和文档进行访问控制配置。

按用户令牌筛选

查询索引时，您可以使用用户令牌根据用户或其群组对文档的访问权限筛选搜索结果。当您发出查询时， Amazon Kendra 提取并验证令牌，提取和检查用户和群组信息，然后运行查询。将返回用户有权访问的所有文档，包括公共文档。有关更多信息，请参阅基于角色的访问控制。

您在UserContext对象中提供用户令牌，然后在查询 API 中传递该令牌。

以下命令说明如何添加用户令牌。

```
response = kendra.query(
    QueryText = query,
    IndexId = index,
    UserToken = {
        Token = "token"
    })
```

您可以将用户映射到群组。使用用户上下文筛选时，无需在发出查询时包含用户所属的所有群组。使用 PutPrincipalMappingAPI，您可以将用户映射到他们的群组。如果您不想使用 `PutPrincipalMapping` API，则必须在发出查询时提供用户名和该用户所属的所有群组。您还可以使用UserGroupResolutionConfiguration对象获取 IAM Identity Center 身份源中群组和用户的访问级别。

按用户 ID 和群组筛选

查询索引时，您可以使用用户 ID 和群组，根据用户或其群组对文档的访问权限筛选搜索结果。当您发出查询时， Amazon Kendra 会检查用户和群组信息并运行查询。将返回与用户有权访问的查询相关的所有文档，包括公共文档。

您还可以按用户和群组有权访问的数据来源筛选搜索结果。如果一个组与多个数据来源相关联，但您只想让该组访问特定数据来源的文档，则指定数据来源非常有用。例如，“研究”、“工程”和“销售和营销”这三个组都与存储在数据来源Confluence和Salesforce中的公司文档相关联。但是，“销售和营销”团队只需要访问存储在Salesforce中的客户相关文档即可。因此，当销售和营销用户搜索与客户相关的文档时，他们可以在搜索结果中看到来自 Salesforce 的文档。不从事销售和市场营销工作的用户不会在搜索结果中看到 Salesforce 文档。

您在UserContext对象中提供用户、群组和数据源信息，然后在 Query API 中传递这些信息。用户 ID 以及组和数据来源列表应与您在 Princip al 对象中指定的名称相匹配，以标识用户、组和数据来源。使用该`Principal`对象，您可以将用户、组或数据来源添加到允许列表或拒绝列表中以访问文档。

您必须提供以下任一信息：

- 用户和群组信息，以及（可选）数据来源信息。
- 仅当您使用 PutPrincipalMappingAPI 将用户映射到群组和数据源时，才会显示用户信息。您还可以使用UserGroupResolutionConfiguration对象获取 IAM Identity Center 身份源中群组和用户的访问级别。

如果查询中未包含此信息，则 Amazon Kendra 返回所有文档。如果您提供此信息，则仅返回具有匹配用户 ID、群组和数据来源的文档。

以下内容显示了如何包括用户 ID、群组和数据来源。

```
response = kendra.query(
    QueryText = query,
    IndexId = index,
    UserId = {
        UserId = "user1"
    },
    Groups = {
        Groups = ["Sales and Marketing"]
    },
    DataSourceGroups = {
```

```
        DataSourceGroups = [{"DataSourceId" : "SalesforceCustomerDocsGroup", "GroupId":
 "Sales and Marketing"}]
    })
```

按属性筛选

查询索引时，您可以使用内置属性，_user_id并_group_id根据用户及其群组对文档的访问权限筛选搜索结果。您最多可以设置 100 个群组标识符。当您发出查询时， Amazon Kendra 会检查用户和群组信息并运行查询。将返回与用户有权访问的查询相关的所有文档，包括公共文档。

你在AttributeFilter对象中提供用户和群组属性，然后在 Query API 中传递这些属性。

以下示例显示了一个请求，该请求根据用户 ID 以及用户所属的“HR”和“IT”组筛选查询响应。该查询将返回允许列表中包含用户或“HR”或“IT”组的所有文档。如果该用户或其中一个组在文档的拒绝列表中，则不会返回该文档。

```
response = kendra.query(
        QueryText = query,
        IndexId = index,
        AttributeFilter = {
            "OrAllFilters": [
                {
                    "EqualsTo": {
                        "Key": "_user_id",
                        "Value": {
                            "StringValue": "user1"
                        }
                     }
                },
                {
                    "EqualsTo": {
                        "Key": "_group_ids",
                        "Value": {
                            "StringListValue": ["HR", "IT"]
                        }
                    }
                }
            ]
        }
        )
```

您还可以在Principal对象中指定群组可以访问哪个数据来源。

> **Note**
>
> 用户上下文筛选不是对内容的身份验证或授权控制。它不会对发送到 Query API 的用户和群组进行用户身份验证。您的应用程序有责任确保发送到 Query API 的用户和群组信息经过身份验证和授权。

每个数据来源都实现了用户上下文筛选。以下部分描述了每种实现。

主题

对直接添加到索引的文档进行用户上下文筛选

使用 BatchPutDocumentAPI 将文档直接添加到索引时，Amazon Kendra 会从文档的AccessControlList字段中获取用户和群组信息。您为文档提供访问控制列表（ACL），ACL 随文档一起提取。

您可以在 BatchPutDocument API 中将 Principal 对象中的 ACL 指定为 Document 对象的一部分。提供以下信息：

- 用户或组应具有的访问权限。您可以说ALLOW或DENY。
- 实体的类型。您可以说USER或GROUP。
- 用户或组的名称。

您最多可以在AccessControlList字段中添加 200 个条目。

筛选用户上下文以查找常见问题

向索引添加常见问题解答时，Amazon Kendra 会从 FAQ JSON 文件的AccessControlList对象/字段中获取用户和群组信息。您也可以使用带有自定义字段或属性的常见问题 CSV 文件进行访问控制。

提供以下信息：

- 用户或组应具有的访问权限。您可以说ALLOW或DENY。

- 实体的类型。您可以说USER或GROUP。
- 用户或组的名称。

有关更多信息，请参阅 常见问题。

数据来源的用户上下文筛选

Amazon Kendra 还会从支持的数据源连接器中搜寻用户和组访问控制列表 (ACL) 信息。这对于用户上下文筛选非常有用，在这种筛选中，搜索结果是根据用户或其群组对文档的访问权限进行筛选的。

主题

- 针对 Adobe 体验管理器数据来源的用户上下文筛选
- Alfresco 数据来源的用户上下文筛选
- 针对 Aurora (MySQL) 数据源的用户上下文筛选
- 筛选 Aurora （ PostgreSQL ）数据来源的用户上下文
- Amazon FSx 数据源的用户上下文筛选
- 数据库数据来源的用户上下文筛选
- Amazon RDS （ Microsoft SQL Server ）数据来源的用户上下文筛选
- 针对 Amazon RDS (MySQL) 数据源的用户上下文筛选
- 筛选 Amazon RDS (Oracle) 数据源的用户上下文
- 筛选 Amazon RDS （ PostgreSQL ）数据来源的用户上下文
- Amazon S3 数据来源的用户上下文筛选
- Amazon WorkDocs 数据源的用户上下文筛选
- Box 数据来源的用户上下文筛选
- Confluence 数据来源的用户上下文筛选
- 针对 Dropbox 数据来源的用户上下文筛选
- Drupal 数据来源的用户上下文筛选
- GitHub 数据源的用户上下文筛选
- Gmail 数据来源的用户上下文筛选
- 筛选 Google 云端硬盘数据来源的用户上下文
- 筛选 IBM DB2 数据来源的用户上下文
- Jira 数据来源的用户上下文筛选

- Microsoft Exchange 数据来源的用户上下文筛选
- 微软 OneDrive 数据源的用户上下文筛选
- 微软 OneDrive v2.0 数据源的用户上下文筛选
- 微软 SharePoint 数据源的用户上下文筛选
- Microsoft SQL Server 数据来源的用户上下文筛选
- Microsoft 团队数据来源的用户上下文筛选
- Microsoft Yammer 数据来源的用户上下文筛选
- 针对 MySQL 数据来源的用户上下文筛选
- 筛选 Oracle 数据库数据来源的用户上下文
- PostgreSQL 数据来源的用户上下文筛选
- Quip 数据来源的用户上下文筛选
- Salesforce 数据来源的用户上下文筛选
- ServiceNow 数据来源的用户上下文筛选
- Slack 数据来源的用户上下文筛选
- Zendesk 数据来源的用户上下文筛选

针对 Adobe 体验管理器数据来源的用户上下文筛选

当您使用 Adobe Experience Manager 数据源时， Amazon Kendra 会从 Adobe Experience Manager 实例获取用户和群组信息。

群组和用户 ID 映射如下：

- `_group_ids`-群组 ID 存在于 Adobe Experience Manager 内容中，其中设置了访问权限。它们是从 Adobe 体验管理器中的群组名称映射出来的。
- `_user_id`-用户 ID 存在于 Adobe Experience Manager 内容中，其中设置了访问权限。它们从用户电子邮件中映射为 Adobe Experience Manager 中的 ID。

您最多可以在`AccessControlList`字段中添加 200 个条目。

Alfresco 数据来源的用户上下文筛选

当你使用 Alfresco 数据源时， Amazon Kendra 会从 Alfresco 实例中获取用户和群组信息。

群组和用户 ID 映射如下：

- `_group_ids`-群组 ID 存在于 Alfresco 中设置了访问权限的文件上。它们是从 Alfresco 中组的系统名称（不是显示名称）映射出来的。
- `_user_id`-Alfresco 中存在已设置访问权限的文件上的用户 ID。它们从用户电子邮件中映射为 Alfresco 中的 ID。

您最多可以在`AccessControlList`字段中添加 200 个条目。

针对 Aurora (MySQL) 数据源的用户上下文筛选

使用 Aurora (MySQL) 数据源时， Amazon Kendra 会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

Aurora (MySQL) 数据库数据源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

筛选 Aurora （PostgreSQL）数据来源的用户上下文

使用 Aurora (PostgreSQL) 数据源时 Amazon Kendra ，会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

Aurora (PostgreSQL) 数据库数据源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

Amazon FSx 数据源的用户上下文筛选

使用 Amazon FSx 数据源时， Amazon Kendra 会从 Amazon FSx 实例的目录服务中获取用户和组信息。

群 Amazon FSx 组和用户 ID 映射如下：

- `_group_ids`-群组 ID 存在 Amazon FSx 于设置了访问权限的文件中。它们是从的目录服务中的系统组名称映射出来的 Amazon FSx。

- `_user_id`—用户 ID 存在 Amazon FSx 于设置了访问权限的文件中。它们是从的目录服务中的系统用户名映射出来的 Amazon FSx。

您最多可以在`AccessControlList`字段中添加 200 个条目。

数据库数据来源的用户上下文筛选

当您使用数据库数据源（例如）时 Amazon Aurora PostgreSQL， Amazon Kendra 会从源表的某一列中获取用户和组信息。您可以在 CreateDataSourceAPI 中将此列指定为AclConfigurationDatabaseConfiguration对象的一部分。

数据库数据来源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

Amazon RDS （ Microsoft SQL Server）数据来源的用户上下文筛选

当你使用 Amazon RDS （Microsoft SQL Server）数据源时， Amazon Kendra 会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为CreateDataSourceAPI 的一部分。

Amazon RDS （微软 SQL Server）数据库数据源有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

针对 Amazon RDS (MySQL) 数据源的用户上下文筛选

使用 Amazon RDS (MySQL) 数据源时， Amazon Kendra 会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

Amazon RDS (MySQL) 数据库数据源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。

- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

筛选 Amazon RDS (Oracle) 数据源的用户上下文

使用 Amazon RDS (Oracle) 数据源时， Amazon Kendra 会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

Amazon RDS (Oracle) 数据库数据源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

筛选 Amazon RDS （ PostgreSQL ）数据来源的用户上下文

使用 Amazon RDS (PostgreSQL) 数据源时 Amazon Kendra ，会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

Amazon RDS (PostgreSQL) 数据库数据源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

Amazon S3 数据来源的用户上下文筛选

您可以使用与文档关联的元数据文件向 Amazon S3 数据源中的文档添加用户上下文筛选。您可以将信息添加到 JSON 文档的`AccessControlList`字段中。有关向从数据来源编制索引的文档中添加元 Amazon S3 数据的更多信息，请参阅 S3 文档元数据。

您提供三条信息：

- 实体应拥有的访问权限。您可以说`ALLOW`或`DENY`。
- 实体的类型。您可以说`USER`或`GROUP`。

- 在 中，将实体命名为 。

您最多可以在AccessControlList字段中添加 200 个条目。

Amazon WorkDocs 数据源的用户上下文筛选

当您使用 Amazon WorkDocs 数据源时， Amazon Kendra 会从该 Amazon WorkDocs 实例获取用户和群组信息。

群 Amazon WorkDocs 组和用户 ID 映射如下：

- _group_ids—群组 ID 存在 Amazon WorkDocs 于设置了访问权限的文件中。它们是根据中组的名称映射出来的 Amazon WorkDocs。
- _user_id—用户 ID 存在 Amazon WorkDocs 于设置了访问权限的文件中。它们是根据中的用户名映射出来的 Amazon WorkDocs。

您最多可以在AccessControlList字段中添加 200 个条目。

Box 数据来源的用户上下文筛选

使用 Box 数据源时， Amazon Kendra 会从 Box 实例获取用户和群组信息。

Box 组和用户 ID 映射如下：

- _group_ids-群组 ID 存在于 Box 中设置了访问权限的文件上。它们是从 Box 中群组的名称映射出来的。
- _user_id-用户 ID 存在于 Box 中设置了访问权限的文件中。它们从用户电子邮件中映射为 Box 中的用户 ID。

您最多可以在AccessControlList字段中添加 200 个条目。

Confluence 数据来源的用户上下文筛选

当您使用 Confluence 数据源时， Amazon Kendra 会从 Confluence 实例获取用户和群组信息。

您可以使用空间权限页面配置用户和群组对空间的访问权限。对于页面和博客，您可以使用限制页面。有关空间权限的更多信息，请参阅 Confluence Support 网站上的空间权限概述。有关页面和博客限制的更多信息，请参阅 Confluence Support 网站上的页面限制。

Confluence 群组和用户名映射如下：

- _group_ids-群组名称会出现在有限制的空间、页面和博客上。它们是根据 Confluence 中的群组名称映射出来的。群组名称始终为小写。

- _user_id- 用户名出现在空间、页面或博客上有限制的地方。它们根据您使用的 Confluence 实例的类型进行映射。

  适用于 Confluence 连接程序 v1.0

  - 服务器-_user_id 是用户名。用户名始终为小写。
  - 云端-_user_id 是用户的账户 ID。

  适用于 Confluence 连接器 v2.0

  - 服务器-_user_id 是用户名。用户名始终为小写。
  - Cloud-_user_id 是用户的电子邮件 ID。

  > **Important**
  >
  > 要使用户上下文筛选功能在 Confluence 连接器上正常运行，您需要确保将获得 Confluence 页面访问权限的用户的可见性设置为“任何人”。有关更多信息，请参阅 Atlassian 开发者文档中的设置电子邮件可见性。

您最多可以在AccessControlList字段中添加 200 个条目。

针对 Dropbox 数据来源的用户上下文筛选

当您使用 Dropbox 数据源时， Amazon Kendra 会从 Dropbox 实例中获取用户和群组信息。

群组和用户 ID 映射如下：

- _group_ids-群组 ID 存在于 Dropbox 中已设置访问权限的文件上。它们是从 Dropbox 中群组的名称映射出来的。
- _user_id-Dropbox 中存在已设置访问权限的文件中的用户 ID。它们从用户的电子邮件中映射为 Dropbox 中的 ID。

您最多可以在AccessControlList字段中添加 200 个条目。

Drupal 数据来源的用户上下文筛选

当你使用 Drupal 数据源时， Amazon Kendra 会从 DrupalInstance 中获取用户和群组信息。

群组和用户 ID 映射如下：

- `_group_ids`- 在 Drupal 中，群组 ID 存在于设置了访问权限的文件上。它们是根据Drupal中群组的名称映射出来的。
- `_user_id`- Drupal 中存在已设置访问权限的文件上的用户 ID。它们从用户电子邮件中映射为 Drupal 中的 ID。

您最多可以在`AccessControlList`字段中添加 200 个条目。

GitHub 数据源的用户上下文筛选

当您使用 GitHub 数据源时， Amazon Kendra 会从该 GitHub 实例获取用户信息。

GitHub 用户 ID 映射如下：

- `_user_id`—用户 ID 存在 GitHub 于设置了访问权限的文件中。它们从用户电子邮件中映射为中的 ID GitHub。

您最多可以在`AccessControlList`字段中添加 200 个条目。

Gmail 数据来源的用户上下文筛选

当你使用 Gmail 数据源时， Amazon Kendra 会从 Gmail 实例中获取用户信息。

用户 ID 映射如下：

- `_user_id`- Gmail 中设置了访问权限的文件中存在用户 ID。它们从用户的电子邮件中映射为 Gmail 中的 ID。

您最多可以在`AccessControlList`字段中添加 200 个条目。

筛选 Google 云端硬盘数据来源的用户上下文

Google Workspace 云端硬盘数据来源会返回谷歌云端硬盘用户和群组的用户和群组信息。组和域成员资格映射到`_group_ids`索引字段。Google 云端硬盘用户名将映射到该`_user_id`字段。

当您在 Query API 中提供一个或多个用户电子邮件地址时，仅返回与这些电子邮件地址共享的文档。以下AttributeFilter参数仅返回与“martha@example.com”共享的文档。

```
"AttributeFilter": {
                "EqualsTo":{
                   "Key": "_user_id",
                   "Value": {
                       "StringValue": "martha@example.com"
                   }
               }
          }
```

如果您在查询中提供一个或多个群组电子邮件地址，则仅返回与群组共享的文档。以下AttributeFilter参数仅返回与“hr@example.com”群组共享的文档。

```
"AttributeFilter": {
                "EqualsTo":{
                   "Key": "_group_ids",
                   "Value": {
                       "StringListValue": ["hr@example.com"]
                   }
               }
          }
```

如果您在查询中提供了域，则会返回与该域共享的所有文档。以下AttributeFilter参数返回与“example.com”域共享的文档。

```
"AttributeFilter": {
                "EqualsTo":{
                   "Key": "_group_ids",
                   "Value": {
                       "StringListValue": ["example.com"]
                   }
               }
          }
```

您最多可以在AccessControlList字段中添加 200 个条目。

筛选 IBM DB2 数据来源的用户上下文

使用 IBM DB2 数据源时， Amazon Kendra 会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

IBM DB2 数据库数据来源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

Jira 数据来源的用户上下文筛选

使用 Jira 数据源时， Amazon Kendra 会从 Jira 实例获取用户和群组信息。

Jira 用户 ID 的映射如下所示：

- `_user_id`-Jira 中存在已设置访问权限的文件上的用户 ID。它们从用户电子邮件中映射为 Jira 中的用户 ID。

您最多可以在`AccessControlList`字段中添加 200 个条目。

Microsoft Exchange 数据来源的用户上下文筛选

当你使用微软 Exchange 数据源时， Amazon Kendra 会从微软 Exchange 实例获取用户信息。

微软 Exchange 用户 ID 映射如下：

- `_user_id`— 用户名存在于微软 Exchange 权限中，用户可以访问某些内容。它们从用户名映射为微软 Exchange 中的 ID。

您最多可以在`AccessControlList`字段中添加 200 个条目。

微软 OneDrive 数据源的用户上下文筛选

Amazon Kendra 在 Microsoft 为站点上的文档编制索引 OneDrive 时，会从 Microsoft 检索用户和群组信息。用户和群组信息取自托管的底层 Microsoft SharePoint 站点 OneDrive。

使用 OneDrive 用户或群组筛选搜索结果时，按如下方式计算 ID：

1. 获取网站名称。例如，`https://host.onmicrosoft.com/sites/siteName`.
2. 以网站名称的 MD5 哈希值为例。例如，`430a6b90503eef95c89295c8999c7981`。
3. 将 MD5 哈希值与竖线 (|) 和 ID 连接起来，创建用户电子邮件或群组 ID。例如，如果群组名称是 localGroupName “”，则群组 ID 将是：

   `"430a6b90503eef95c89295c8999c7981 | localGroupName"`

   > **Note**
   >
   > 在竖线前后各留一个空格。竖条用于标识`localGroupName`其 MD5 哈希值。

   对于用户名“someone@host.onmicrosoft.com”，用户 ID 将如下所示：

   `"430a6b90503eef95c89295c8999c7981 | someone@host.onmicrosoft.com"`

在调用查询 API 时，将用户`_user_id`或群组 ID Amazon Kendra 作为或`_group_id`属性发送到。例如，使用群组筛选搜索结果的 AWS CLI 命令如下所示：

```
aws kendra  query \
                --index-id index ID
                --query-text "query text"
                --attribute-filter '{
                   "EqualsTo":{
                     "Key": "_group_id",
                     "Value": {"StringValue": "430a6b90503eef95c89295c8999c7981 |
 localGroupName"}
                  }}'
```

您最多可以在`AccessControlList`字段中添加 200 个条目。

微软 OneDrive v2.0 数据源的用户上下文筛选

Microsoft OneDrive v2.0 数据源返回来自 OneDrive 访问控制列表 (ACL) 实体的分区和页面信息。Amazon Kendra 使用 OneDrive 租户域连接到 OneDrive 实例，然后可以根据用户或群组对分区和文件名的访问权限筛选搜索结果。

对于标准对象，`_user_id`和`_group_id`的使用方式如下：

- `_user_id`— 你的 Microsoft OneDrive 用户电子邮件 ID 已映射到该`_user_id`字段。

- _group_id— 你的 Microsoft OneDrive 群组电子邮件已映射到该_group_id字段。

您最多可以在AccessControlList字段中添加 200 个条目。

微软 SharePoint 数据源的用户上下文筛选

Amazon Kendra 在 Microsoft 为站点文档编制索引 SharePoint 时，会从 Microsoft 检索用户和群组信息。要根据用户或群组访问权限筛选搜索结果，请在调用 Query API 时提供用户和群组信息。

要使用用户名进行筛选，请使用该用户的电子邮件地址。例如，johnstiles@example.com。

使用 SharePoint 群组筛选搜索结果时，按如下方式计算群组 ID：

对于本地团体

1. 获取网站名称。例如，https://host.onmicrosoft.com/sites/siteName.
2. 取站点名称的 SHA256 哈希。例如，430a6b90503eef95c89295c8999c7981。
3. 通过将 SHA256 哈希值与竖线 (\|) 和群组名称连接起来来创建群组 ID。例如，如果群组名为 localGroupName “”，则群组 ID 将是：

   ```
   "430a6b90503eef95c89295c8999c7981 | localGroupName"
   ```

   > **Note**
   >
   > 在竖线前后各留一个空格。竖条用于标识localGroupName其 SHA256 哈希值。

调用查询 API 时，将群组 ID Amazon Kendra 作为_group_id属性发送到。例如，该 AWS CLI 命令如下所示：

```
aws kendra  query \
                --index-id index ID
                --query-text "query text"
                --attribute-filter '{
                   "EqualsTo":{
                     "Key": "_group_id",
                     "Value": {"StringValue": "430a6b90503eef95c89295c8999c7981 |
 localGroupName"}
                  }}'
```

适用于 AD 组

1. 使用 AD 组 ID 来配置对搜索结果的筛选。

调用查询 API 时，将群组 ID Amazon Kendra 作为_group_id属性发送到。例如，该 AWS CLI 命令如下所示：

```
aws kendra  query \
                --index-id index ID
                --query-text "query text"
                --attribute-filter '{
                   "EqualsTo":{
                     "Key": "_group_id",
                     "Value": {"StringValue": "AD group"}
                  }}'
```

您最多可以在AccessControlList字段中添加 200 个条目。

Microsoft SQL Server 数据来源的用户上下文筛选

当你使用 Microsoft SQL Server 数据源时， Amazon Kendra 会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

Microsoft SQL Server 数据库数据来源有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

Microsoft 团队数据来源的用户上下文筛选

Amazon Kendra 在为文档编制索引时，会从 Microsoft Teams 中检索用户信息。用户信息取自底层的 Microsoft Teams 实例。

您最多可以在AccessControlList字段中添加 200 个条目。

Microsoft Yammer 数据来源的用户上下文筛选

Amazon Kendra 在为文档编制索引时，会从 Microsoft Yammer 中检索用户信息。用户和群组信息取自底层的 Microsoft Yammer 实例。

Microsoft Yammer 用户 ID 映射如下：

- `_email_id`— 映射到该`_user_id`字段的 Microsoft 电子邮件 ID。

您最多可以在`AccessControlList`字段中添加 200 个条目。

针对 MySQL 数据来源的用户上下文筛选

使用 MySQL 数据源时， Amazon Kendra 会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

MySQL 数据库数据来源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

筛选 Oracle 数据库数据来源的用户上下文

使用 Oracle 数据库数据源时， Amazon Kendra 会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

Oracle 数据库数据来源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

PostgreSQL 数据来源的用户上下文筛选

使用 PostgreSQL 数据源时 Amazon Kendra ，会从源表中的一列中获取用户和组信息。您可以在控制台中指定此列，或者使用TemplateConfiguration对象作为 CreateDataSourceAPI 的一部分。

PostgreSQL 数据库数据来源具有以下限制：

- 您只能为数据库数据来源指定允许列表。您无法指定拒绝列表。
- 您只能指定群组。您不能为允许列表指定单个用户。
- 数据库列应是一个包含以分号分隔的组列表的字符串。

Quip 数据来源的用户上下文筛选

当您使用 Quip 数据源时，Amazon Kendra 会从 Quip 实例获取用户信息。

Quip 用户 ID 的映射如下所示：

- `_user_id`-Quip 中存在已设置访问权限的文件上的用户 ID。它们从用户电子邮件中映射为 Quip 中的 ID。

您最多可以在`AccessControlList`字段中添加 200 个条目。

Salesforce 数据来源的用户上下文筛选

Salesforce 数据来源返回来自 Salesforce 访问控制列表（ACL）实体的用户和群组信息。您可以将用户上下文筛选应用于 Salesforce 标准对象和聊天提要。用户上下文筛选不适用于 Salesforce 知识文章。

如果您将任何 Salesforce 字段映射到 Amazon Kendra 文档标题和文档正文字段，Amazon Kendra 将在搜索响应中使用来自文档标题和正文字段的数据。

对于标准对象，`_user_id`和`_group_ids`的使用方式如下：

- `_user_id`- Salesforce 用户的用户名。
- `_group_ids`—
  - Salesforce 的名字 `Profile`
  - Salesforce 的名字 `Group`
  - Salesforce 的名字 `UserRole`
  - Salesforce 的名字 `PermissionSet`

对于 chatter Feed，`_user_id`和`_group_ids`的使用方式如下：

- `_user_id`- Salesforce 用户的用户名。仅当该项目发布在用户的 Feed 中时才可用。
- `_group_ids`-群组 ID 的使用方式如下。仅当 Feed 项目在聊天或协作群组中发布时才可用。

- 聊天群组或协作群组的名称。
- 如果该群组是公开的，PUBLIC:ALL.

您最多可以在AccessControlList字段中添加 200 个条目。

ServiceNow 数据来源的用户上下文筛选

只有 TemplateConfiguration API 和 ServiceNow ServiceNow Connector v2.0 支持用户上下文筛选。ServiceNowConfigurationAPI 和 ServiceNow 连接器 v1.0。不支持用户上下文筛选。

当您使用 ServiceNow 数据源时， Amazon Kendra 会从该 ServiceNow 实例获取用户和群组信息。

群组和用户 ID 映射如下：

- _group_ids—群组 ID 存在 ServiceNow 于设置了访问权限的文件中。它们是根据中的角色名称映射而来sys_ids的 ServiceNow。
- _user_id—用户 ID 存在 ServiceNow 于设置了访问权限的文件中。它们从用户电子邮件中映射为中的 ID ServiceNow。

您最多可以在AccessControlList字段中添加 200 个条目。

Slack 数据来源的用户上下文筛选

当你使用 Slack 数据源时， Amazon Kendra 会从 Slack 实例中获取用户信息。

Slack 用户 ID 的映射如下所示：

- _user_id- Slack 中存在设置访问权限的消息和频道上的用户 ID。它们从用户电子邮件中映射为 Slack 中的 ID。

您最多可以在AccessControlList字段中添加 200 个条目。

Zendesk 数据来源的用户上下文筛选

当你使用 Zendesk 数据源时， Amazon Kendra 会从 Zendesk 实例中获取用户和群组信息。

群组和用户 ID 映射如下：

- _group_ids-群组 ID 存在于设置访问权限的 Zendesk 工单和文章中。它们是从 Zendesk 中的群组名称映射出来的。

- _user_id-群组 ID 存在于设置访问权限的 Zendesk 工单和文章中。它们从用户电子邮件中映射为 Zendesk 中的 ID。

您最多可以在AccessControlList字段中添加 200 个条目。

查询响应和响应类型

Amazon Kendra 支持不同的查询响应和响应类型。

查询响应

调用 Query API 会返回有关搜索结果的信息。结果以QueryResultItem对象数组 (ResultItems) 的形式出现。每QueryResultItem份都包含结果摘要。包括与查询结果关联的文档属性。

摘要信息

摘要信息因结果类型而异。在每种情况下，它都包含与搜索词匹配的文档文本。它还包括突出显示信息，可用于突出显示应用程序输出中的搜索文本。例如，如果搜索词是太空针塔的高度是多少？ ，摘要信息包括高度和太空针塔这两个词的文本位置。有关响应卡的信息，请参阅 查询响应和响应类型。

文档属性

每个结果都包含与查询相匹配的文档的文档属性。有些属性是预定义的，例如DocumentIdDocumentTitle、和DocumentUri。其他属性是您定义的自定义属性。您可以使用文档属性来筛选来自 Query API 的响应。例如，您可能只需要由特定作者撰写的文档或文档的特定版本。有关更多信息，请参阅 筛选和分面搜索。在向索引中添加文档时，可以指定文档属性。有关更多信息，请参阅自定义字段和属性。

以下是查询结果的 JSON 代码示例。请注意DocumentAttributes和中的文档属性AdditionalAttributes。

```
{
    "QueryId": "query-id",
    "ResultItems": [
        {
            "Id": "result-id",
            "Type": "ANSWER",
            "AdditionalAttributes": [
                {
                    "Key": "AnswerText",
                    "ValueType": "TEXT_WITH_HIGHLIGHTS_VALUE",
```

```
                "Value": {
                    "TextWithHighlightsValue": {
                        "Text": "text",
                        "Highlights": [
                            {
                                "BeginOffset": 55,
                                "EndOffset": 90,
                                "TopAnswer": false
                            }
                        ]
                    }
                }
            }
        ],
        "DocumentId": "document-id",
        "DocumentTitle": {
            "Text": "title"
        },
        "DocumentExcerpt": {
            "Text": "text",
            "Highlights": [
                {
                    "BeginOffset": 0,
                    "EndOffset": 300,
                    "TopAnswer": false
                }
            ]
        },
        "DocumentURI": "uri",
        "DocumentAttributes": [],
        "ScoreAttributes": "score",
        "FeedbackToken": "token"
    },
    {
        "Id": "result-id",
        "Type": "ANSWER",
        "Format": "TABLE",
        "DocumentId": "document-id",
        "DocumentTitle": {
            "Text": "title"
        },
        "TableExcerpt": {
            "Rows": [{
                "Cells": [{
```

```
                    "Header": true,
                    "Highlighted": false,
                    "TopAnswer": false,
                    "Value": "value"
                }, {
                    "Header": true,
                    "Highlighted": false,
                    "TopAnswer": false,
                    "Value": "value"
                }, {
                    "Header": true,
                    "Highlighted": false,
                    "TopAnswer": false,
                    "Value": "value"
                }, {
                    "Header": true,
                    "Highlighted": false,
                    "TopAnswer": false,
                    "Value": "value"
                }]
            }, {
                "Cells": [{
                    "Header": false,
                    "Highlighted": false,
                    "TopAnswer": false,
                    "Value": "value"
                }, {
                    "Header": false,
                    "Highlighted": false,
                    "TopAnswer": false,
                    "Value": "value"
                }, {
                    "Header": false,
                    "Highlighted": true,
                    "TopAnswer": true,
                    "Value": "value"
                }, {
                    "Header": false,
                    "Highlighted": false,
                    "TopAnswer": false,
                    "Value": "value"
                ]}
              }],
                "TotalNumberofRows": number
```

```
    },
            "DocumentURI": "uri",
            "ScoreAttributes": "score",
            "FeedbackToken": "token"
        },
        {
            "Id": "result-id",
            "Type": "DOCUMENT",
            "AdditionalAttributes": [],
            "DocumentId": "document-id",
            "DocumentTitle": {
                "Text": "title",
                "Highlights": []
            },
            "DocumentExcerpt": {
                "Text": "text",
                "Highlights": [
                    {
                        "BeginOffset": 74,
                        "EndOffset": 77,
                        "TopAnswer": false
                    }
                ]
            },
            "DocumentURI": "uri",
            "DocumentAttributes": [
                {
                    "Key": "_source_uri",
                    "Value": {
                        "StringValue": "uri"
                    }
                }
            ],
            "ScoreAttributes": "score",
            "FeedbackToken": "token",
        }
    ],
    "FacetResults": [],
    "TotalNumberOfResults": number
}
```

响应时间

Amazon Kendra 返回三种类型的查询响应。

- 答案（包括表格答案）
- 文档
- 问答

响应的类型将在QueryResultItem对象的Type响应字段中返回。

回答

Amazon Kendra 在响应中检测到一个或多个问题答案。事实是对谁、什么、何时或何地问题的回应，例如离我最近的服务中心在哪里？ Amazon Kendra 返回索引中与查询最匹配的文本。文本位于AnswerText字段中，并在响应文本中包含搜索词的突出显示信息。 AnswerText包括带有突出显示文本的完整文档摘录，同时DocumentExcerpt包括带有突出显示文本的截断（290 个字符）文档摘录。

Amazon Kendra 每个文档只返回一个答案，这是可信度最高的答案。要从一个文档返回多个答案，必须将该文档拆分为多个文档。

```
{
    'AnswerText': {
        'TextWithHighlights': [
            {
                'BeginOffset': 271,
                'EndOffset': 279,
                'TopAnswer': False
            },
            {
                'BeginOffset': 481,
                'EndOffset': 489,
                'TopAnswer': False
            },
            {
                'BeginOffset': 547,
                'EndOffset': 555,
                'TopAnswer': False
            },
            {
```

```
                'BeginOffset': 764,
                'EndOffset': 772,
                'TopAnswer': False
            }
        ],
        'Text': 'Asynchronousoperationscan\n''alsoprocess
\n''documentsthatareinPDF''format.UsingPDFformatfilesallowsyoutoprocess''multi-
page\n''documents.\n''Forinformationabouthow''AmazonTextractrepresents
\n''documentsasBlockobjects,
        ''seeDocumentsandBlockObjects.
\n''\n''\n''\n''Forinformationaboutdocument''limits,
        seeLimitsinAmazonTextract.
\n''\n''\n''\n''TheAmazonTextractsynchronous''operationscanprocessdocumentsstoredinanAmazon
\n''S3Bucketoryoucanpass''base64encodedimagebytes.\n''Formoreinformation,

 see''CallingAmazonTextractSynchronousOperations.''Asynchronousoperationsrequireinputdocuments
\n''tobesuppliedinanAmazon''S3Bucket.'
    },
    'DocumentExcerpt': {
        'Highlights': [
            {
                'BeginOffset': 0,
                'EndOffset': 300,
                'TopAnswer': False
            }
        ],
        'Text': 'Asynchronousoperationscan\n''alsoprocess
\n''documentsthatareinPDF''format.UsingPDFformatfilesallowsyoutoprocess''multi-page
\n''documents.\n''ForinformationabouthowAmazon''Textractrepresents\n'''
    },
    'Type': 'ANSWER'
}
```

文档

Amazon Kendra 返回与搜索词匹配的文档的排名文档。排名基于对搜索结果准确性的 Amazon Kendra 置信度。有关匹配文档的信息将在中返回QueryResultItem。它包括文档标题。摘录包括搜索文本的突出显示信息以及文档中匹配文本的部分。匹配文档的 URI 位于SourceURI文档属性中。以下示例 JSON 显示了匹配文档的文档摘要。

```
{
    'DocumentTitle': {
        'Highlights': [
```

```
            {
                'BeginOffset': 7,
                'EndOffset': 15,
                'TopAnswer': False
            },
            {
                'BeginOffset': 97,
                'EndOffset': 105,
                'TopAnswer': False
            }
        ],
        'Text': 'AmazonTextractAPIPermissions: Actions,
        \n''Permissions,
        andResourcesReference-''AmazonTextract'
    },
    'DocumentExcerpt': {
        'Highlights': [
            {
                'BeginOffset': 68,
                'EndOffset': 76,
                'TopAnswer': False
            },
            {
                'BeginOffset': 121,
                'EndOffset': 129,
                'TopAnswer': False
            }
        ],
        'Text': '...LoggingandMonitoring\tMonitoring
\n''\tCloudWatchMetricsforAmazonTextract
\n''\tLoggingAmazonTextractAPICallswithAWSCloudTrail\n''\tAPIReference\tActions
\tAnalyzeDocument\n''\tDetectDocumentText\n''\tGetDocumentAnalysis...'
    },
    'Type': 'DOCUMENT'
}
```

问答

当一个问题与索引中的一个常见问题 Amazon Kendra 匹配时，将返回问答响应。答案包括QueryResultItem字段中匹配的问题和答案。它还包括查询字符串中检测到的查询词的突出显示信息。以下 JSON 显示了问题和答案的回答。请注意，答案中包含问题文本。

```
{
```

```
    'AnswerText': {
        'TextWithHighlights': [

        ],
        'Text': '605feet'
    },
    'DocumentExcerpt': {
        'Highlights': [
            {
                'BeginOffset': 0,
                'EndOffset': 8,
                'TopAnswer': False
            }
        ],
        'Text': '605feet'
    },
    'Type': 'QUESTION_ANSWER',
    'QuestionText': {
        'Highlights': [
            {
                'BeginOffset': 12,
                'EndOffset': 18,
                'TopAnswer': False
            },
            {
                'BeginOffset': 26,
                'EndOffset': 31,
                'TopAnswer': False
            },
            {
                'BeginOffset': 32,
                'EndOffset': 38,
                'TopAnswer': False
            }
        ],
        'Text': 'whatistheheightoftheSpaceNeedle?'
    }
}
```

有关向索引添加问题和答案文本的信息，请参阅创建常见问题解答。

调整和排序响应

您可以通过相关性调整来修改字段或属性对搜索相关性的影响。您也可以按特定属性或字段对搜索结果进行排序。

主题

优化响应

您可以通过相关性调整来修改字段或属性对搜索相关性的影响。要快速测试相关性调整，请使用查询 API 在查询中传递调整配置。然后，您可以看到从不同配置中获得的不同搜索结果。控制台不支持在查询级别进行相关性调整。您也可以仅`StringList`在索引级别调整该类型的字段或属性。有关更多信息，请参阅调整搜索相关性。

默认情况下，查询响应按响应中每个结果的 Amazon Kendra 相关性分数排序。

您可以调整以下类型的任何内置或自定义属性/字段的结果：

- 日期值
- 长型值。
- 字符串值

您无法对以下类型的属性进行排序：

- 字符串列表值

对文档结果进行排名和调整 (AWS SDK)

将`Searchable`参数设置为 true 以增强文档元数据配置。

要调整查询中的属性，请设置 Query API 的`DocumentRelevanceOverrideConfigurations`参数并指定要调整的属性的名称。

以下 JSON 示例显示了一个`DocumentRelevanceOverrideConfigurations`对象，该对象会覆盖对索引中名为“部门”的属性的调整。

```
"DocumentRelevanceOverrideConfigurations" : [
    "Name": "department",
    "Relevance": {
        "Importance": 1,
        "ValueImportanceMap": {
            "IT": 3,
            "HR": 7
        }
    }
]
```

对响应进行排序

Amazon Kendra 使用排序属性或字段作为查询返回的文档的标准的一部分。例如，按“_created_at”排序的查询返回的结果可能与按“_version”排序的查询所包含的结果不一样。

默认情况下，查询响应按响应中每个结果的 Amazon Kendra 相关性分数排序。要更改排序顺序，请将文档属性设置为可排序，然后配置 Amazon Kendra 为使用该属性对响应进行排序。

您可以对以下类型的任何内置或自定义属性/字段的结果进行排序：

- 日期值
- 长型值。
- 字符串值

您无法对以下类型的属性进行排序：

- 字符串列表值

您可以在每个查询中根据一个或多个文档属性进行排序。查询返回 100 个结果。如果设置了排序属性的文档少于 100 个，则在结果末尾返回没有排序属性值的文档，按与查询的相关性进行排序。

对文档结果进行排序 (AWS SDK)

1. 要使用 UpdateIndexAPI 使属性可排序，请将`Sortable`参数设置为。`true`以下 JSON 示例用于`DocumentMetadataConfigurationUpdates`向索引添加名为“部门”的属性并使其可排序。

```
"DocumentMetadataConfigurationUpdates": [
   {
```

```
        "Name": "Department",
        "Type": "STRING_VALUE",
        "Search": {
            "Sortable": "true"
        }
    }
]
```

2. 要在查询中使用一个可排序的属性，请设置查询 API 的SortingConfiguration参数。指定要排序的属性的名称以及是按升序还是降序对响应进行排序。

   以下 JSON 示例显示了您用来按“部门”属性按升序对查询结果进行排序的SortingConfiguration参数。

```
   "SortingConfiguration": {
      "DocumentAttributeKey": "Department",
      "SortOrder": "ASC"
   }
```

3. 要在查询中使用多个可排序属性，请设置查询 API 的SortingConfigurations参数。您最多可以设置 3 个对结果 Amazon Kendra 进行排序的字段。您还可以指定结果是按升序还是降序排序。排序字段配额可以增加。

   如果您不提供排序配置，则按 Amazon Kendra 决定结果的相关性对结果进行排序。如果结果排序有平局，则按相关性对结果进行排序。

   以下 JSON 示例显示了您用来按属性“名称”和“价格”升序对查询结果进行排序的SortingConfigurations参数。

```
"CollapseConfiguration" : {
    "DocumentAttributeKey": "Name",
    "SortingConfigurations": [
        {
            "DocumentAttributeKey": "Price",
            "SortOrder": "ASC"
        }
    ],
    "MissingAttributeKeyStrategy": "IGNORE"
}
```

对文档结果排序（控制台）

> **Note**
>
> 目前不支持多属性排序。 AWS Management Console

1. 要使属性可在控制台中排序，请在属性定义中选择“可排序”。可以在创建属性时对属性进行排序，也可以稍后对其进行修改。
2. 要在控制台中对查询响应进行排序，请从“排序”菜单中选择要对响应进行排序的属性。只有在数据来源配置期间标记为可排序的属性才会出现在列表中。

折叠/展开查询结果

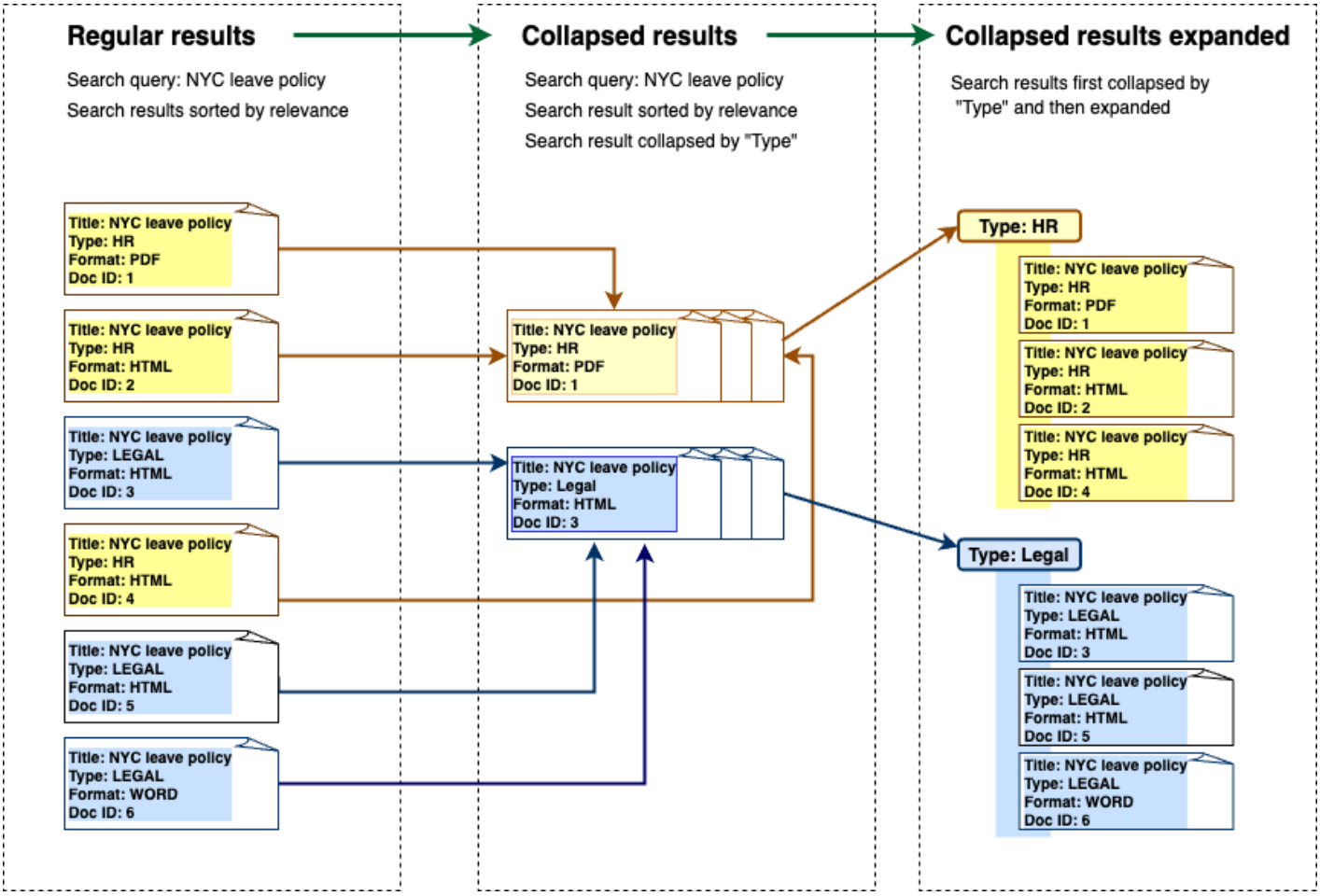
当您 Amazon Kendra 连接到数据时，它会抓取文档元数据属性（例如`_document_title`、`_created_at`、和`_document_id`），并在查询期间使用这些属性或字段提供高级搜索功能。

Amazon Kendra的“折叠和展开查询结果”功能允许您使用常用文档属性对搜索结果进行分组，并在指定的主文档下显示搜索结果（折叠或部分展开）。

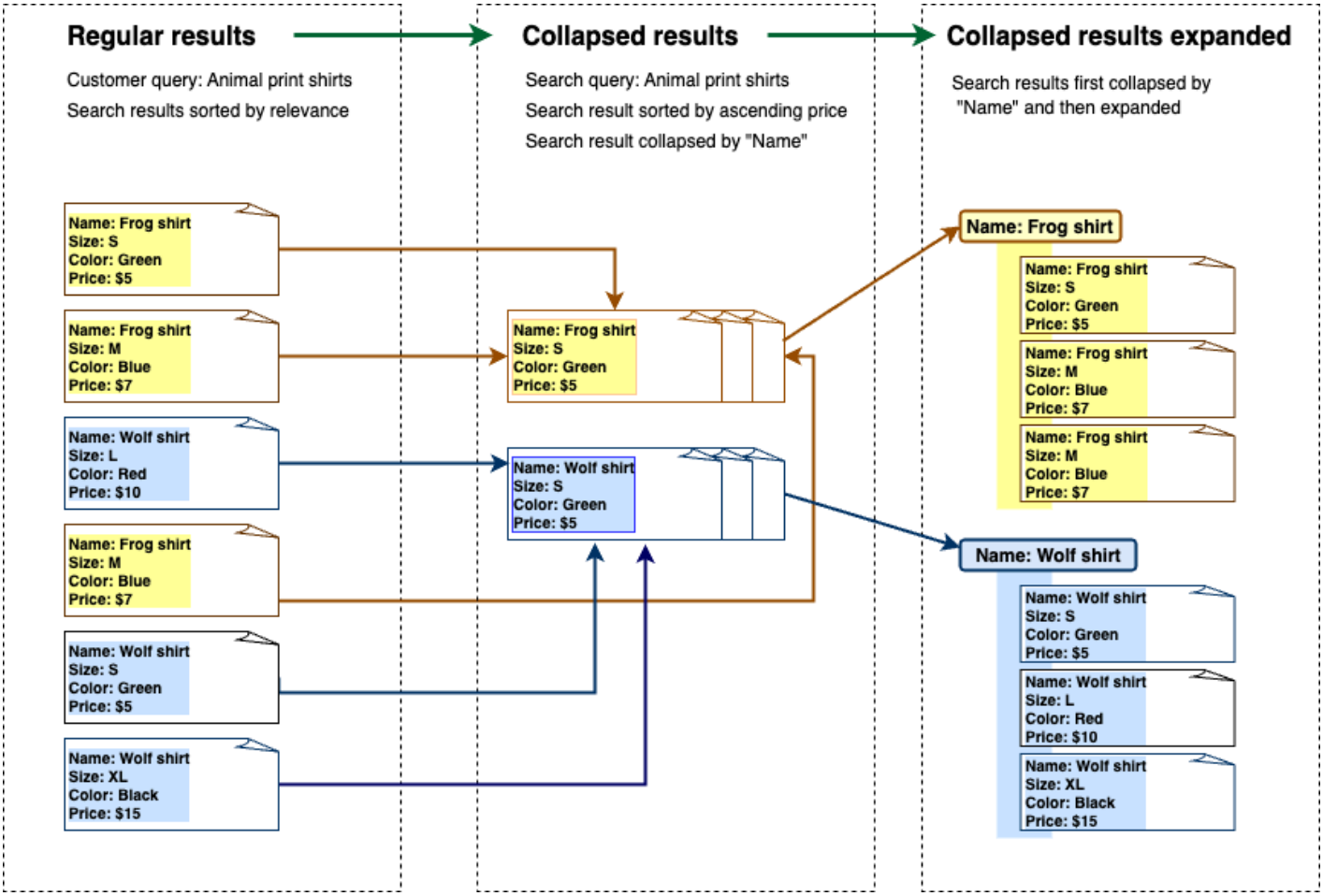
> **Note**
>
> 折叠和展开查询结果功能目前只能通过 Amazon Kendra API 使用。

这在以下几种搜索情况下很有用：

- 索引中的文档中存在多个版本的内容。当您的最终用户查询索引时，您希望他们看到包含隐藏/折叠重复项的最相关的文档版本。例如，如果您的索引包含名为 “NYC 休假政策” 的文档的多个版本，则可以选择使用 “类型” 属性/字段折叠特定群组 “HR” 和 “Legal” 的文档。



- 您的索引包含多个文档，其中包含有关一种商品或对象（例如产品库存）的唯一信息。为了方便地捕获和排序项目信息，您希望最终用户能够将项目或对象链接的所有文档作为一个搜索结果进行访问。在以下示例中，买家搜索 “动物印花衬衫” 会返回按名称分组并按价格升序排序的结果。



折叠结果

要将相似或相关的文档组合在一起，必须指定要折叠的属性（例如，您可以折叠/分组文档）。`_category`为此，请调用 Query API 并使用CollapseConfiguration对象指定`DocumentAttributeKey`要折叠的。`DocumentAttributeKey`控制哪些字段的搜索结果将处于折叠状态。支持的属性键字段包括`String`和`Number`。 `String list`不支持和`Date`类型。

使用排序顺序选择主文档

要将主文档配置为在折叠的组中显示，请使用下面的`SortingConfigurations`参数CollapseConfiguration。例如，要获取文档的最新版本，您可以按以下顺序对每个折叠的组进行排序`_version`。您最多可以指定 3 个要作为排序依据的属性/字段以及每个属性/字段的排序顺序。`SortingConfigurations`您可以请求增加排序属性的数量的限额。

默认情况下， Amazon Kendra 按其为响应中的每个结果确定的相关性分数对查询响应进行排序。要更改默认排序顺序，请将文档属性设置为可排序，然后配置 Amazon Kendra 为使用这些属性对响应进行排序。有关更多信息，请参阅 响应。

缺少文档密钥策略

如果您的文档没有折叠属性值，则会 Amazon Kendra 提供三个自定义选项：

- 选择一个组中包含空值或缺失值COLLAPSE的所有文档。这是默认配置。
- 选择值为空或缺失值的IGNORE文档。被忽略的文档不会出现在查询结果中。
- 将EXPAND每个具有空值或缺失值的文档选择为一组。

扩大结果

您可以使用CollapseConfiguration对象中的Expand参数选择是否展开折叠的搜索结果组。展开的结果与为该组选择主文档时使用的排序顺序相同。

要配置要展开的折叠搜索结果组的数量，请使用ExpandConfiguration对象中的MaxResultItemstoExpand参数。例如，如果将此值设置为 10，则只有 100 个结果组中的前 10 个具有扩展功能。

要配置每个折叠的主文档显示的展开结果数，请使用MaxExpandResultsPerItem参数。例如，如果您将此值设置为 3，则每个折叠的组最多会显示 3 个结果。

与其他 Amazon Kendra 功能的互动

- 折叠和展开结果不会改变分面的数量，也不会影响显示的结果总数。
- Amazon Kendra 即使精选搜索结果的字段值与您配置的折叠字段相同，也不会折叠。
- 折叠和展开结果仅适用于该类型的DOCUMENT结果。

调整搜索相关性

Amazon Kendra 查询生成按相关性排名的搜索结果。索引中的可搜索字段或属性都对排名有影响。

您可以通过相关性调整来修改字段或属性对搜索相关性的影响。要调整搜索相关性，您可以在索引级别手动调整，也可以在索引级别为索引设置调整配置，还可以在查询级别覆盖在索引级别设置的配置，从而进行调整。

当使用相关性调整时，如果查询包含与字段或属性匹配的字词，则给出结果的响应速度将得到提升。您还可以指定有匹配时，文档将获得多少提升。相关性调整不会 Amazon Kendra 导致在查询响应中包含文档，它只是 Amazon Kendra 用于确定文档相关性的因素之一。

您可以提升索引中的特定字段或属性，从而为特定响应分配更高的重要性。例如，如果有人搜索“re:Invent 是在何时？” 你可以提高该_last_update_at领域文档新鲜度的相关性。或者，在研究报告索引中，您可以提升“source”字段中的特定数据来源。

您还可以根据投票或查看次数来提升文档，这在论坛和其他支持知识库中很常见。例如，您可以组合提升，从而提高查看次数和最近查看次数较多的文档。

您可以使用 Importance 参数设置文档获得的提升量。Importance 越高，字段或属性越能提升文档的相关性。当您调整索引或在查询级别调整时，请以较小的增量来增加 Importance 参数的值，直到实现所需的效果。要确定您是否改善了搜索结果，请执行搜索并将结果与之前的查询进行比较。

您可以指定日期、数字或字符串属性来调整索引或在查询级别进行调整。StringList 类型的字段或属性只能在索引级别进行调整。对于何时提升结果，每个字段或属性都有特定的条件。

- 日期字段或属性 - 日期字段有三个特定的条件：Freshness、Duration 和 RankOrder。
  - Duration 设置应用提升的时间段。例如，如果您将时间段设置为 86400 秒（即一天），则提升在一天后会开始减弱。重要性越高，效果下降的速度越快。
  - Freshness 确定将文档应用于字段或属性时的新鲜程度。如果您将 Freshness 应用到创建日期或上次更新日期的字段，则认为最近创建或上次更新的文档比旧文档“更新鲜”。例如，如果文档 1 是在 11 月 14 日创建的，而文档 2 是在 11 月 5 日创建的，则文档 1 比文档 2 的新鲜度“更高”。另外，如果文档 1 上次更新是在 11 月 14 日，而文档 2 上次更新是在 11 月 20 日，则文档 2 比文档 1 的新鲜度“更高”。文档越新鲜，应用这种提升的次数就越多。索引中只能有一个 Freshness 字段。
  - RankOrder 按升序或降序应用提升。如果指定 ASCENDING，则更晚的日期优先。如果指定 DESCENDING，则更早的日期优先。

- 数值字段或属性-对于数值字段或属性，您可以指定在确定字段或属性的相关性时 Amazon Kendra 应使用的等级顺序。如果指定 ASCENDING，则较大的数字优先。如果指定 DESCENDING，则较小的数字优先。
- 字符串字段或属性 - 对于字符串字段或属性，您可以创建字段的类别，以便为每个类别分配不同的提升。例如，如果您提升名为“Department”的字段或属性，则可以对“HR”中的文档和“Legal”中的文档分配不同的提升。您可以提升类型 String 的字段或属性。只能在索引级别提升 StringList 字段。

在索引级别进行相关性调整

您可以使用控制台来调整索引详细信息或 UpdateIndexAPI，从而在索引级别调整字段或属性的相关性。

以下示例将该_last_updated_at字段设置为文档的Freshness字段。

```
"DocumentMetadataConfigurationUpdates" : [
    {
        "Name": "_last_updated_at",
        "Type": "DATE_VALUE",
        "Relevance": {
            "Freshness": TRUE,
            "Importance": 2
        }
    }
]
```

以下示例对“Department”字段中的不同类别应用不同的重要性。

```
"DocumentMetadataConfigurationUpdates" : [
    {
        "Name": "department",
        "Type": "STRING_VALUE",
        "Relevance": {
            "Importance": 2,
            "ValueImportanceMap": {
                "HR": 3,
                "Legal": 1
            }
        }
    }
```

```
]
```

在查询级别进行相关性调整

您可以使用查询 API 在查询级别调整字段或属性的相关性。

控制台不支持在查询级别进行相关性调整。

在查询级别进行调整可加快相关性调整的测试过程，因为您无需为每个测试手动更新索引中的调整配置。您可以通过在查询中传递调整配置来调整文档的相关性。然后，您可以查看从不同的配置获得的不同结果。在查询中传递的配置会覆盖在索引级别设置的配置。

以下示例覆盖了应用于“Department”字段和在索引级别设置的每个部门类别的重要性，如上面的示例所示。当用户输入搜索查询时，“Department”字段的重要性相当高，而 Legal 部门的重要性高于 HR 部门。

```
"DocumentRelevanceOverrideConfigurations" : [
    {
        "Name": "department",
        "Type": "STRING_VALUE",
        "Relevance": {
            "Importance": 2,
            "ValueImportanceMap": {
                "HR": 2,
                "Legal": 8
            }
        }
    }
]
```

通过搜索分析获得见解

您可以使用 se Amazon Kendra arch Analytics 来深入了解您的搜索应用程序如何成功或失败地帮助用户查找信息。

Amazon Kendra Analytics 提供了用户与搜索应用程序的交互方式以及搜索应用程序配置的有效性的快照。您可以使用 GetSnapshotsAPI 或在控制台的导航面板上选择 Analytics 来查看指标数据。

您可以使用 GetSnapshots 在自己的自定义构建控制面板上显示生成的数据。或者，您可以使用控制台中提供的指标控制面板，其中包含可视化图表。借助可视化控制面板，您可以查看一段时间内用户行为的趋势或模式，或者发现搜索应用程序配置所存在的问题。例如，显示每天查询数量稳定且稳步增长的折线图可能表明采用率和使用率有所提高。另一方面，突然下降可能表明存在必须调查的问题。

您可以使用这些指标在不同的数据点之间建立联系，以解决用户如何查询信息或发现商机的问题。例如，文档“人工智能的工作原理是什么？” 是搜索结果中点击次数最多的文档，搜索次数最多的查询是“机器学习的工作原理是什么”。这会可以让您了解用户首选的术语和语言。您可以将这些术语编写到您的文档中，也可以为这些术语使用自定义同义词，以便让用户更容易搜索您的文档。

搜索指标

有 10 个指标可用于分析您的搜索应用程序的性能或用户正在搜索的信息。要检索指标数据，请在调用 GetSnapshots 时指定要检索的指标数据的字符串名称。

您还必须提供查看指标数据的时间间隔或时段。时间间隔使用索引的时区。您可以查看以下时段内的数据：

- THIS_WEEK：本周，从星期日开始到当日的前一天结束。
- ONE_WEEK_AGO：上周，从星期日开始到下一个星期六结束。
- TWO_WEEKS_AGO：上上周，从星期日开始到下一个星期六结束。
- THIS_MONTH：本月，从当月开始第一天开始到当日的前一天结束。
- ONE_MONTH_AGO：上月，从当月的第一天开始到当月最后一天结束。
- TWO_MONTHS_AGO：上上月，从当月的第一天开始，到当月最后一天结束。

在控制台中，支持的时段包括本周、上月、本月、上月。

点击率

搜索结果中产生文档点击的查询比例。这可以让您了解搜索应用程序配置是否可以帮助用户找到与其查询相关的信息。对于返回即时答案的查询，用户可能无需点击文档即可获得更多信息。有关更多信息，请参阅the section called “即时回答率”。您必须致电SubmitFeedback以确保收集到点击反馈。

要使用 GetSnapshots API 检索关于点击率的数据，请将 metricType 指定为 AGG_QUERY_DOC_METRICS。在控制台中选择导航面板上的 Analytics，您也可以查看该指标。

点击次数为零

在搜索结果中点击次数为零的查询比例。这可以帮助您了解内容有哪些差距，为何会提供相关性不大的搜索结果。对于返回即时答案的查询，用户可能无需点击文档即可获得更多信息。有关更多信息，请参阅the section called “即时回答率”。此外，您的搜索设置（例如，调整配置）可能会影响搜索结果中文档的返回方式。

要使用 GetSnapshots API 检索点击率为零的数据，请将 metricType 指定为 AGG_QUERY_DOC_METRICS。在控制台中选择导航面板上的 Analytics，您也可以查看该指标。

搜索结果率为零

导致搜索结果率为零的查询比例。这可以帮助您了解内容有哪些差距，为何会提供不相关的搜索结果。

要使用 GetSnapshots API 检索搜索结果率为零的数据，请将 metricType 指定为 AGG_QUERY_DOC_METRICS。在控制台中选择导航面板上的 Analytics，您也可以查看该指标。

即时回答率

返回具有即时回答或常见问题解答的查询的比例。这可以帮助您了解即时回答在提供信息方面的作用。

要使用 GetSnapshots API 检索即时回答率数据，请将 metricType 指定为 AGG_QUERY_DOC_METRICS。在控制台中选择导航面板上的 Analytics，您也可以查看该指标。

主要查询

您的用户最常搜索的前 100 个查询。这可以帮助您了解哪些查询很受欢迎，以及您的用户最感兴趣的信息类型。

指标包括搜索查询的次数、点击文档的比例、未点击文档的比例、查询搜索结果中的平均点击深度、查询的即时回答比例以及查询前 10 个搜索结果的平均置信度。

要使用 GetSnapshots API 检索有关主要查询的数据，请将 metricType 指定为 QUERIES_BY_COUNT。在控制台中选择控制台导航面板上的 Analytics，然后选择查询列表下的主要查询，您也可以查看该指标。

点击次数为零的主要查询

在搜索结果中点击次数为零的前 100 个查询。这可以帮助您了解内容有哪些差距，即缺少与某些查询相关的文档，或者您的搜索应用程序配置返回的搜索结果不相关。对于返回即时答案的查询，用户可能无需点击文档即可获得更多信息。有关更多信息，请参阅the section called “即时回答率”。

指标包括产生的点击次数为零的查询数量、点击次数为零的查询比例、即时回答的查询比例以及产生前 10 个搜索结果的查询的平均置信度。

要使用 GetSnapshots API 检索有关点击次数为零的主要查询的数据，请将 metricType 指定为 QUERIES_BY_ZERO_CLICK_RATE。在控制台中选择控制台导航面板上的 Analytics，然后选择查询列表下的点击次数为零的主要查询，您也可以查看该指标。

搜索结果为零的主要查询

在搜索结果为零的前 100 个查询。这可以帮助您了解内容有哪些差距，即没有与某些查询相关的文档。或者，您的用户可能会使用专业术语进行查询，而可能导致没有搜索结果，这表明您需要创建自定义同义词来解决问题。

指标包括导致搜索结果为零的查询数量，搜索结果为零的查询比例，以及该查询的搜索数量相对于所有查询的比例。

要使用 GetSnapshots API 检索有关搜索结果为零的主要查询的数据，请将 metricType 指定为 QUERIES_BY_ZERO_RESULT_RATE。在控制台中选择控制台导航面板上的 Analytics，然后选择查询列表下的结果为零的主要查询，您也可以查看该指标。

点击次数最多的文档

搜索结果中点击次数最多的前 100 个文档。这可以帮助您了解在用户查询信息时，哪些文档或搜索结果与他们最相关。

指标包括点击文档的次数、文档获得用户点喜欢的次数（竖起大拇指）、文档获得用户点不喜欢的次数（倒竖大拇指）。

要使用 GetSnapshots API 检索点击次数最多的文档的数据，请将 metricType 指定为 DOCS_BY_CLICK_COUNT。在控制台中选择控制台导航面板上的 Analytics，然后选择查询列表下的点击次数最多的文档，您也可以查看该指标。

查询总数

您的用户搜索的查询总数。这可以帮助您了解用户对搜索应用程序的参与程度。

要使用 GetSnapshots API 检索有关查询总数的数据，请将 metricType 指定为 AGG_QUERY_DOC_METRICS。在控制台中选择导航面板上的 Analytics，您也可以查看该指标。

文档总数

您的索引中的文档总数。这可以帮助您将索引大小与查询总数进行比较，从而检查文档数量是否有适合查询量。

要使用 GetSnapshots API 检索有关文档总数的数据，请将 metricType 指定为 AGG_QUERY_DOC_METRICS。在控制台中选择导航面板上的 Analytics，您也可以查看该指标。

检索指标数据的示例

以下代码是有关上月主要查询的检索数据的示例。

Console

检索上月的主要查询

1. 在左侧导航窗格的索引下，选择您的索引，然后选择 Analytics。
2. 在 Analytics 页面上，选择本周按钮，将检索数据的时段更改为上月。
3. 在 Analytics 页面的查询列表下，选择主要查询。

CLI

检索上月的主要查询

```
aws kendra get-snapshots \
--index-id index-id \
--interval "ONE_MONTH_AGO" \
--metric-type "QUERIES_BY_COUNT"
```

Python

检索上月的主要查询

```
import boto3

kendra = boto3.client("kendra")

index_id = "index-id"
interval = "ONE_MONTH_AGO"
metric_type = "QUERIES_BY_COUNT"

snapshots_response = kendra.get_snapshots(
    IndexId = index_id,
    Interval = interval,
    MetricType = metric_type
)

print("Top queries data: " + snapshots_response["snapshotsData"])
```

Java

检索上月的主要查询

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.GetSnapshotsRequest;
import software.amazon.awssdk.services.kendra.model.GetSnapshotsResponse;

public class TopQueriesExample {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();

        String indexId = "indexID";
        String interval = "ONE_MONTH_AGO";
        String metricType = "QUERIES_BY_COUNT";

        GetSnapshotsRequest getSnapshotsRequest = GetSnapshotsRequest
            .builder()
            .indexId(indexId)
            .interval(interval)
            .metricType(metricType)
            .build();

        GetSnapshotsResponse getSnapshotsResponse =
 kendra.getSnapshots(GetSnapshotsRequest);
```

```
        System.out.println(String.format("Top queries data: ",
getSnapshotsResponse.snapshotsData()))
```

从指标到可行见解

可行见解是从原始数据中提取的有意义的信息，用于指导您的行动或决策。要从指标中提取含义并用来获得可行的见解，重要的是不仅要分析每个指标，还要在指标之间建立联系。

例如，点击次数为零的主要查询是“当前有哪些区域可用？” 但是，它的即时回答率为 100%。这表明您的用户无需点击搜索结果或点击提供可用区域信息的文档即可获得该问题的答案。如果您只看到点击次数为零，就无法了解全貌，对于搜索应用程序配置能够成功处理该查询，您也可能得出错误的结论。

另一个可行见解的例子是发现商机。通常，企业通过分析搜索指标来寻找获客机会。点击次数最多的文档是“可用区域”。此外，大多数搜索的主要查询都与大洋洲区域的产品供应问题相关，答案中的即时回答率为 100%，有关可用区域的更多信息的点击率很高。这表明该区域对您的产品或服务有兴趣和需求。

可视化和报告搜索分析

有五个指标，包括趋势数据，可供您可视化和查看一段时间内的趋势或模式。如果您使用控制台，则会提供趋势数据图表。如果您使用 API，则可以检索趋势数据来创建自己的图表或可视化结果。控制台中的大多数图表都绘制了所选时段内的每日数据点。

控制台提供了一个指标控制面板，您可以在其中选择要查看的图表和排名。在 Analytics 主页上选择导出，将控制面板上显示的指标以 CSV 格式导出。您可以将这些报告包含在您的业务文档或演示文稿中。

您可以可视化以下指标：

查询总数图表

每天查询数量折线图。该图表可帮助您可视化每日用户参与模式。一些例子包括用户参与度稳步增加或下降，或者由于搜索应用程序崩溃或网站问题而导致查询量急剧下降至 0。

如果您使用 API，则可以通过指定 TREND_QUERY_DOC_METRICS 来检索这些数据。您可以使用这些数据来创建自己的图表，也可以使用控制台中提供的图表。

点击率图表

每日点击比例折线图。该图表可帮助您可视化每日点击率的模式。一些例子包括点击率的稳步增加或降低，或者即时回答的减少可能会影响点击率的增加。

如果您使用 API，则可以通过指定 TREND_QUERY_DOC_METRICS 来检索这些数据。您可以使用这些数据来创建自己的图表，也可以使用控制台中提供的图表。

零点击率图表

每日零点击比例折线图。该图表可帮助您可视化每日零点击率的模式。一些例子包括零点击率的稳步增加或降低，或者即时回答的增加可能会影响零点击量的增加。

如果您使用 API，则可以通过指定 TREND_QUERY_DOC_METRICS 来检索这些数据。您可以使用这些数据来创建自己的图表，也可以使用控制台中提供的图表。

零搜索结果率图表

每日零搜索结果比例折线图。该图表可帮助您可视化每日零搜索结果的模式。一些例子包括零搜索结果率的稳步增加或降低，或者索引中文档数量的急剧减少可能会影响零搜索结果的增加。

如果您使用 API，则可以通过指定 TREND_QUERY_DOC_METRICS 来检索这些数据。您可以使用这些数据来创建自己的图表，也可以使用控制台中提供的图表。

即时回答率图表

返回具有即时回答或常见问题解答的查询比例的折线图。该图表可帮助您可视化每日即时回答率的模式。一些例子包括问答类型查询的稳步增加或减少，或者点击次数的减少可能会影响即时回答的增加。

如果您使用 API，则可以通过指定 TREND_QUERY_DOC_METRICS 来检索这些数据。您可以使用这些数据来创建自己的图表，也可以使用控制台中提供的图表。

为渐进式学习提交反馈

Amazon Kendra 使用增量学习来改善搜索结果。凭借来自查询的反馈，渐进式学习可以改进排名算法和优化搜索结果，从而提高准确性。

例如，假设您的用户搜索“医疗保健福利”一词。如果用户总是从列表中选择第二个结果，则随着时间的推移， Amazon Kendra 会将该结果的排名提升到第一名。随着时间的推移，提升效果会降低，因此，如果用户停止选择结果， Amazon Kendra 最终会将其删除，而是显示另一个更受欢迎的结果。这有助于根据相关性、年龄和内容对结果进行 Amazon Kendra 优先排序。

为所有索引和所有支持的文档类型激活渐进式学习。

Amazon Kendra 在您提供反馈后立即开始学习，但可能需要超过 24 小时才能看到反馈的结果。Amazon Kendra 提供了三种提交反馈的方法： AWS 控制台、可以包含在搜索结果页面上的 JavaScript 库以及可以使用的 API。

Amazon Kendra 接受两种类型的用户反馈：

- 点击 - 有关用户选择了哪些查询结果的信息。反馈包括结果 ID 以及选择搜索结果的日期和时间的 Unix 时间戳。

  要提交点击反馈，您的应用程序必须从用户活动收集点击信息，然后将这些信息提交给 Amazon Kendra。您可以使用控制台、 JavaScript 库和 Amazon Kendra API 收集点击信息。
- 相关性 - 有关搜索结果相关性的信息，通常由用户提供。反馈包含结果 ID 和相关性指标（`RELEVANT` 或 `NOT_RELEVANT`）。相关性信息由用户决定。

  要提交相关性反馈，您的应用程序必须提供一种反馈机制，从而允许用户为查询结果选择适当的相关性，然后将这些信息提交给 Amazon Kendra。您只能使用控制台和 Amazon Kendra API 收集相关信息。

当索引处于活动状态时可使用反馈。反馈仅影响将其提交到的索引，不能跨索引使用，也不能用于不同的账户。

在查询 Amazon Kendra 索引时，应提供其他用户上下文。当您提供用户上下文时， Amazon Kendra 能够判断反馈是由单个用户还是由多个用户提供的，并相应地调整搜索结果。

当您提供用户上下文时，查询的反馈与上下文中提供的特定用户相关联。如果您未指定用户上下文，则可以提供用于对查询进行分组和汇总的访问者 ID。

如果您不提供用户上下文或访问者 ID，则反馈将是匿名的，并与其他匿名反馈汇总在一起。

以下代码展示如何将用户上下文作为令牌或访问者 ID 包括在内。

```
response = kendra.query(
    QueryText = query,
    IndexId = index,
    UserToken = {
        Token = "token"
    })

    OR

    response = kendra.query(
    QueryText = query,
    IndexId = index,
    VisitorId = "visitor-id")
```

对于 Web 应用程序，您可以使用 Cookie、位置或浏览器用户为每个用户生成访问者 ID。

对于头部查询，由于查询量极大，因而点击率反馈可以提供足够的信息来提高整体准确性。对于尾部查询，由于查询量极小，主题专家应提交相关和不相关的反馈，从而提高这些查询的准确性。

除了控制台之外，您还可以使用以下两种方法之一： JavaScript 库或 SubmitFeedbackAPI。您只能使用一种反馈收集方法。为了获得最佳结果，您应在发起查询后的 24 小时内提交反馈。

主题

- 使用 Amazon Kendra JavaScript 库提交反馈
- 使用 Amazon Kendra API 提交反馈

使用 Amazon Kendra JavaScript 库提交反馈

Amazon Kendra 提供了一个 JavaScript 库，您可以使用该库向搜索结果页面添加点击反馈。要使用该库，请在显示搜索结果的客户端代码中插入脚本标签，然后向结果列表中的每个文档链接添加信息。当用户选择查看文档的链接时，点击信息就会发送到 Amazon Kendra。

该库适用于支持 ES6/ES2015 JavaScript 版本的浏览器。

步骤 1：在 Amazon Kendra 搜索应用程序中插入脚本标签

在呈现 Amazon Kendra 搜索结果的客户端代码中，插入一个<script>标签并添加对 JavaScript 库的引用：

```
<script>
 (function(w, d, s, c, g, n) {
   if(!w[n]) {
     w[n] = w[n] || function () {
           (w[n].q = w[n].q || []).push(arguments);
     }
     w[n].st = new Date().getTime();
     w[n].ep = g;
     var e = document.createElement(s),
         j = document.getElementsByTagName(s)[0];
     e.async = 1;
     e.src = c;
     e.type = 'module';
     j.parentNode.insertBefore(e, j);
   }
 })(window, document, 'script',
 'library download URL',
 'feedback endpoint',
 'kendraFeedback');
</script>
```

该脚本从 Amazon Kendra 托管 CDN 异步下载 JavaScript 库，并初始化一个名为的全局变量`kendraFeedback`，该变量允许您设置可选参数。

根据托管 Amazon Kendra 索引的区域，将*#####*和*####*替换为下表中的标识符。

区域	下载 URL	反馈端点
us-east-1	https://d2zm0lpns956f8.cloudfront.net/ksf-v1.js	https://ujxwp5s92h.execute-api.us-east-1.amazonaws.com/prod/submit
us-east-2	https://d2crv7fufeg244.cloudfront.net/ksf-v1.js	https://i6h76zwzf3.execute-api.us-east-2.amazonaws.com/prod/submit
us-west-2	https://d2iezfpnpcoujy.cloudfront.net/ksf-v1.js	https://wg6nim909c.execute-api.us-west-2.amazonaws.com/prod/submit

区域	下载 URL	反馈端点
ca-central-1	https://d1zbkfomowykaq.cloudfront.net/ksf-v1.js	https://budi8txevj.execute-api.ca-central-1.amazonaws.com/prod/submit
eu-west-1	https://d3gptlxtulu4us.cloudfront.net/ksf-v1.js	https://po2b11740b.execute-api.eu-west-1.amazonaws.com/prod/submit
ap-southeast-1	https://d1vvuam7g4taoe.cloudfront.net/ksf-v1	https://9je5uw7t5l.execute-api.ap-southeast-1.amazonaws.com/prod/submit
ap-southeast-2	https://dopqntoe6z0ce.cloudfront.net/ksf-v1.js	https://oovf4nvjj7.execute-api.ap-southeast-2.amazonaws.com/prod/submit
ap-south-1	https://d1ts9ouelsmk3g.cloudfront.net/ksf-v1.js	https://k1abnmd43b.execute-api.ap-south-1.amazonaws.com/prod/submit
ap-northeast-1	https://d3w0ybsa293kb4.cloudfront.net/ksf-v1.js	https://wg7rz0uzjh.execute-api.ap-northeast-1.amazonaws.com/prod/submit
eu-west-2	https://d1tsrujswld1d1.cloudfront.net/ksf-v1.js	https://qi7mct3x7f.execute-api.eu-west-2.amazonaws.com/prod/submit

例如，如果您的索引位于美国东部（弗吉尼亚北部），则*### URL* 为 `https://d2zm0lpns956f8.cloudfront.net/ksf-v1.js`，*####*为 `https://ujxwp5s92h.execute-api.us-east-1.amazonaws.com/prod/submit`。

您可以为 Amazon Kendra JavaScript 库进行两个可选设置：

- `disableCookies`— 默认情况下， Amazon Kendra 设置一个唯一标识用户的 Cookie。将其设置为 `true` 可禁用 Cookie。

```
kendraFeedback('disableCookie', 'true | false');
```

`searchDivClassName` - 默认情况下， Amazon Kendra 监控搜索结果页面上的所有链接是否有点击。将其设置为 类名可仅监视指定类中的链接。

```
kendraFeedback('searchDivClassName', 'class name');
```

步骤 2：将反馈令牌添加到搜索结果中

在结果页面上，将一个名 `data-kendra-token` 的 HTML 属性添加到锚点标签或直属父 div 标签，该标签包含查询响应中指向文档的链接。例如：

```
<a href="document location" data-kendra-token="feedback token value"></a>
OR
<div data-url="document location" data-kendra-token="feedback token value"></div>
```

查询响应的 `feedbackToken` 字段中包含一个令牌。如果用户选择响应，则令牌会唯一标识该响应。将令牌的值分配给 `data-kendra-token` 属性。当用户选择结果并将其作为反馈提交给 Amazon Kendra 端点时， Amazon Kendra JavaScript 库会查找此令牌。

Amazon Kendra JavaScript 库仅提交反馈令牌和其他元数据，例如选择结果的时间和唯一的访客 ID。

步骤 3：测试反馈脚本

要确保 JavaScript 库配置正确并向正确的端点发送反馈，请执行以下操作。此示例使用 Chrome 浏览器。

1. 在浏览器中打开 Web 开发者工具。在 Chrome 上，打开浏览器右上角的 Chrome 菜单，选择更多工具，然后选择开发者工具。
2. 确保控制台选项卡中没有与 Amazon Kendra JavaScript 库相关的错误。
3. 搜索并选择任何结果。在开发者工具的网络选项卡中，您会看到发送到反馈端点的请求、结果的令牌以及 200 OK 状态。

使用 Amazon Kendra API 提交反馈

要使用 Amazon Kendra API 提交查询反馈，请使用 SubmitFeedbackAPI。要识别查询，请提供查询适用的索引的索引 ID 以及查询 API 的响应中返回的查询 ID。

以下示例说明如何使用 Amazon Kendra 提交点击和相关性反馈。您可以通过 `ClickFeedbackItems` 和 `RelevanceFeedbackItems` 数组提交多组反馈。此示例提交了一次点击和一个相关性反馈项目。反馈提交使用当前时间。

提交搜索反馈 (AWS SDK)

1. 您可以使用以下带有所需值的示例代码：

   a. `index id`— 查询所适用的索引的 ID。

   b. `query id`— 您要提供反馈的查询。

   c. `result id`— 您要提供反馈的查询结果的 ID。查询响应包含结果 ID。

   d. `relevance value`—RELEVANT（查询结果相关）或NOT_RELEVANT（查询结果不相关）。

   Python

```
import boto3
import time

kendra = boto3.client("kendra")

# Provide the index ID
index_id = "index-id"
# Provide the query ID
query_id = "query-id"
# Provide the search result ID
result_id = "result-id"

# Configure the feedback item
feedback_item = {"ClickTime": int(time.time()),
    "ResultId":result_id}

# Configure the relevance value
relevance_value = "RELEVANT"
relevance_item = {"RelevanceValue": relevance_value,
    "ResultId": result_id
```

```
    }

response = kendra.submit_feedback(
    QueryId = query_id,
    IndexId = index_id,
    ClickFeedbackItems = [feedback_item],
    RelevanceFeedbackItems = [relevance_item]
)


print("Submitted feedback for query: " + query_id)
```

Java

```
package com.amazonaws.kendra;

import java.time.Instant;
import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.ClickFeedback;
import software.amazon.awssdk.services.kendra.model.RelevanceFeedback;
import software.amazon.awssdk.services.kendra.model.RelevanceType;
import software.amazon.awssdk.services.kendra.model.SubmitFeedbackRequest;
import software.amazon.awssdk.services.kendra.model.SubmitFeedbackResponse;

public class SubmitFeedbackExample {
    public static void main(String[] args) {
        KendraClient kendra = KendraClient.builder().build();

        SubmitFeedbackRequest submitFeedbackRequest = SubmitFeedbackRequest
            .builder()
            .indexId("IndexId")
            .queryId("QueryId")
            .clickFeedbackItems(
                ClickFeedback
                .builder()
                .clickTime(Instant.now())
                .resultId("ResultId")
                .build())
```

```
            .relevanceFeedbackItems(
                RelevanceFeedback
                .builder()
                .relevanceValue(RelevanceType.RELEVANT)
                .resultId("ResultId")
                .build())
            .build();

        SubmitFeedbackResponse response =
 kendra.submitFeedback(submitFeedbackRequest);

        System.out.println("Feedback is submitted");
    }
}
```

2. 运行该代码。提交反馈后，代码会显示一条消息。

将自定义同义词添加到索引中

要将自定义同义词添加到索引中，请在同义词库文件中指定这些同义词。在 Amazon Kendra 使用同义词时，可以包含特定于业务的术语或专业术语。通用英语同义词（例如）内置在同义词库文件中 `leader, head` Amazon Kendra ，不应包含在同义词库文件中，包括使用连字符的通用同义词。 Amazon Kendra 支持所有响应类型的同义词，包括DOCUMENT响应类型和 QUESTION_ANSWER / 或ANSWER响应类型。 Amazon Kendra 目前不支持添加标记为停用词的同义词。这些同义词将包含在未来版本中。

Amazon Kendra 在同义词之间建立关联。例如，使用同义词对`Dynamo, Amazon DynamoDB`，将 Dynamo Amazon Kendra 与相关联。 Amazon DynamoDB查询“什么是 Dynamo？” 然后返回诸如 “什么是 Amazon DynamoDB？” 之类的文档。使用同义词， Amazon Kendra 可以更轻松地获取相关性。

同义词库文件是存储在存储桶中的文本文件。 Amazon S3 请参阅 将同义词库添加到索引中。

同义词库文件使用 Solr 同义词格式。 Amazon Kendra 对每个索引的同义词库数量有限制。请参阅限额。

同义词可能在以下场景中很有用：

- 不是传统英语同义词的专业术语，例如，`NLP, Natural Language Processing`。
- 具有复杂语义关联性的专有名词。例如，在机器学习中，`cost, loss, model performance` 是公众可能不太理解的名词。
- 不同形式的产品名称，例如，`Elastic Compute Cloud, EC2`。
- 特定领域或特定业务的术语，例如，产品名称。例如，`Route53, DNS`。

在以下情况下，请勿使用同义词：

- 通用英语同义词，例如，`leader, head`。这些同义词不是特定于领域的，在这些场景中使用同义词可能会产生意想不到的影响。
- 打字错误，例如，`teh => the`。
- 词态变化，例如名词的复数和所有格形式，形容词的比较级和最高级形式，以及动词的过去时、过去分词和进行时。`good, better, best` 是比较级和最高级形容词的一个例子。
- 一元（单个词）停用词，例如，`WHO`。一元停用词不能出现在同义词库中，因此将其排除在搜索范围之外。例如，`WHO => World Health Organization` 会被拒绝。但是，您可以使用 `W.H.O.` 作为同义词，也可以将停用词用作多词同义词的一部分。例如，不允许使用 `of`，但可以使用 `United States of America`。

自定义同义词可以扩展查询范围以涵盖特定于业务的同义词，从而轻松提高 Amazon Kendra对特定业务术语的理解。尽管同义词可以提高搜索准确性，但重要的是要了解同义词如何影响延迟，这样才能对其进行优化。

同义词的一般规则是：查询中与同义词匹配和扩展的词汇越多，对延迟的潜在影响就越大。影响延迟的其他因素包括已编入索引的文档的平均大小、索引的大小、对搜索结果的任何筛选以及 Amazon Kendra 索引的总体负载。与任何同义词都不匹配的查询不受影响。

关于同义词如何影响延迟的一般指南：

使用案例	延迟增加*
典型的自然语言或关键词查询，每个查询包含 3 到 5 个单词	少于 15%
1 个查询词扩展为 3 个同义词	
大约 500,000 个文档（平均每个文档有 10.48 KB 的提取文本）或 30,000 个常见问题解答/问题对的索引	

*性能因在索引上对同义词和配置的具体使用而异。最好测试搜索性能，以便针对您的特定用例获得更准确的基准。

如果您的同义词库很大，词汇扩展率很高，并且延迟增加不在可接受的范围内，则可以尝试以下一两种方法：

- 修剪同义词库以降低扩展率（每个词汇的同义词数量）。
- 减少词汇的总体覆盖范围（同义词库中的行数）。

或者，您可以增加配置容量（虚拟存储单位）以抵消延迟的增加。

主题

- 在搜索结果中突出显示

创建同义词库文件

Amazon Kendra 同义词库文件是一个 UTF-8 编码的文件，其中包含 Solr 同义词列表格式的同义词列表。同义词库文件必须小于 5 MB。

指定同义词映射的方法有两种：

- 双向同义词是逗号分隔的词汇列表。如果您的用户查询任何词汇，则列表中的所有词汇都将用于搜索文档，其中包括原始查询词。
- 单向同义词是用符号 “=>” 分隔的词汇，意思是将词汇映射到其同义词。如果用户查询符号 “=>” 左边的词汇，则该词汇会映射到右边的词汇以使用同义词搜索文档。反之亦然，它不被映射，因而是单向的。

同义词本身区分大小写，但它们映射到的词汇不区分大小写。例如，`ML => Machine Learning` 表示如果您的用户查询“ML”或“ml”或使用其他大小写，它将映射到 “Machine Learning”。反之，如果您要这样映射 `Machine Learning => ML`，那么“Machine Learning”、“machine learning”或其他情况将映射到“ML”。

同义词不会在特殊字符上搜索完全匹配的字符。例如，如果搜索 dead-letter-queue “”，则 Amazon Kendra 可以返回与 “死信队列”（无连字符）匹配的文档。如果您的文档包含连字符（例如 dead-letter-queue “”），则会在搜索过程中 Amazon Kendra 处理文档以删除连字符。对于内置于同义词库文件中 Amazon Kendra 且不应包含在同义词库文件中的通用英语同义词术语， Amazon Kendra 可以同时搜索该术语的连字符版本和该术语的非连字符版本。例如，如果您搜索 “第三方” 和 “第三方”，则 Amazon Kendra 返回与这两个术语的任一版本相匹配的文档。

对于包含停用词或常用词的同义词， Amazon Kendra 返回与包括停用词在内的术语匹配的文档。例如，您可以创建同义词规则来映射 “登机” 和 “入职”。不能单独使用停用词作为同义词。例如，如果搜索 “on”，则 Amazon Kendra 无法返回所有包含 “on” 的文档。

某些同义词规则会被忽略。例如，`a => b`是一条规则，但`a => a`会被忽略并且不算作规则。

词汇计数是同义词库文件中唯一词汇的数量。下面的示例文件包括术语`AWS CodeStar`、`ML`、`Machine Learning`、`autoscaling group`、`ASG`、等。

每个同义词库有最大数量的同义词规则，每个术语的同义词数量有上限。有关更多信息，请参阅 的配额 Amazon Kendra。

以下示例显示了包含同义词规则的同义词库文件。每行都包含一条同义词规则。搜索会忽略空行和注释。

```
# Lines starting with pound are comments and blank lines are ignored.

# Synonym relationships can be defined as unidirectional or bidirectional
 relationships.

# Unidirection relationships are represented by any term sequence
# on the left hand side (LHS) of "=>" followed by synonyms on the right hand side (RHS)
CodeStar => AWS CodeStar
# This will map CodeStar to AWS CodeStar, but not vice-versa

# To map terms vice versa
ML => Machine Learning
Machine Learning => ML

# Multiple synonym relationships may be defined in one line as well by comma
 seperation.
autoscaling group, ASG => Auto Scaling group, autoscaling
# The above is equivalent to:
# autoscaling group => Auto Scaling group, autoscaling
# ASG => Auto Scaling group, autoscaling

# Bi-directional synonyms are comma separated terms with no "=>"
DNS, Route53, Route 53
# DNS, Route53, and Route 53 map to one another and are interchangeable at match time
# The above is equivalent to:
# DNS => Route53, Route 53
# Route53 => DNS, Route 53
# Route 53 => DNS, Route53

# Overlapping LHS terms will be merged
Beta => Alpha
Beta => Gamma
Beta, Delta
# is equivalent to:
# Beta => Alpha, Gamma, Delta
# Delta => Beta

# Each line contains a single synonym rule.
# Synonym rule count is the total number of lines defining synonym relationships
# Term count is the total number of unique terms for all rules.
```

```
# Comments and blanks lines do not count.
```

将同义词库添加到索引中

以下过程说明如何将包含同义词的同义词库文件添加到索引中。要了解更新后的同义词库文件的效果，最多可能需要 30 分钟。有关同义词库文件的更多信息，请参阅创建同义词库文件。

Console

添加同义词库

1. 在左侧导航窗格中要添加同义词列表的索引下，请选择同义词。
2. 在同义词页面上，选择添加同义词库。
3. 在定义同义词库中，为同义词库指定名称和可选描述。
4. 在同义词库设置中，提供同义词库文件的 Amazon S3 路径。文件必须小于 5 MB。
5. 对于 IAM 角色，选择一个角色或选择创建新角色并指定角色名称以创建新角色。 Amazon Kendra 使用此角色代表您访问 Amazon S3 资源。IAM 角色的前缀为 “AmazonKendra-”。
6. 选择保存以保存配置并添加同义词库。收录同义词库后，它就会处于活动状态，结果中会突出显示同义词。要了解同义词库文件的效果，最多可能需要 30 分钟。

CLI

要使用将同义词添加到索引中 AWS CLI，请调用：`create-thesaurus`

```
aws kendra create-thesaurus \
--index-id index-id \
--name "thesaurus-name" \
--description "thesaurus-description" \
--source-s3-path "Bucket=bucket-name,Key=thesaurus/synonyms.txt" \
--role-arn role-arn
```

调用 `list-thesauri` 以查看同义词库列表：

```
aws kendra list-thesauri \
--index-id index-id
```

要查看同义词库的详细信息，请调用 `describe-thesaurus`：

```
aws kendra describe-thesaurus \
--index-id index-id \
--index-id thesaurus-id
```

要了解同义词库文件的效果，最多可能需要 30 分钟。

Python

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Create a thesaurus")

thesaurus_name = "thesaurus-name"
thesaurus_description = "thesaurus-description"
thesaurus_role_arn = "role-arn"

index_id = "index-id"

s3_bucket_name = "bucket-name"
s3_key = "thesaurus-file"
source_s3_path= {
    'Bucket': s3_bucket_name,
    'Key': s3_key
}

try:
    thesaurus_response = kendra.create_thesaurus(
        Description = thesaurus_description,
        Name = thesaurus_name,
        RoleArn = thesaurus_role_arn,
        IndexId = index_id,
        SourceS3Path = source_s3_path
    )

    pprint.pprint(thesaurus_response)

    thesaurus_id = thesaurus_response["Id"]
```

```
    print("Wait for Kendra to create the thesaurus.")

    while True:
        # Get thesaurus description
        thesaurus_description = kendra.describe_thesaurus(
            Id = thesaurus_id,
            IndexId = index_id
        )
        # If status is not CREATING quit
        status = thesaurus_description["Status"]
        print("Creating thesaurus. Status: " + status)
        if status != "CREATING":
            break
        time.sleep(60)

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.CreateThesaurusRequest;
import software.amazon.awssdk.services.kendra.model.CreateThesaurusResponse;
import software.amazon.awssdk.services.kendra.model.DescribeThesaurusRequest;
import software.amazon.awssdk.services.kendra.model.DescribeThesaurusResponse;
import software.amazon.awssdk.services.kendra.model.S3Path;
import software.amazon.awssdk.services.kendra.model.ThesaurusStatus;

public class CreateThesaurusExample {

  public static void main(String[] args) throws InterruptedException {

    KendraClient kendra = KendraClient.builder().build();

    String thesaurusName = "thesaurus-name";
    String thesaurusDescription = "thesaurus-description";
    String thesaurusRoleArn = "role-arn";

    String s3BucketName = "bucket-name";
```

```
    String s3Key = "thesaurus-file";
    String indexId = "index-id";

    System.out.println(String.format("Creating a thesaurus named %s",
thesaurusName));
    CreateThesaurusRequest createThesaurusRequest = CreateThesaurusRequest
        .builder()
        .name(thesaurusName)
        .indexId(indexId)
        .description(thesaurusDescription)
        .roleArn(thesaurusRoleArn)
        .sourceS3Path(S3Path.builder()
            .bucket(s3BucketName)
            .key(s3Key)
            .build())
        .build();
    CreateThesaurusResponse createThesaurusResponse =
kendra.createThesaurus(createThesaurusRequest);
    System.out.println(String.format("Thesaurus response %s",
createThesaurusResponse));

    String thesaurusId = createThesaurusResponse.id();

    System.out.println(String.format("Waiting until the thesaurus with ID %s is
created.", thesaurusId));

    while (true) {
      DescribeThesaurusRequest describeThesaurusRequest =
DescribeThesaurusRequest.builder()
          .id(thesaurusId)
          .indexId(indexId)
          .build();
      DescribeThesaurusResponse describeThesaurusResponse =
kendra.describeThesaurus(describeThesaurusRequest);
      ThesaurusStatus status = describeThesaurusResponse.status();
      if (status != ThesaurusStatus.CREATING) {
        break;
      }

      TimeUnit.SECONDS.sleep(60);
    }

    System.out.println("Thesaurus creation is complete.");
  }
```

```
}
```

更新同义词库

在创建同义词库后，您可以更改其配置。您可以更改同义词库名称和 IAM 信息之类的详细信息。您也可以更改同义词库文件位置的 Amazon S3 路径。如果更改同义词库文件的路径，则 Amazon Kendra 会将现有同义词库替换为更新后的路径中指定的同义词库。

要了解更新后的同义词库文件的效果，最多可能需要 30 分钟。

> **Note**
>
> 如果同义词库文件中存在验证或语法错误，则会保留之前上传的同义词库文件。

以下过程说明如何修改同义词库的详细信息。

Console

修改同义词库的详细信息

1. 在左侧导航窗格中要修改的索引下，选择同义词。
2. 在同义词页面上，选择要修改的同义词库，然后选择编辑。
3. 在更新同义词库页面上，更新同义词库详细信息。
4. （可选）选择 “更改同义词库文件路径”，然后指定新同义词库文件的 Amazon S3 路径。您的现有同义词库文件将替换为您指定的文件。如果不更改路径，则从现有路径 Amazon Kendra 重新加载同义词库。

   如果选择 “保留当前同义词库文件”，则 Amazon Kendra 不会重新加载同义词库文件。
5. 选择保存以保存配置。

您也可以从现有同义词库路径重新加载同义词库。

从现有路径重新加载同义词库

1. 在左侧导航窗格中要修改的索引下，选择同义词。
2. 在同义词页面上，选择要重新加载的同义词库，然后选择刷新。

3. 在重新加载同义词库文件页面上，确认要刷新同义词库文件。

CLI

要更新同义词库，请调用 update-thesaurus：

```
aws kendra update-thesaurus \
--index-id index-id \
--name "thesaurus-name" \
--description "thesaurus-description" \
--source-s3-path "Bucket=bucket-name,Key=thesaurus/synonyms.txt" \
--role-arn role-arn
```

Python

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra = boto3.client("kendra")

print("Update a thesaurus")

thesaurus_name = "thesaurus-name"
thesaurus_description = "thesaurus-description"
thesaurus_role_arn = "role-arn"

thesaurus_id = "thesaurus-id"
index_id = "index-id"

s3_bucket_name = "bucket-name"
s3_key = "thesaurus-file"
source_s3_path= {
    'Bucket': s3_bucket_name,
    'Key': s3_key
}

try:
    kendra.update_thesaurus(
        Id = thesaurus_id,
        IndexId = index_id,
```

```
        Description = thesaurus_description,
        Name = thesaurus_name,
        RoleArn = thesaurus_role_arn,
        SourceS3Path = source_s3_path
    )

    print("Wait for Kendra to update the thesaurus.")

    while True:
        # Get thesaurus description
        thesaurus_description = kendra.describe_thesaurus(
            Id = thesaurus_id,
            IndexId = index_id
        )
        # If status is not UPDATING quit
        status = thesaurus_description["Status"]
        print("Updating thesaurus. Status: " + status)
        if status != "UPDATING":
            break
        time.sleep(60)

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.UpdateThesaurusRequest;
import software.amazon.awssdk.services.kendra.model.DescribeThesaurusRequest;
import software.amazon.awssdk.services.kendra.model.DescribeThesaurusResponse;
import software.amazon.awssdk.services.kendra.model.S3Path;
import software.amazon.awssdk.services.kendra.model.ThesaurusStatus;

public class UpdateThesaurusExample {

  public static void main(String[] args) throws InterruptedException {

    KendraClient kendra = KendraClient.builder().build();
```

```
    String thesaurusName = "thesaurus-name";
    String thesaurusDescription = "thesaurus-description";
    String thesaurusRoleArn = "role-arn";

    String s3BucketName = "bucket-name";
    String s3Key = "thesaurus-file";

    String thesaurusId = "thesaurus-id";
    String indexId = "index-id";

    UpdateThesaurusRequest updateThesaurusRequest = UpdateThesaurusRequest
        .builder()
        .id(thesaurusId)
        .indexId(indexId)
        .name(thesaurusName)
        .description(thesaurusDescription)
        .roleArn(thesaurusRoleArn)
        .sourceS3Path(S3Path.builder()
            .bucket(s3BucketName)
            .key(s3Key)
            .build())
        .build();
    kendra.updateThesaurus(updateThesaurusRequest);

    System.out.println(String.format("Waiting until the thesaurus with ID %s is
updated.", thesaurusId));

    // a new source s3 path requires re-consumption by Kendra
    // and so can take as long as a Create Thesaurus operation
    while (true) {
      DescribeThesaurusRequest describeThesaurusRequest =
DescribeThesaurusRequest.builder()
          .id(thesaurusId)
          .indexId(indexId)
          .build();
      DescribeThesaurusResponse describeThesaurusResponse =
kendra.describeThesaurus(describeThesaurusRequest);
      ThesaurusStatus status = describeThesaurusResponse.status();
      if (status != ThesaurusStatus.UPDATING) {
        break;
      }

      TimeUnit.SECONDS.sleep(60);
    }
```

```
    System.out.println("Thesaurus update is complete.");
  }
}
```

更新同义词库

以下过程说明如何删除同义词库。

Console

1. 在左侧导航窗格中要修改的索引下，选择同义词。
2. 在同义词页面上，选择要删除的同义词库。
3. 在同义词库详细信息页面上，选择删除，然后确认删除。

CLI

要使用删除索引的同义词，请调用： AWS CLIdelete-thesaurus

```
aws kendra delete-thesaurus \
--index-id index-id \
--id thesaurus-id
```

Python

```
import boto3
from botocore.exceptions import ClientError

kendra = boto3.client("kendra")

print("Delete a thesaurus")

thesaurus_id = "thesaurus-id"
index_id = "index-id"

try:
    kendra.delete_thesaurus(
        Id = thesaurus_id,
        IndexId = index_id
    )
```

```
except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

```
package com.amazonaws.kendra;

import software.amazon.awssdk.services.kendra.KendraClient;
import software.amazon.awssdk.services.kendra.model.DeleteThesaurusRequest;

public class DeleteThesaurusExample {

  public static void main(String[] args) throws InterruptedException {

    KendraClient kendra = KendraClient.builder().build();

    String thesaurusId = "thesaurus-id";
    String indexId = "index-id";

    DeleteThesaurusRequest updateThesaurusRequest = DeleteThesaurusRequest
        .builder()
        .id(thesaurusId)
        .indexId(indexId)
        .build();
    kendra.deleteThesaurus(updateThesaurusRequest);
  }
}
```

在搜索结果中突出显示

默认情况下，突出显示同义词处于启用状态。亮点信息包含在 Amazon Kendra SDK 和 CLI 查询结果中。如果您 Amazon Kendra 使用 SDK 或 CLI 进行交互，则由您决定如何显示结果。

同义词突出显示将使用突出显示类型 THESAURUS_SYNONYM。有关突出显示的更多信息，请参阅突出显示对象。

教程：使用 Amazon Kendra 构建扩充元数据的智能搜索解决方案

本教程向您展示如何使用 Amazon Kendra、Amazon Comprehend、Amazon Simprehend、Amazon 简单存储服务（S3）和 AWS CloudShell 为您的企业数据构建一个富含元数据、基于自然语言的智能搜索解决方案。

Amazon Kendra 是一项智能搜索服务，可以为您的非结构化自然语言数据存储库构建搜索索引。为了让您的客户更轻松地找到和筛选相关答案，您可以使用 Amazon Comprehend 从您的数据中提取元数据，然后将其提取到您的 Amazon Kendra 搜索索引中。

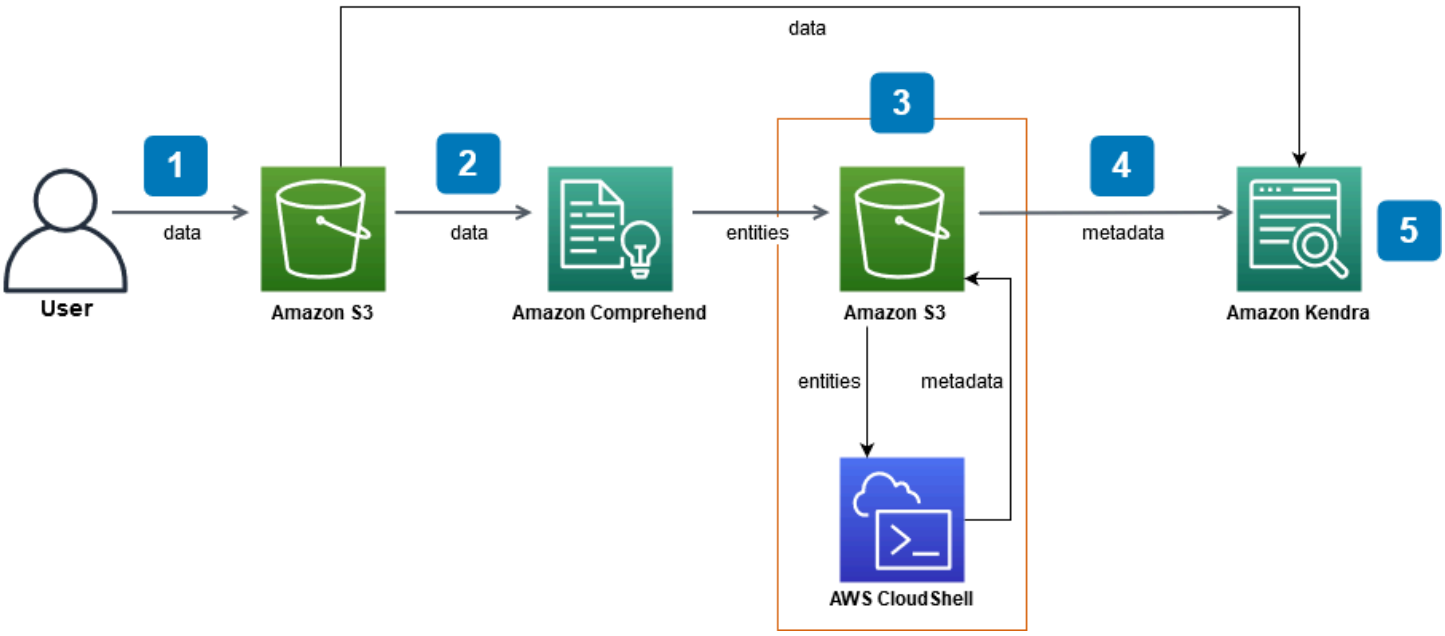
Amazon Comprehend 是一项可识别实体的自然语言处理（NLP）服务。实体是对数据中的人物、地点、位置、组织和对象的引用。

本教程使用新闻文章的示例数据集来提取实体，将其转换为元数据，然后将其提取到您的 Amazon Kendra 索引中以进行搜索。添加的元数据允许您使用这些实体的任何子集筛选搜索结果，并提高搜索准确性。通过学习本教程，您将学习如何在没有任何专业机器学习知识的情况下为您的企业数据创建搜索解决方案。

本教程向您展示如何使用以下步骤构建搜索解决方案：

1. 在 Amazon S3 中存储新闻文章的示例数据集。
2. 使用 Amazon Comprehend 从您的数据中提取实体。
3. 运行 Python 3 脚本将实体转换为 Amazon Kendra 索引元数据格式，并将此元数据存储在 S3 中。
4. 创建 Amazon Kendra 搜索索引并提取数据和元数据。
5. 查询搜索索引。

以下图表显示了工作流程：



完成本教程的预计时间：1 小时

预计费用：本教程中的某些操作会对您的 AWS 账户产生费用。有关每项服务的费用的更多信息，请参阅 Amazon S3、Amazon Comprehend、AWS CloudShell 和 Amazon Kendra 的价格页面。

主题

先决条件

完成本教程需要以下资源：

- 一个 AWS 账户。如果您没有 AWS 账户，请按照设置 Amazon Kendra 中的步骤设置您的 AWS 账户。

- 运行 Windows、macOS 或 Linux 的开发计算机（用于访问 AWS 管理控制台）。有关更多信息，请参阅配置 AWS 管理控制台。
- AWS Identity and Access Management（IAM）用户。要了解如何为您的账户设置 IAM 用户和群组，请参阅《IAM 用户指南》中的入门部分。

  如果您使用的是 AWS Command Line Interface，则还需要将以下策略附加到您的 IAM 用户，以授予其完成本教程所需的基本权限。

  有关更多信息，请参阅创建 IAM 策略和添加和移除 IAM 身份权限。
- AWS 区域服务列表。要减少延迟，请选择 Amazon Comprehend 和 Amazon Kendra 支持的离您的地理位置最近的 AWS 区域。
- （可选）一个 AWS Key Management Service。虽然本教程不使用加密，但您可能需要针对您的特定用例使用加密最佳实践。
- （可选）Amazon Virtual Private Cloud。虽然本教程不使用 VPC，但您可能需要使用 VPC 最佳实践来确保特定用例的数据安全。

步骤 1：向 Amazon S3 添加文档

在对数据集运行 Amazon Comprehend 实体分析任务之前，您需要创建一个 Amazon S3 存储桶来托管数据、元数据和 Amazon Comprehend 实体分析输出。

主题

下载示例数据集

在 Amazon Comprehend 可以对您的数据运行实体分析任务之前，您必须下载并提取数据集并将其上传到 S3 存储桶。

下载并提取数据集（控制台）

1. 在您的设备上下载 tutorial-dataset.zip 文件夹。

2. 解压tutorial-dataset文件夹以访问 data 文件夹。

下载并提取数据集（终端）

1. 要下载 tutorial-dataset，请在终端窗口中运行以下命令：

   Linux

   ```
   curl -o path/tutorial-dataset.zip https://docs.aws.amazon.com/kendra/latest/dg/
   samples/tutorial-dataset.zip
   ```

   其中：

   - *path/* 是您要保存 zip 文件夹的位置的本地文件路径。

   macOS

   ```
   curl -o path/tutorial-dataset.zip https://docs.aws.amazon.com/kendra/latest/dg/
   samples/tutorial-dataset.zip
   ```

   其中：

   - *path/* 是您要保存 zip 文件夹的位置的本地文件路径。

   Windows

   ```
   curl -o path/tutorial-dataset.zip https://docs.aws.amazon.com/kendra/latest/dg/
   samples/tutorial-dataset.zip
   ```

   其中：

   - *path/* 是您要保存 zip 文件夹的位置的本地文件路径。

2. 要从 zip 文件夹中提取数据，请在终端窗口中运行以下命令：

   Linux

   ```
   unzip path/tutorial-dataset.zip -d path/
   ```

其中：

- *path/* 是您保存的 zip 文件夹的本地文件路径。

macOS

```
unzip path/tutorial-dataset.zip -d path/
```

其中：

- *path/* 是您保存的 zip 文件夹的本地文件路径。

Windows

```
tar -xf path/tutorial-dataset.zip -C path/
```

其中：

- *path/* 是您保存的 zip 文件夹的本地文件路径。

在此步骤结束时，您应该将解压缩的文件放在名为 tutorial-dataset 的解压缩文件夹中。此文件夹包含一个带有 Apache 2.0 开源属性的 README 文件和一个名为 data 的文件夹，其中包含本教程的数据集。该数据集由 100 个带有 .story 扩展名的文件组成。

创建 Amazon S3 存储桶

下载和提取示例数据文件夹后，您可以将其存储在 Amazon S3 存储桶中。

> **Important**
>
> 在所有 AWS中，Amazon S3 存储桶的名称必须是唯一的。

创建 S3 存储桶（控制台）

1. 登录 AWS Management Console 并打开 Amazon S3 控制台，网址为 https://console.aws.amazon.com/s3/。

2. 在存储桶中，选择创建存储桶。
3. 对于 Bucket name（存储桶名称），请输入唯一名称。
4. 对于区域，选择要在其中创建存储桶的 AWS 区域。

> **Note**
>
> 您必须选择同时支持 Amazon Comprehend 和 Amazon Kendra 的区域。创建存储桶后无法更改其区域。

5. 保留此存储桶的“阻止公共访问”设置、存储桶版本控制和标签的默认设置。
6. 对于默认加密，请选择禁用。
7. 保留高级设置的默认设置。
8. 查看您的存储桶配置，然后选择创建存储桶。

创建 S3 存储桶（AWS CLI）

1. 要创建 S3 存储桶，请使用 AWS CLI中的 create-bucket 命令：

   Linux

   ```
   aws s3api create-bucket \
           --bucket DOC-EXAMPLE-BUCKET \
           --region aws-region \
           --create-bucket-configuration LocationConstraint=aws-region
   ```

   其中：

   - DOC-EXAMPLE-BUCKET 是存储桶的名称，
   - *aws-region* 是要在其中创建存储桶的区域。

   macOS

   ```
   aws s3api create-bucket \
           --bucket DOC-EXAMPLE-BUCKET \
           --region aws-region \
           --create-bucket-configuration LocationConstraint=aws-region
   ```

其中：

- DOC-EXAMPLE-BUCKET 是存储桶的名称，
- *aws-region* 是要在其中创建存储桶的区域。

Windows

```
aws s3api create-bucket ^
        --bucket DOC-EXAMPLE-BUCKET ^
        --region aws-region ^
        --create-bucket-configuration LocationConstraint=aws-region
```

其中：

- DOC-EXAMPLE-BUCKET 是存储桶的名称，
- *aws-region* 是要在其中创建存储桶的区域。

> **Note**
>
> 您必须选择同时支持 Amazon Comprehend 和 Amazon Kendra 的区域。创建存储桶后无法更改其区域。

2. 要确保您的存储桶已成功创建，请使用 list 命令：

Linux

```
aws s3 ls
```

macOS

```
aws s3 ls
```

Windows

```
aws s3 ls
```

在 S3 存储桶中创建数据和元数据文件夹

创建 S3 存储桶后，您可以在其中创建数据和元数据文件夹。

在 S3 存储桶中创建文件夹（控制台）

1. 打开 Amazon S3 控制台，网址为：https://console.aws.amazon.com/s3/。
2. 在存储桶中，单击存储桶列表中您的存储桶的名称。
3. 从对象选项卡中，选择创建文件夹。
4. 对于新文件夹名称，输入 **`data`**。
5. 对于加密设置，选择禁用。
6. 请选择 Create folder（创建文件夹）。
7. 重复步骤 3 到 6，创建另一个用于存储 Amazon Kendra 元数据的文件夹，并命名步骤 4 **`metadata`** 中创建的文件夹。

在 S3 存储桶中创建文件夹（AWS CLI）

1. 要在您的 S3 存储桶中创建 `data` 文件夹，请在 AWS CLI中使用 put-object 命令：

   Linux

   ```
   aws s3api put-object \
           --bucket DOC-EXAMPLE-BUCKET \
           --key data/
   ```

   其中：

   - DOC-EXAMPLE-BUCKET 是存储桶的名称。

   macOS

   ```
   aws s3api put-object \
           --bucket DOC-EXAMPLE-BUCKET \
           --key data/
   ```

   其中：

   - DOC-EXAMPLE-BUCKET 是存储桶的名称。

Windows

```
aws s3api put-object ^
        --bucket DOC-EXAMPLE-BUCKET ^
        --key data/
```

其中：

- DOC-EXAMPLE-BUCKET 是存储桶的名称。

2. 要在您的 S3 存储桶中创建 `metadata` 文件夹，请在 AWS CLI中使用 put-object 命令：

Linux

```
aws s3api put-object \
        --bucket DOC-EXAMPLE-BUCKET \
        --key metadata/
```

其中：

- DOC-EXAMPLE-BUCKET 是存储桶的名称。

macOS

```
aws s3api put-object \
        --bucket DOC-EXAMPLE-BUCKET \
        --key metadata/
```

其中：

- DOC-EXAMPLE-BUCKET 是存储桶的名称。

Windows

```
aws s3api put-object ^
        --bucket DOC-EXAMPLE-BUCKET ^
        --key metadata/
```

其中：

- DOC-EXAMPLE-BUCKET 是存储桶的名称。

3. 为确保您的文件夹已成功创建，请使用 list 命令检查存储桶中的内容：

Linux

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/
```

其中：

- DOC-EXAMPLE-BUCKET 是存储桶的名称。

macOS

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/
```

其中：

- DOC-EXAMPLE-BUCKET 是存储桶的名称。

Windows

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/
```

其中：

- DOC-EXAMPLE-BUCKET 是存储桶的名称。

上传输入数据

创建数据和元数据文件夹后，将示例数据集上传到 data 文件夹中。

将示例数据集上传到数据文件夹中（控制台）

1. 打开 Amazon S3 控制台，网址为：https://console.aws.amazon.com/s3/。
2. 在存储桶中，从存储桶列表中单击存储桶的名称，然后单击 data。

3. 选择上传，然后选择添加文件。
4. 在对话框中，在本地设备中导航到 tutorial-dataset 文件夹内的文件夹 data，选择所有文件，然后选择打开。
5. 保留目标、权限和属性的默认设置。
6. 选择上传。

将示例数据集上传到数据文件夹中（AWS CLI）

1. 要将示例数据上传到 data 文件夹，请使用以下 AWS CLI中的 copy 命令：

   Linux

   ```
   aws s3 cp path/tutorial-dataset/data s3://DOC-EXAMPLE-BUCKET/data/ --recursive
   ```

   其中：

   - *path/* 是设备上 tutorial-dataset 文件夹的文件路径，
   - DOC-EXAMPLE-BUCKET 是存储桶的名称。

   macOS

   ```
   aws s3 cp path/tutorial-dataset/data s3://DOC-EXAMPLE-BUCKET/data/ --recursive
   ```

   其中：

   - *path/* 是设备上 tutorial-dataset 文件夹的文件路径，
   - DOC-EXAMPLE-BUCKET 是存储桶的名称。

   Windows

   ```
   aws s3 cp path/tutorial-dataset/data s3://DOC-EXAMPLE-BUCKET/data/ --recursive
   ```

   其中：

   - *path/* 是设备上 tutorial-dataset 文件夹的文件路径，
   - DOC-EXAMPLE-BUCKET 是存储桶的名称。

2. 要确保您的数据集文件已成功上传到 data 文件夹中，请使用 AWS CLI中的 list 命令：

Linux

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/data/
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

macOS

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/data/
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

Windows

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/data/
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

在此步骤结束时，您将有一个 S3 存储桶，其中的数据集存储在 data 文件夹中，还有一个用于存储您的 Amazon Kendra 元数据的空 metadata 文件夹。

步骤 2：在 Amazon Comprehend 上运行实体分析任务 Amazon Comprehend

将示例数据集存储在 S3 存储桶中后，您可以运行 Amazon Comprehend 实体分析任务，从您的文档中提取实体。这些实体将形成 Amazon Kendra 的自定义属性，并帮助您筛选索引中的搜索结果。有关更多信息，请参阅检测事件。

主题

正在运行 Amazon Comprehend 实体分析任务

要从数据集中提取实体，您需要运行 Amazon Comprehend 实体分析作业。

如果您在此步骤中使用 AWS CLI，则首先要为 Amazon Comprehend 创建并附加一个 AWS IAM 角色和策略，然后运行实体分析任务。要对您的示例数据运行实体分析任务，Amazon Comprehend 需要：

- 将其识别为可信实体的 AWS Identity and Access Management (IAM) 角色
- 附加到 AWS IAM 角色的 IAM 策略，该策略授予其访问您的 S3 存储桶的权限

有关更多信息，请参阅如何将 Amazon Comprehend 与 IAM 配合使用和适用于 Amazon Comprehend 的基于身份的策略。

运行 Amazon Comprehend 实体分析任务（控制台）

1. 打开 the Amazon Comprehend 控制台，网址：https://console.aws.amazon.com/comprehend/。

   > **Important**
   >
   > 确保您所在的区域与您创建 Amazon S3 存储桶时所在的区域相同。如果您在其他区域，请从顶部导航栏的 AWS 区域选择器中选择您创建 S3 存储桶的区域。

2. 选择启动 Amazon Comprehend。
3. 在左侧导航窗格中，选择分析作业。
4. 请选择创建作业。
5. 在作业设置部分，执行以下操作：

   a. 对于名称，请输入 **`data-entities-analysis`**。
   b. 对于分析类型，选择实体。
   c. 在语言中，选择英语。
   d. 保持作业加密处于关闭状态。

6. 请在输入数据部分，执行以下操作：

   a. 在数据来源中，选择我的文档。

b. 对于 S3 位置，选择浏览 S3。

c. 在选择资源中，从存储桶列表中单击存储桶的名称。

d. 对于对象，选择 data 的选项按钮，然后单击选择。

e. 对于输入格式，选择每行一个文档。

7. 请在输出数据部分，执行以下操作：

a. 对于 S3 位置，选择浏览 S3，然后从存储桶列表中选择存储桶的选项框并单击选择。

b. 保持加密处于关闭状态。

8. 在访问权限部分，执行以下操作：

a. 对于 IAM 角色，选择创建 IAM 角色。

b. 在访问权限中，请选择输入和输出 S3 存储桶。

c. 在名称后缀中，输入 **comprehend-role**。该角色提供对 Amazon S3 存储桶的访问权限。

9. 保持默认的 VPC 设置。

10. 请选择创建作业。

运行 Amazon Comprehend 实体分析任务（AWS CLI）

1. 要为 Amazon Comprehend 创建并附加将其识别为可信实体的 IAM 角色，请执行以下操作：

a. 在本地设备上的文本编辑器中，将以下信任策略另存为名为 comprehend-trust-policy.json 的 JSON 文件。

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": "comprehend.amazonaws.com"
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

b. 要创建名为 comprehend-role 的 IAM 角色，并将您保存的 comprehend-trust-policy.json 文件附加到该角色上，请使用 create-role 命令：

Linux

```
aws iam create-role \
          --role-name comprehend-role \
          --assume-role-policy-document file://path/comprehend-trust-
policy.json
```

其中：

- *path/* 是本地设备上 comprehend-trust-policy.json 的文件路径。

macOS

```
aws iam create-role \
          --role-name comprehend-role \
          --assume-role-policy-document file://path/comprehend-trust-
policy.json
```

其中：

- *path/* 是本地设备上 comprehend-trust-policy.json 的文件路径。

Windows

```
aws iam create-role ^
          --role-name comprehend-role ^
          --assume-role-policy-document file://path/comprehend-trust-
policy.json
```

其中：

- *path/* 是本地设备上 comprehend-trust-policy.json 的文件路径。

c. 将 Amazon Resource Name（ARN）复制到您的文本编辑器中，并将其作为 comprehend-role-arn 保存到本地。

> **Note**
>
> ARN 的格式类似于 *arn:aws:iam::123456789012:role/comprehend-role*。您需要保存的 comprehend-role-arn 的 ARN 才能运行 Amazon Comprehend 分析任务。

2. 要创建 IAM 策略并将其附加到您的 IAM 角色以授予其访问您的 S3 存储桶的权限，请执行以下操作：

   a. 在本地设备上的文本编辑器中，将以下信任策略另存为名为 comprehend-S3-access-policy.json 的 JSON 文件。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::DOC-EXAMPLE-BUCKET/*"
            ],
            "Effect": "Allow"
        },
        {
            "Action": [
                "s3:ListBucket"
            ],
            "Resource": [
                "arn:aws:s3:::DOC-EXAMPLE-BUCKET"
            ],
            "Effect": "Allow"
        },
        {
            "Action": [
                "s3:PutObject"
            ],
            "Resource": [
                "arn:aws:s3:::DOC-EXAMPLE-BUCKET/*"
            ],
            "Effect": "Allow"
```

```
        }
    ]
}
```

b. 要创建名为 comprehend-S3-access-policy 的 IAM 策略以访问您的 S3 存储桶，请使用 create-policy 命令：

Linux

```
aws iam create-policy \
          --policy-name comprehend-S3-access-policy \
          --policy-document file://path/comprehend-S3-access-policy.json
```

其中：

- *path/* 是本地设备上 comprehend-S3-access-policy.json 的文件路径。

macOS

```
aws iam create-policy \
          --policy-name comprehend-S3-access-policy \
          --policy-document file://path/comprehend-S3-access-policy.json
```

其中：

- *path/* 是本地设备上 comprehend-S3-access-policy.json 的文件路径。

Windows

```
aws iam create-policy ^
          --policy-name comprehend-S3-access-policy ^
          --policy-document file://path/comprehend-S3-access-policy.json
```

其中：

- *path/* 是本地设备上 comprehend-S3-access-policy.json 的文件路径。

c. 将 Amazon Resource Name（ARN）复制到您的文本编辑器中，并将其作为 comprehend-S3-access-arn 保存到本地。

> **Note**
>
> 该 ARN 的格式类似于 *arn:aws:iam::123456789012:role/comprehend-S3-access-policy*。您需要保存的 ARN 才能将 comprehend-S3-access-arn 附加到您的 comprehend-S3-access-policy IAM 角色。

d. 要将附加comprehend-S3-access-policy到您的 IAM 角色，请使用以下attach-role-policy命令：

Linux

```
aws iam attach-role-policy \
          --policy-arn policy-arn \
          --role-name comprehend-role
```

其中：

- *policy-arn* 另存为 comprehend-S3-access-arn 的 ARN。

macOS

```
aws iam attach-role-policy \
          --policy-arn policy-arn \
          --role-name comprehend-role
```

其中：

- *policy-arn* 另存为 comprehend-S3-access-arn 的 ARN。

Windows

```
aws iam attach-role-policy ^
          --policy-arn policy-arn ^
          --role-name comprehend-role
```

其中：

- *policy-arn* 另存为 comprehend-S3-access-arn 的 ARN。

3. 要运行 Amazon Comprehend 实体分析任务，请使用以下命令：start-entities-detection-job

Linux

```
aws comprehend start-entities-detection-job \
        --input-data-config S3Uri=s3://DOC-EXAMPLE-BUCKET/
data/,InputFormat=ONE_DOC_PER_FILE \
        --output-data-config S3Uri=s3://DOC-EXAMPLE-BUCKET/ \
        --data-access-role-arn role-arn \
        --job-name data-entities-analysis \
        --language-code en \
        --region aws-region
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。
- *role-arn* 是另存为 comprehend-role-arn 的 ARN。
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws comprehend start-entities-detection-job \
        --input-data-config S3Uri=s3://DOC-EXAMPLE-BUCKET/
data/,InputFormat=ONE_DOC_PER_FILE \
        --output-data-config S3Uri=s3://DOC-EXAMPLE-BUCKET/ \
        --data-access-role-arn role-arn \
        --job-name data-entities-analysis \
        --language-code en \
        --region aws-region
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。
- *role-arn* 是另存为 comprehend-role-arn 的 ARN。
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws comprehend start-entities-detection-job ^
```

```
        --input-data-config S3Uri=s3://DOC-EXAMPLE-BUCKET/
data/,InputFormat=ONE_DOC_PER_FILE ^
        --output-data-config S3Uri=s3://DOC-EXAMPLE-BUCKET/ ^
        --data-access-role-arn role-arn ^
        --job-name data-entities-analysis ^
        --language-code en ^
        --region aws-region
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。
- *role-arn* 是另存为 comprehend-role-arn 的 ARN。
- *aws-re* gion 就是您的区域。 AWS

4. 复制实体分析 JobId 并在文本编辑器中将其另存为 comprehend-job-id。JobId 可帮助您跟踪实体的分析作业的状态。
5. 要跟踪实体分析任务的进度，请使用以下describe-entities-detection-job命令：

Linux

```
aws comprehend describe-entities-detection-job \
        --job-id entities-job-id \
        --region aws-region
```

其中：

- *entities-job-id*你得救了吗comprehend-job-id，
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws comprehend describe-entities-detection-job \
        --job-id entities-job-id \
        --region aws-region
```

其中：

- *entities-job-id*你得救了吗comprehend-job-id，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws comprehend describe-entities-detection-job ^
        --job-id entities-job-id ^
        --region aws-region
```

其中：

- *entities-job-id*你得救了吗comprehend-job-id，
- *aws-re* gion 就是您的区域。 AWS

将 JobStatus 更改为 COMPLETED 需要几分钟。

在此步骤结束时，Amazon Comprehend 将实体分析结果作为压缩的 output.tar.gz 文件存储在 S3 存储桶中自动生成的文件夹内的 output 文件夹中。确保您的分析作业状态已完成，然后再继续下一步。

步骤 3：将实体分析输出格式化为 Amazon Kendra 元数据

要将 Amazon Comprehend 提取的实体转换为 Amazon Kendra 索引所需的元数据格式，您需要运行 Python 3 脚本。转换结果存储在 Amazon S3 存储桶中的 metadata 文件夹中。

有关 Amazon Kendra 元数据格式和结构的更多信息，请参阅 S3 文档元数据。

主题

- 下载和提取 Amazon Comprehend 的输出
- 将输出上传到 S3 存储桶
- 将输出转换为 Amazon Kendra 元数据格式
- 清理 Amazon S3 存储桶

下载和提取 Amazon Comprehend 的输出

要格式化 Amazon Comprehend 实体分析输出，您必须先下载 Amazon Comprehend 实体分析 output.tar.gz 档案并提取实体分析文件。

下载和解压缩输出文件（控制台）

1. 在 Amazon Comprehend 控制台导航窗格中，导航至分析作业。
2. 选择您的实体分析作业 `data-entities-analysis`。
3. 在输出下，选择输出数据位置旁边显示的链接。这会将您重定向到 S3 存储桶中的 `output.tar.gz` 存档。
4. 在概述选项卡中，选择下载。

> **Tip**
>
> 所有 Amazon Comprehend 分析作业的输出具有相同的名称。重命名存档将帮助您更轻松地对其进行跟踪。

5. 将下载的 Amazon Comprehend 文件解压并提取到您的设备上。

下载并解压缩文件（AWS CLI）。

1. 要访问您的 S3 存储桶中包含实体分析任务结果的 Amazon Comprehend 自动生成文件夹的名称，请使用以下命令：describe-entities-detection-job

   Linux

   ```
   aws comprehend describe-entities-detection-job \
             --job-id entities-job-id \
             --region aws-region
   ```

   其中：

   - *entities-job-id*你是comprehend-job-id从哪里救出来的the section called “步骤 2：检测实体”，
   - *aws-re* gion 就是您的区域。 AWS

   macOS

   ```
   aws comprehend describe-entities-detection-job \
             --job-id entities-job-id \
             --region aws-region
   ```

其中：

- *entities-job-id*你是comprehend-job-id从哪里救出来的the section called “步骤 2：检测实体”，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws comprehend describe-entities-detection-job ^
          --job-id entities-job-id ^
          --region aws-region
```

其中：

- *entities-job-id*你是comprehend-job-id从哪里救出来的the section called “步骤 2：检测实体”，
- *aws-re* gion 就是您的区域。 AWS

2. 在实体作业描述中的 OutputDataConfig 对象中，复制 S3Uri 值并将其保存为文本编辑器中的 comprehend-S3uri。

> **Note**
>
> 该S3Uri值的格式类似于 *s3://DOC-EXAMPLE-BUCKET/...* /output/output.tar.gz。

3. 要下载实体输出存档，请使用 copy 命令：

Linux

```
aws s3 cp s3://DOC-EXAMPLE-BUCKET/.../output/output.tar.gz path/output.tar.gz
```

其中：

- *s3://DOC-EXAMPLE-BUCKET/...* /output/output.tar.gz 是你另存为的S3Uri值comprehend-S3uri，
- *path/* 是要保存输出的本地目录。

macOS

```
aws s3 cp s3://DOC-EXAMPLE-BUCKET/.../output/output.tar.gz path/output.tar.gz
```

其中：

- *s3://DOC-EXAMPLE-BUCKET/...* /output/output.tar.gz 是你另存为的S3Uri值comprehend-S3uri，
- *path/* 是要保存输出的本地目录。

Windows

```
aws s3 cp s3://DOC-EXAMPLE-BUCKET/.../output/output.tar.gz path/output.tar.gz
```

其中：

- *s3://DOC-EXAMPLE-BUCKET/...* /output/output.tar.gz 是你另存为的S3Uri值comprehend-S3uri，
- *path/* 是要保存输出的本地目录。

4. 要提取实体输出，请在终端窗口中运行以下命令：

Linux

```
tar -xf path/output.tar.gz -C path/
```

其中：

- *path/* 是本地设备上已下载 output.tar.gz 存档的文件路径。

macOS

```
tar -xf path/output.tar.gz -C path/
```

其中：

- *path/* 是本地设备上已下载 output.tar.gz 存档的文件路径。

Windows

```
tar -xf path/output.tar.gz -C path/
```

其中：

- *path/* 是本地设备上已下载 output.tar.gz 存档的文件路径。

完成此步骤后，您的设备上应该有一个名为 output 的文件，其中包含已识别的 Amazon Comprehend 实体的列表。

将输出上传到 S3 存储桶

下载并解压 Amazon Comprehend 实体分析文件后，您可以将提取的 output 文件上传到您的 Amazon S3 存储桶。

上传提取的 Amazon Comprehend 输出文件（控制台）

1. 打开 Amazon S3 控制台，网址为：https://console.aws.amazon.com/s3/。
2. 在存储桶中，单击存储桶的名称，然后选择上传。
3. 在文件和文件夹中，选择添加文件。
4. 在对话框中，导航到设备中提取的 output 文件，将其选中，然后选择打开。
5. 保留目标、权限和属性的默认设置。
6. 选择上传。

上传提取的 Amazon Comprehend 输出文件（AWS CLI）

1. 要将提取的 output 文件上传到您的存储桶，请使用 copy 命令：

   Linux

   ```
   aws s3 cp path/output s3://DOC-EXAMPLE-BUCKET/output
   ```

   其中：

   - *path/* 是您提取 output 的文件的本地文件路径，

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

macOS

```
aws s3 cp path/output s3://DOC-EXAMPLE-BUCKET/output
```

其中：

- *path/* 是您提取 output 的文件的本地文件路径，
- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

Windows

```
aws s3 cp path/output s3://DOC-EXAMPLE-BUCKET/output
```

其中：

- *path/* 是您提取 output 的文件的本地文件路径，
- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

2. 为确保 output 文件已成功上传到您的 S3 存储桶，请使用 list 命令检查其内容：

Linux

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

macOS

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

Windows

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

将输出转换为 Amazon Kendra 元数据格式

要将 Amazon Comprehend 输出转换为 Amazon Kendra 元数据，您需要运行 Python 3 脚本。如果您使用的是控制台，则使用 AWS CloudShell 此步骤。

运行 Python 3 脚本（控制台）

1. 将 converter.py.zip 下载到您的设备上。
2. 提取 Python 3 文件 `converter.py`。
3. 登录AWS 管理控制台，确保将您的 AWS 区域设置为与 S3 存储桶和 Amazon Comprehend 分析任务相同的区域。
4. 选择AWS CloudShell 图标或AWS CloudShell在顶部导航栏的搜索框中键入以启动环境。

> **Note**
>
> 首次在新的浏览器窗口中 AWS CloudShell 启动时，将显示一个欢迎面板并列出主要功能。关闭此面板后，表示 Shell 已经准备就绪，可以进行交互。

5. 终端准备就绪后，从导航窗格中选择操作，然后从菜单中选择上传文件。
6. 在打开的对话框中点击选择文件，然后从您的设备中选择下载的 Python 3 文件 `converter.py`。选择上传。
7. 在 AWS CloudShell 环境中，输入以下命令：

```
python3 converter.py
```

8. 当 Shell 界面提示您输入 S3 存储桶的名称时，输入您的 S3 存储桶的名称并按 Enter。
9. 当 Shell 界面提示您输入 Comprehend 输出文件的完整文件路径时，输入 **`output`** 并按 Enter。

10. 当 Shell 界面提示您输入元数据文件夹的完整文件路径时，输入 **`metadata/`** 并按 Enter。

> **Important**
>
> 要使元数据格式正确，步骤 8-10 中的输入值必须精确。

运行 Python 3 脚本（AWS CLI）

1. 要下载 Python 3 文件 `converter.py`，请在终端窗口中运行以下命令：

   Linux

   ```
   curl -o path/converter.py.zip https://docs.aws.amazon.com/kendra/latest/dg/
   samples/converter.py.zip
   ```

   其中：

   - *`path/`* 是您要将压缩文件保存到的位置的文件路径。

   macOS

   ```
   curl -o path/converter.py.zip https://docs.aws.amazon.com/kendra/latest/dg/
   samples/converter.py.zip
   ```

   其中：

   - *`path/`* 是您要将压缩文件保存到的位置的文件路径。

   Windows

   ```
   curl -o path/converter.py.zip https://docs.aws.amazon.com/kendra/latest/dg/
   samples/converter.py.zip
   ```

   其中：

   - *`path/`* 是您要将压缩文件保存到的位置的文件路径。

2. 要提取 Python 3 文件，请在终端窗口中运行以下命令：

Linux

```
unzip path/converter.py.zip -d path/
```

其中：

- *path/* 是您保存的 converter.py.zip 文件的路径。

macOS

```
unzip path/converter.py.zip -d path/
```

其中：

- *path/* 是您保存的 converter.py.zip 文件的路径。

Windows

```
tar -xf path/converter.py.zip -C path/
```

其中：

- *path/* 是您保存的 converter.py.zip 文件的路径。

3. 通过运行以下命令确保已将 Boto3 安装在您的设备上。

Linux

```
pip3 show boto3
```

macOS

```
pip3 show boto3
```

Windows

```
pip3 show boto3
```

> **Note**
>
> 如果尚未安装 Boto3，请运行 `pip3 install boto3` 进行安装。

4. 要运行 Python 3 脚本以转换 `output` 文件，请运行以下命令。

   Linux

   ```
   python path/converter.py
   ```

   其中：

   - *path/* 是您保存的 `converter.py.zip` 文件的路径。

   macOS

   ```
   python path/converter.py
   ```

   其中：

   - *path/* 是您保存的 `converter.py.zip` 文件的路径。

   Windows

   ```
   python path/converter.py
   ```

   其中：

   - *path/* 是您保存的 `converter.py.zip` 文件的路径。

5. 当 AWS CLI 提示您输入时`Enter the name of your S3 bucket`，输入您的 S3 存储桶的名称，然后按 Enter。
6. 当 AWS CLI 提示您这样做时`Enter the full filepath to your Comprehend output file`，输入**`output`**并按 Enter。
7. 当 AWS CLI 提示您这样做时`Enter the full filepath to your metadata folder`，输入**`metadata/`**并按 Enter。

> **Important**
>
> 要使元数据格式正确，步骤 5-7 中的输入值必须精确。

在此步骤结束时，格式化的元数据将存放在您的 S3 存储桶的 `metadata` 文件夹中。

清理 Amazon S3 存储桶

由于 Amazon Kendra 索引会同步存储在存储桶中的所有文件，因此我们建议您清理 Amazon S3 存储桶，以防止出现冗余的搜索结果。

清理 Amazon S3 存储桶（控制台）。

1. 打开 Amazon S3 控制台，网址为：https://console.aws.amazon.com/s3/。
2. 在存储桶中，选择您的存储桶，然后选择 Amazon Comprehend 实体分析输出文件夹、Amazon Comprehend 实体分析 `.temp` 文件和提取的 Amazon Comprehend `output` 文件。
3. 从概览选项卡中选择删除。
4. 在删除对象中，选择永久删除对象？，然后在文本输入字段中输入 **`permanently delete`**。
5. 选择删除对象。

清理 Amazon S3 存储桶（AWS CLI）

1. 要删除 S3 存储桶中除 `data` 和 `metadata`文件夹之外的所有文件，请在 AWS CLI中使用 remove 命令：

   Linux

   ```
   aws s3 rm s3://DOC-EXAMPLE-BUCKET/ --recursive --exclude "data/*" --exclude
    "metadata/*"
   ```

   其中：

   - DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

macOS

```
aws s3 rm s3://DOC-EXAMPLE-BUCKET/ --recursive --exclude "data/*" --exclude
 "metadata/*"
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

Windows

```
aws s3 rm s3://DOC-EXAMPLE-BUCKET/ --recursive --exclude "data/*" --exclude
 "metadata/*"
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

2. 为确保已成功从您的 S3 存储桶删除对象，请使用 list 命令检查其内容：

Linux

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

macOS

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

Windows

```
aws s3 ls s3://DOC-EXAMPLE-BUCKET/
```

其中：

- DOC-EXAMPLE-BUCKET 是 S3 存储桶的名称。

在本步骤结束时，您已将 Amazon Comprehend 实体分析输出转换为 Amazon Kendra 元数据。现在，您可以创建 Amazon Kendra 索引。

步骤 4：创建 Amazon Kendra 索引并提取元数据

要实施您的智能搜索解决方案，您需要创建 Amazon Kendra 索引并将您的 S3 数据和元数据提取到该索引中。

在向 Amazon Kendra 索引添加元数据之前，您需要创建与自定义文档属性相对应的自定义索引字段，这些字段又对应于 Amazon Comprehend 实体类型。Amazon Kendra 使用您创建的索引字段和自定义文档属性来搜索和筛选您的文档。

有关更多信息，请参阅索引和创建自定义文档属性。

主题

- 创建 Amazon Kendra 索引
- 更新 Amazon S3 访问的 IAM 角色
- 创建 Amazon Kendra 自定义搜索索引字段
- 添加 Amazon S3 存储桶作为索引的数据来源
- 同步 Amazon Kendra 索引

创建 Amazon Kendra 索引

要查询您的源文档，您需要创建 Amazon Kendra 索引。

如果您使用的是本步骤中的 AWS CLI ，则可以创建并附加一个 AWS IAM 角色和策略，允许 Amazon Kendra 在创建索引之前访问您的 CloudWatch 日志。有关更多信息，请参阅先决条件。

创建 Amazon Kendra 索引（控制台）

1. 打开 Amazon SES 控制台，URL 为：https://console.aws.amazon.com/sesv2/。

> **Important**
>
> 确保您所在的区域与您创建 Amazon Comprehend 实体分析任务和 Amazon S3 存储桶所在的区域相同。如果您在其他区域，请从顶部导航栏的 AWS 区域选择器中选择您创建 Amazon S3 存储桶的区域。

2. 选择创建索引。
3. 要在指定索引详细信息页面上查看索引详细信息，请执行以下操作：
   a. 对于 Index name (索引名称)，输入 **kendra-index**。
   b. 将描述字段留空。
   c. 对于 IAM Role (IAM 角色)，选择 Create a new role (创建新角色)。该角色提供对 Amazon S3 存储桶的访问权限。
   d. 对于角色名称，输入 **kendra-role**。IAM 角色将带有前缀 AmazonKendra-。
   e. 将加密和标签保留默认设置，然后选择下一步。
4. 对于配置用户访问控制页面上的访问控制设置，选择否，然后选择下一步。
5. 对于预配详细信息页面上的预配版本，请选择开发者版本并选择创建。

创建 Amazon Kendra 索引（AWS CLI）

1. 要为 Amazon Kendra 创建并附加将其识别为可信实体的 IAM 角色，请执行以下操作：
   a. 在本地设备上的文本编辑器中，将以下信任策略另存为名为 kendra-trust-policy.json 的 JSON 文件。

```
{
    "Version": "2012-10-17",
    "Statement": {
        "Effect": "Allow",
        "Principal": {
            "Service": "kendra.amazonaws.com"
        },
        "Action": "sts:AssumeRole"
    }
```

```
}
```

b. 要创建名为 kendra-role 的 IAM 角色，并将您保存的 kendra-trust-policy.json 文件附加到该角色上，请使用 create-role 命令：

Linux

```
aws iam create-role \
          --role-name kendra-role \
          --assume-role-policy-document file://path/kendra-trust-policy.json
```

其中：

- *path/* 是本地设备上 kendra-trust-policy.json 的文件路径。

macOS

```
aws iam create-role \
          --role-name kendra-role \
          --assume-role-policy-document file://path/kendra-trust-policy.json
```

其中：

- *path/* 是本地设备上 kendra-trust-policy.json 的文件路径。

Windows

```
aws iam create-role ^
          --role-name kendra-role ^
          --assume-role-policy-document file://path/kendra-trust-policy.json
```

其中：

- *path/* 是本地设备上 kendra-trust-policy.json 的文件路径。

c. 将 Amazon Resource Name（ARN）复制到您的文本编辑器中，并将其作为 kendra-role-arn 保存到本地。

> **Note**
>
> ARN 的格式类似于 *arn:aws:iam::123456789012:role/kendra-role*。您需要保存为 kendra-role-arn 的 ARN 才能运行 Amazon Kendra 作业。

2. 在创建索引之前，必须提供写入 CloudWatch 日志的权限。kendra-role为此，请完成以下步骤：

   a. 在本地设备上的文本编辑器中，将以下信任策略另存为名为 kendra-cloudwatch-policy.json 的 JSON 文件。

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Action":"cloudwatch:PutMetricData",
         "Resource":"*",
         "Condition":{
            "StringEquals":{
               "cloudwatch:namespace":"Kendra"
            }
         }
      },
      {
         "Effect":"Allow",
         "Action":"logs:DescribeLogGroups",
         "Resource":"*"
      },
      {
         "Effect":"Allow",
         "Action":"logs:CreateLogGroup",
         "Resource":"arn:aws:logs:aws-region:aws-account-id:log-group:/aws/
kendra/*"
      },
      {
         "Effect":"Allow",
         "Action":[
            "logs:DescribeLogStreams",
            "logs:CreateLogStream",
            "logs:PutLogEvents"
```

```
        ],
        "Resource":"arn:aws:logs:aws-region:aws-account-id:log-group:/aws/
kendra/*:log-stream:*"
      }
   ]
}
```

将 *aw* s-region 替换 AWS 为您所在的地区*aws-account-id*和 12 位数的账户 ID。 AWS

b. 要创建访问 CloudWatch 日志的 IAM 策略，请使用 create-policy 命令：

Linux

```
aws iam create-policy \
         --policy-name kendra-cloudwatch-policy \
         --policy-document file://path/kendra-cloudwatch-policy.json
```

其中：

- *path/* 是本地设备上 kendra-cloudwatch-policy.json 的文件路径。

macOS

```
aws iam create-policy \
         --policy-name kendra-cloudwatch-policy \
         --policy-document file://path/kendra-cloudwatch-policy.json
```

其中：

- *path/* 是本地设备上 kendra-cloudwatch-policy.json 的文件路径。

Windows

```
aws iam create-policy ^
         --policy-name kendra-cloudwatch-policy ^
         --policy-document file://path/kendra-cloudwatch-policy.json
```

其中：

- *path/* 是本地设备上 kendra-cloudwatch-policy.json 的文件路径。

c. 将 Amazon Resource Name（ARN）复制到您的文本编辑器中，并将其作为 `kendra-cloudwatch-arn` 保存到本地。

> **Note**
>
> ARN 的格式类似于 arn: aws: iam:: 123456789012*: role/*。kendra-cloudwatch-policy您需要保存的 ARN 才能将 `kendra-cloudwatch-arn` 附加到您的 `kendra-cloudwatch-policy` IAM 角色。

d. 要将附加`kendra-cloudwatch-policy`到您的 IAM 角色，请使用以下attach-role-policy命令：

Linux

```
aws iam attach-role-policy \
          --policy-arn policy-arn \
          --role-name kendra-role
```

其中：

- *policy-arn* 是您保存的 `kendra-cloudwatch-arn`。

macOS

```
aws iam attach-role-policy \
          --policy-arn policy-arn \
          --role-name kendra-role
```

其中：

- *policy-arn* 是您保存的 `kendra-cloudwatch-arn`。

Windows

```
aws iam attach-role-policy ^
          --policy-arn policy-arn ^
          --role-name kendra-role
```

其中：

- *policy-arn* 是您保存的 kendra-cloudwatch-arn。

3. 要创建索引，请使用 create-index 命令：

Linux

```
aws kendra create-index \
        --name kendra-index \
        --edition DEVELOPER_EDITION \
        --role-arn role-arn \
        --region aws-region
```

其中：

- *role-arn* 是您保存的 kendra-role-arn，
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra create-index \
        --name kendra-index \
        --edition DEVELOPER_EDITION \
        --role-arn role-arn \
        --region aws-region
```

其中：

- *role-arn* 是您保存的 kendra-role-arn，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra create-index ^
        --name kendra-index ^
        --edition DEVELOPER_EDITION ^
        --role-arn role-arn ^
        --region aws-region
```

其中：

- *role-arn* 是您保存的 kendra-role-arn，
- *aws-re* gion 就是您的区域。 AWS

4. 复制索引 Id 并将其作为 kendra-index-id 保存在文本编辑器中。Id 可帮助您跟踪索引创建的状态。
5. 要跟踪索引创建任务的进度，请使用 describe-index 命令：

Linux

```
aws kendra describe-index \
        --id kendra-index-id \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra describe-index \
        --id kendra-index-id \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra describe-index ^
        --id kendra-index-id ^
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

索引创建过程平均需要 15 分钟，但可能需要更长的时间。当索引的状态为活动时，您的索引就已经准备就绪。在创建索引的同时，您可以开始下一步。

如果您在此步骤 AWS CLI 中使用的是，则可以创建一个 IAM 策略并将其附加到您的 Amazon Kendra IAM 角色，该策略授予您的索引访问您的 S3 存储桶的权限。

更新 Amazon S3 访问的 IAM 角色

在创建索引的同时，您可以更新您的 Amazon Kendra IAM 角色以允许您创建的索引从 Amazon S3 存储桶中读取数据。有关更多信息，请参阅 Amazon Kendra 的 IAM 访问角色。

更新 IAM 角色（控制台）

1. 通过 https://console.aws.amazon.com/iam/ 打开 IAM 控制台。
2. 在左侧导航窗格中，选择角色，然后在角色名称上方的搜索框中输入 **kendra-role**。
3. 在建议的选项中，单击kendra-role。
4. 在摘要中，选择附加策略。
5. 在附加权限的搜索框中，从建议的选项中输入**S3**并选中 AmazonS3 ReadOnlyAccess 策略旁边的复选框。
6. 选择附加策略。现在，在摘要页面上，您将看到两个附加到 IAM 角色的策略。
7. 返回 Amazon Kendra 控制台，URL 为 https://console.aws.amazon.com/kendra/，并等待索引的状态从正在创建变成活动，然后再继续下一步。

更新 IAM 角色（AWS CLI）

1. 在本地设备上的文本编辑器中，将以下文本另存为名为 kendra-S3-access-policy.json 的 JSON 文件。

```
{
   "Version":"2012-10-17",
   "Statement":[
```

```
      {
         "Action":[
            "s3:GetObject"
         ],
         "Resource":[
            "arn:aws:s3:::DOC-EXAMPLE-BUCKET/*"
         ],
         "Effect":"Allow"
      },
      {
         "Action":[
            "s3:ListBucket"
         ],
         "Resource":[
            "arn:aws:s3:::DOC-EXAMPLE-BUCKET"
         ],
         "Effect":"Allow"
      },
      {
         "Effect":"Allow",
         "Action":[
            "kendra:BatchPutDocument",
            "kendra:BatchDeleteDocument",
            "kendra:ListDataSourceSyncJobs"
         ],
         "Resource":[
            "arn:aws:kendra:aws-region:aws-account-id:index/kendra-index-id"
         ]
      }
   ]
}
```

将 DOC-EXAMPLE-BUCKET 替换为您的 S3 存储桶名称，将 aws-region 替换为您*#### AWS #*，将 12 AWS 位数的账户 ID *aws-account-id*替换为已保存的。*kendra-index-id*kendra-index-id

2. 要创建 IAM policy 以访问 S3 桶，请使用 create-policy 命令：

   Linux

```
aws iam create-policy \
          --policy-name kendra-S3-access-policy \
          --policy-document file://path/kendra-S3-access-policy.json
```

其中：

- *path/* 是本地设备上 kendra-S3-access-policy.json 的文件路径。

macOS

```
aws iam create-policy \
          --policy-name kendra-S3-access-policy \
          --policy-document file://path/kendra-S3-access-policy.json
```

其中：

- *path/* 是本地设备上 kendra-S3-access-policy.json 的文件路径。

Windows

```
aws iam create-policy ^
          --policy-name kendra-S3-access-policy ^
          --policy-document file://path/kendra-S3-access-policy.json
```

其中：

- *path/* 是本地设备上 kendra-S3-access-policy.json 的文件路径。

3. 将 Amazon Resource Name（ARN）复制到您的文本编辑器中，并将其作为 kendra-S3-access-arn 保存到本地。

> **Note**
>
> 该 ARN 的格式类似于 *arn:aws:iam::123456789012:role/kendra-S3-access-policy*。您需要保存的 ARN 才能将 kendra-S3-access-arn 附加到您的 kendra-S3-access-policy IAM 角色。

4. 要将附加kendra-S3-access-policy到您的 Amazon Kendra IAM 角色，请使用以下attach-role-policy命令：

Linux

```
aws iam attach-role-policy \
```

```
          --policy-arn policy-arn \
          --role-name kendra-role
```

其中：

- *policy-arn* 是您保存的 kendra-S3-access-arn。

macOS

```
aws iam attach-role-policy \
          --policy-arn policy-arn \
          --role-name kendra-role
```

其中：

- *policy-arn* 是您保存的 kendra-S3-access-arn。

Windows

```
aws iam attach-role-policy ^
          --policy-arn policy-arn ^
          --role-name kendra-role
```

其中：

- *policy-arn* 是您保存的 kendra-S3-access-arn。

创建 Amazon Kendra 自定义搜索索引字段

要让 Amazon Kendra 做好将您的元数据识别为自定义文档属性的准备，您需要创建与 Amazon Comprehend 实体类型对应的自定义字段。您输入以下九种 Amazon Comprehend 实体类型作为自定义字段：

- COMMERCIAL_ITEM
- DATE
- EVENT
- LOCATION

- ORGANIZATION
- OTHER
- PERSON
- QUANTITY
- TITLE

> **Important**
>
> 索引将无法识别拼写错误的实体类型。

为您的 Amazon Kendra 索引创建自定义字段（控制台）

1. 打开 Amazon SES 控制台，URL 为：https://console.aws.amazon.com/sesv2/。
2. 从索引列表中，单击`kendra-index`。
3. 在左侧导航面板的数据管理下，选择分面定义。
4. 从索引字段菜单中，选择添加字段。
5. 在添加索引字段对话框中，执行以下操作：
   a. 在字段名称中，输入 **`COMMERCIAL_ITEM`**。
   b. 在数据类型中，选择字符串列表。
   c. 在使用类型中，选择可分面、可搜索和可显示，然后选择添加。
   d. 对每种 Amazon Comprehend 实体类型重复步骤 a 到 c：COMMERCIAL_ITEM、DATE、EVENT、LOCATION、ORGANIZATION、OTHER、PERSON、QU

控制台显示成功添加字段的消息。在继续下一步之前，您可以选择将其关闭。

为您的 Amazon Kendra 索引（AWS CLI）创建自定义字段

1. 在本地设备上的文本编辑器中，将以下文本另存为名为 `custom-attributes.json` 的 JSON 文件。

```
[
   {
       "Name": "COMMERCIAL_ITEM",
       "Type": "STRING_LIST_VALUE",
```

```
        "Search": {
            "Facetable": true,
            "Searchable": true,
            "Displayable": true
        }
    },
    {
        "Name": "DATE",
        "Type": "STRING_LIST_VALUE",
        "Search": {
            "Facetable": true,
            "Searchable": true,
            "Displayable": true
        }
    },
    {
        "Name": "EVENT",
        "Type": "STRING_LIST_VALUE",
        "Search": {
            "Facetable": true,
            "Searchable": true,
            "Displayable": true
        }
    },
    {
        "Name": "LOCATION",
        "Type": "STRING_LIST_VALUE",
        "Search": {
            "Facetable": true,
            "Searchable": true,
            "Displayable": true
        }
    },
    {
        "Name": "ORGANIZATION",
        "Type": "STRING_LIST_VALUE",
        "Search": {
            "Facetable": true,
            "Searchable": true,
            "Displayable": true
        }
    },
    {
        "Name": "OTHER",
```

```
        "Type": "STRING_LIST_VALUE",
        "Search": {
            "Facetable": true,
            "Searchable": true,
            "Displayable": true
        }
    },
    {
        "Name": "PERSON",
        "Type": "STRING_LIST_VALUE",
        "Search": {
            "Facetable": true,
            "Searchable": true,
            "Displayable": true
        }
    },
    {
        "Name": "QUANTITY",
        "Type": "STRING_LIST_VALUE",
        "Search": {
            "Facetable": true,
            "Searchable": true,
            "Displayable": true
        }
    },
    {
        "Name": "TITLE",
        "Type": "STRING_LIST_VALUE",
        "Search": {
            "Facetable": true,
            "Searchable": true,
            "Displayable": true
        }
    }
]
```

2. 要在索引中创建自定义字段，请使用 update-index 命令：

   Linux

```
aws kendra update-index \
        --id kendra-index-id \
        --document-metadata-configuration-updates file://path/custom-
attributes.json \
```

```
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *path/* 是本地设备上 custom-attributes.json 的文件路径。
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra update-index \
        --id kendra-index-id \
        --document-metadata-configuration-updates file://path/custom-
attributes.json \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *path/* 是本地设备上 custom-attributes.json 的文件路径。
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra update-index ^
        --id kendra-index-id ^
        --document-metadata-configuration-updates file://path/custom-
attributes.json ^
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *path/* 是本地设备上 custom-attributes.json 的文件路径。
- *aws-re* gion 就是您的区域。 AWS

3. 要验证自定义属性是否已添加到您的索引中，请使用 describe-index 命令：

Linux

```
aws kendra describe-index \
        --id kendra-index-id \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra describe-index \
        --id kendra-index-id \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra describe-index ^
        --id kendra-index-id ^
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

添加 Amazon S3 存储桶作为索引的数据来源

必须先将 S3 数据来源连接到索引，然后才能同步索引。

将 S3 存储桶连接到您的 Amazon Kendra 索引（控制台）

1. 打开 Amazon SES 控制台，URL 为：https://console.aws.amazon.com/sesv2/。
2. 从索引列表中，单击kendra-index。
3. 在左侧导航菜单的数据管理下，选择数据来源。
4. 在选择数据来源连接器类型部分下，导航到 Amazon S3，然后选择添加连接器。
5. 在指定数据来源详细信息页面中，执行以下操作：
   a. 在名称和描述下，对于数据来源名称，输入 **S3-data-source**。
   b. 将描述部分留空。
   c. 保留标签的默认设置。
   d. 选择下一步。
6. 在配置同步设置页面的同步范围部分，执行以下操作：
   a. 在输入数据来源位置中，选择浏览 S3。
   b. 在选择资源中，选择您的 S3 存储桶，然后点击选择。
   c. 在元数据文件前缀文件夹位置中，选择浏览 S3。
   d. 在选择资源中，从存储桶列表中单击存储桶的名称。
   e. 对于对象，选择 metadata 的选项框，然后单击选择。现在，位置字段应该显示 metadata/。
   f. 保留访问控制列表配置文件位置、选择解密密钥和其他配置的默认设置。
7. 对于 IAM 角色，在配置同步设置页面上，选择kendra-role。
8. 在配置同步设置页面的同步运行计划下，选择频率，选择按需运行，然后选择下一步。
9. 在审核和创建页面中，查看数据来源详细信息的选择，然后选择添加数据来源。

将 S3 存储桶连接到您的 Amazon Kendra 索引（AWS CLI）

1. 在本地设备上的文本编辑器中，将以下文本另存为名为 S3-data-connector.json 的 JSON 文件。

```
{
   "S3Configuration":{
      "BucketName":"DOC-EXAMPLE-BUCKET",
      "DocumentsMetadataConfiguration":{
         "S3Prefix":"metadata"
```

```
      }
   }
}
```

将 DOC-EXAMPLE-BUCKET 替换为您的 S3 存储桶的名称。

2. 要将 S3 存储桶连接到索引，请使用以下create-data-source命令：

Linux

```
aws kendra create-data-source \
        --index-id kendra-index-id \
        --name S3-data-source \
        --type S3 \
        --configuration file://path/S3-data-connector.json \
        --role-arn role-arn \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *path/* 是本地设备上 S3-data-connector.json 的文件路径。
- *role-arn* 是您保存的 kendra-role-arn，
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra create-data-source \
        --index-id kendra-index-id \
        --name S3-data-source \
        --type S3 \
        --configuration file://path/S3-data-connector.json \
        --role-arn role-arn \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *path/* 是本地设备上 S3-data-connector.json 的文件路径。

- *role-arn* 是您保存的 kendra-role-arn，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra create-data-source ^
        --index-id kendra-index-id ^
        --name S3-data-source ^
        --type S3 ^
        --configuration file://path/S3-data-connector.json ^
        --role-arn role-arn ^
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *path/* 是本地设备上 S3-data-connector.json 的文件路径。
- *role-arn* 是您保存的 kendra-role-arn，
- *aws-re* gion 就是您的区域。 AWS

3. 复制连接器 Id 并将其作为 S3-connector-id 保存在文本编辑器中。Id 可帮助您跟踪数据连接过程的状态。
4. 要确保您的 S3 数据源已成功连接，请使用以下describe-data-source命令：

Linux

```
aws kendra describe-data-source \
        --id S3-connector-id \
        --index-id kendra-index-id \
        --region aws-region
```

其中：

- *S3-connector-id* 是您保存的 S3-connector-id，
- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra describe-data-source \
        --id S3-connector-id \
        --index-id kendra-index-id \
        --region aws-region
```

其中：

- *S3-connector-id* 是您保存的 S3-connector-id，
- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra describe-data-source ^
        --id S3-connector-id ^
        --index-id kendra-index-id ^
        --region aws-region
```

其中：

- *S3-connector-id* 是您保存的 S3-connector-id，
- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

在此步骤结束时，您的 Amazon S3 数据来源已连接到索引。

同步 Amazon Kendra 索引

添加了 Amazon S3 数据来源后，您现在可以将您的 Amazon Kendra 索引同步到该数据来源。

同步您的 Amazon Kendra 索引（控制台）

1. 打开 Amazon SES 控制台，URL 为：https://console.aws.amazon.com/sesv2/。
2. 从索引列表中，单击kendra-index。

3. 从左侧导航菜单中选择数据来源。
4. 在数据来源中选择 S3-data-source。
5. 从顶部导航栏中，选择立即同步。

同步您的 Amazon Kendra 索引（AWS CLI）

1. 要同步索引，请使用 start-data-source-sync-job 命令：

   Linux

   ```
   aws kendra start-data-source-sync-job \
           --id S3-connector-id \
           --index-id kendra-index-id \
           --region aws-region
   ```

   其中：

   - *S3-connector-id* 是您保存的 S3-connector-id，
   - *kendra-index-id*你得救了吗kendra-index-id，
   - *aws-re* gion 就是您的区域。 AWS

   macOS

   ```
   aws kendra start-data-source-sync-job \
           --id S3-connector-id \
           --index-id kendra-index-id \
           --region aws-region
   ```

   其中：

   - *S3-connector-id* 是您保存的 S3-connector-id，
   - *kendra-index-id*你得救了吗kendra-index-id，
   - *aws-re* gion 就是您的区域。 AWS

   Windows

   ```
   aws kendra start-data-source-sync-job ^
   ```

```
        --id S3-connector-id ^
        --index-id kendra-index-id ^
        --region aws-region
```

其中：

- *S3-connector-id* 是您保存的 S3-connector-id，
- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

2. 要检查索引同步的状态，请使用 list-data-source-sync-jobs 命令：

Linux

```
aws kendra list-data-source-sync-jobs \
        --id S3-connector-id \
        --index-id kendra-index-id \
        --region aws-region
```

其中：

- *S3-connector-id* 是您保存的 S3-connector-id，
- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra list-data-source-sync-jobs \
        --id S3-connector-id \
        --index-id kendra-index-id \
        --region aws-region
```

其中：

- *S3-connector-id* 是您保存的 S3-connector-id，
- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra list-data-source-sync-jobs ^
        --id S3-connector-id ^
        --index-id kendra-index-id ^
        --region aws-region
```

其中：

- *S3-connector-id* 是您保存的 S3-connector-id，
- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

在本步骤结束时，您已经为您的数据集创建了一个可搜索和可筛选的 Amazon Kendra 索引。

步骤 5：查询 Amazon Kendra 索引

您的 Amazon Kendra 索引现在可以进行自然语言查询了。当您搜索索引时，Amazon Kendra 会使用您提供的所有数据和元数据为您的搜索查询返回最准确的答案。

Amazon Kendra 可以回答三种查询：

- 事实类查询（“谁”、“什么”、“何时”或“何处”问题）
- 描述性查询（“如何”问题）
- 关键词搜索（意图和范围不清楚的问题）

主题

查询您的 Amazon Kendra 索引

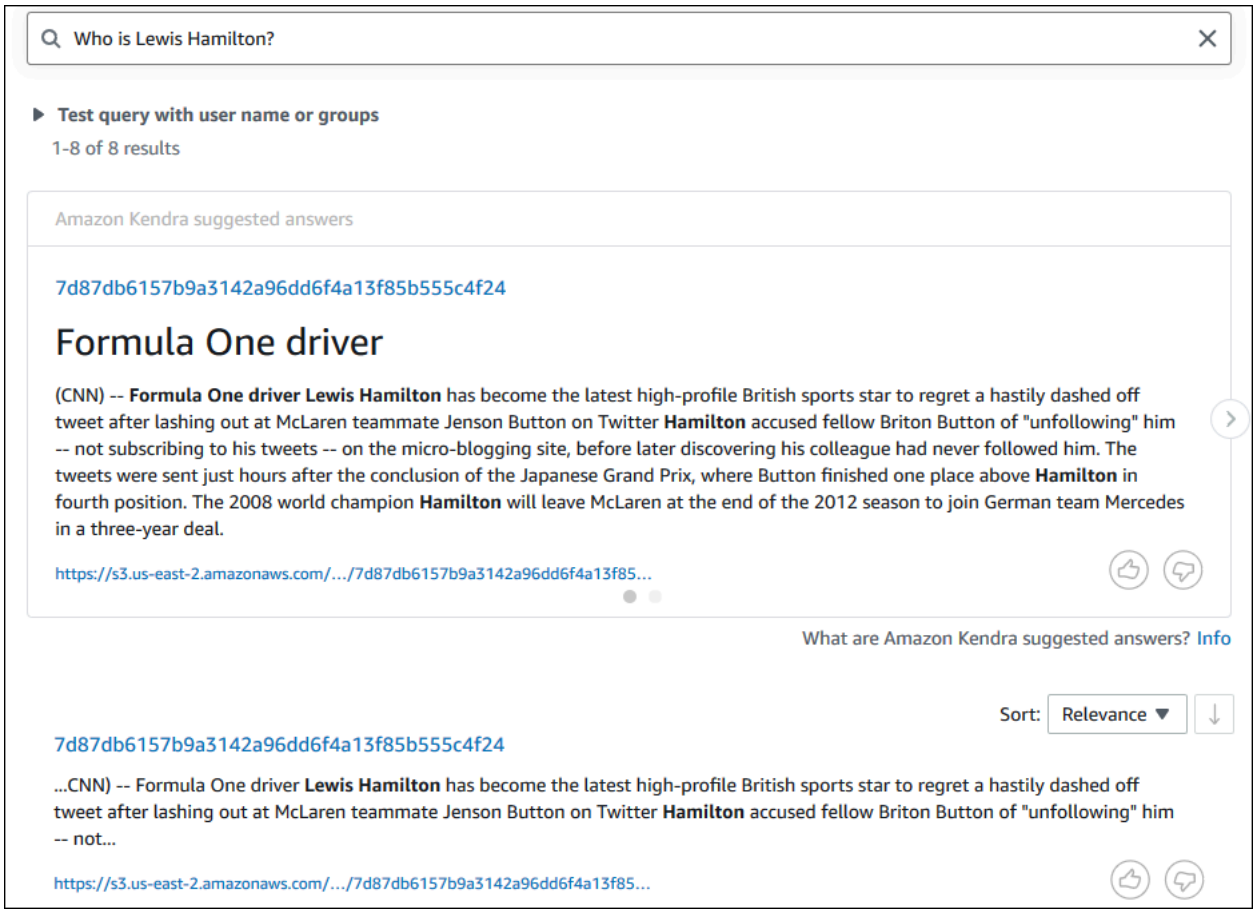
您可以使用与 Amazon Kendra 支持的三种查询相对应的问题来查询您的 Amazon Kendra 索引。有关更多信息，请参阅查询。

本节中的示例问题是根据样本数据集选择的。

查询您的 Amazon Kendra 索引（控制台）

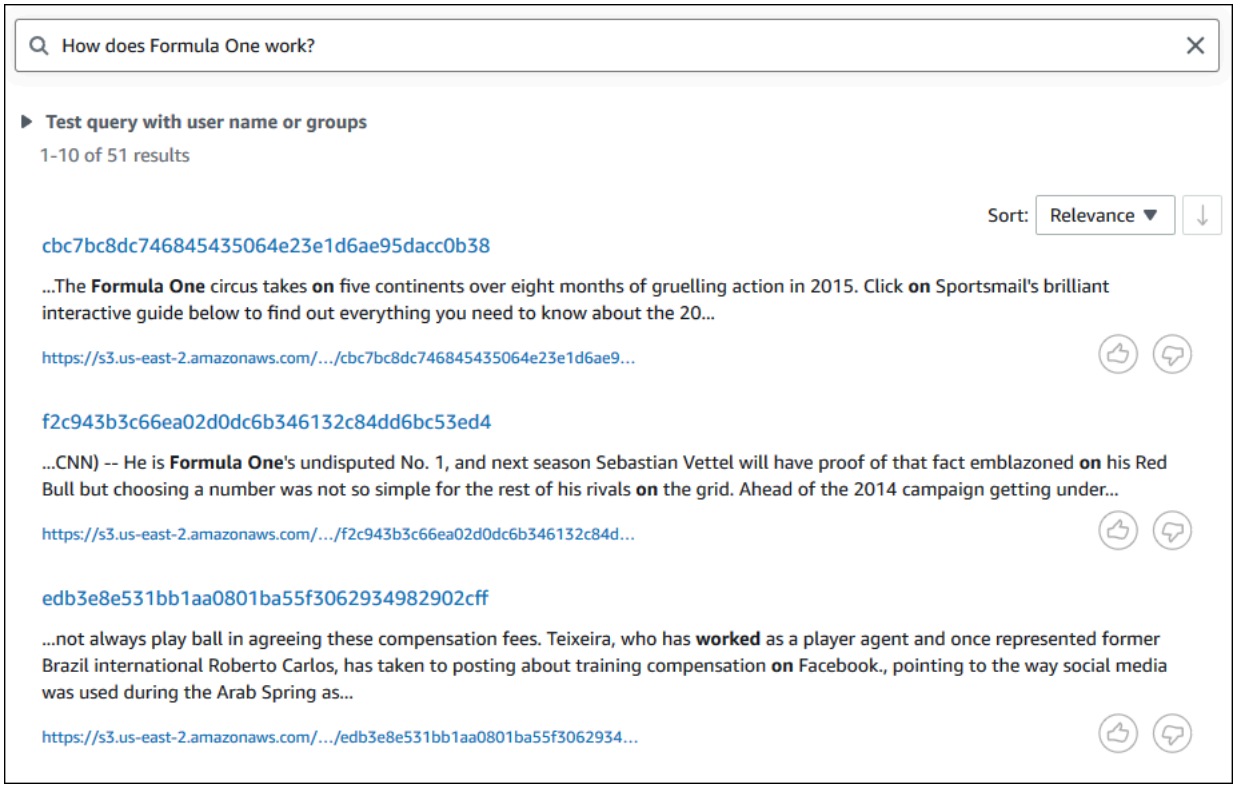
1. 打开 Amazon SES 控制台，URL 为：https://console.aws.amazon.com/sesv2/。
2. 从索引列表中，单击kendra-index。
3. 从左侧导航菜单中，选择搜索索引的选项。
4. 要运行示例事实查询，请在搜索框中输入 **`Who is Lewis Hamilton?`** 并按 Enter。

   返回的第一个结果是 Amazon Kendra 建议的答案以及包含答案的数据文件。其余结果构成一组推荐文档。



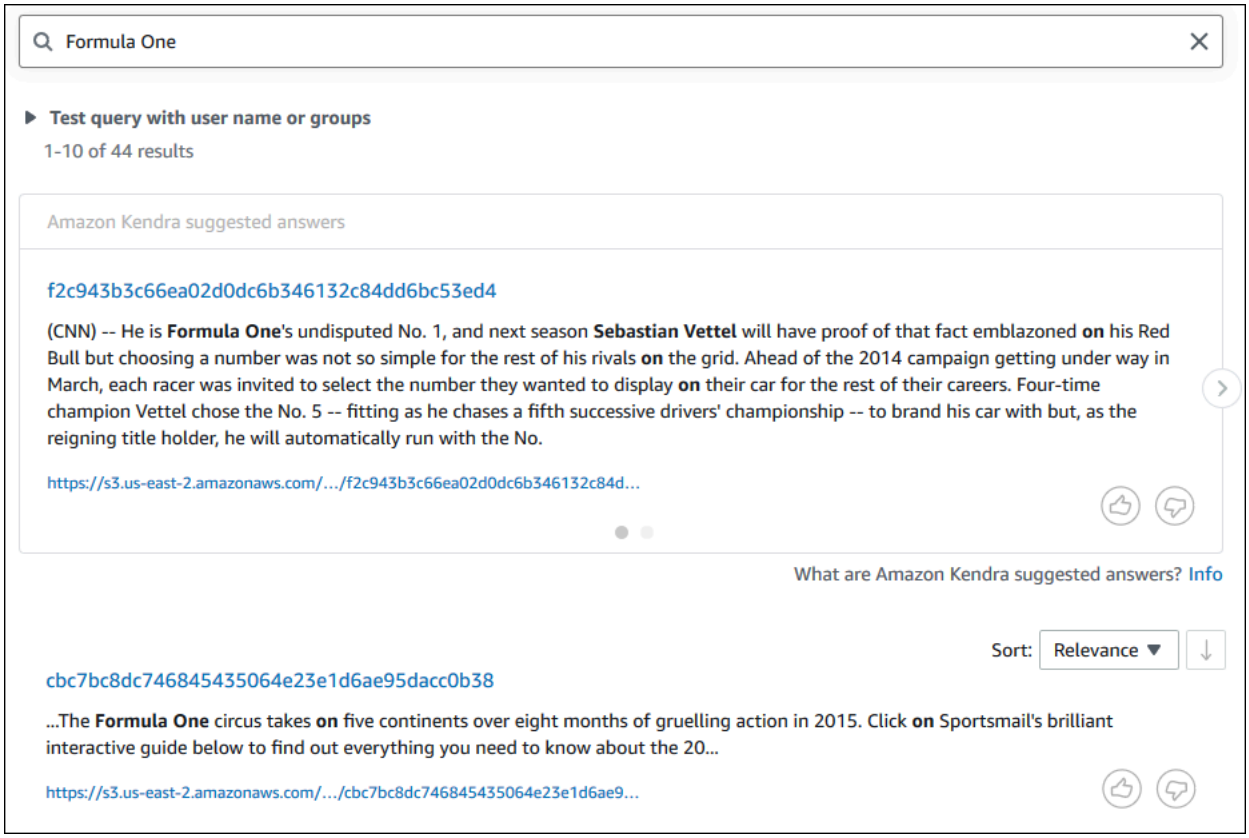
5. 要运行描述性查询，请在搜索框中输入 **`How does Formula One work?`** 并按 Enter。

   您将看到 Amazon Kendra 控制台返回的另一个结果，这次是突出显示了相关的短语。



6. 要进行关键字搜索，请在搜索框中输入 **`Formula One`** 并按 Enter。

   您将看到 Amazon Kendra 控制台返回的另一个结果，然后是数据集中所有其他提及该短语的结果。



查询您的 Amazon Kendra 索引（AWS CLI）

1. 要运行示例事实查询，请使用 query 命令：

   Linux

   ```
   aws kendra query \
           --index-id kendra-index-id \
           --query-text "Who is Lewis Hamilton?" \
           --region aws-region
   ```

   其中：

   - *kendra-index-id*你得救了吗kendra-index-id，
   - *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra query \
        --index-id kendra-index-id \
        --query-text "Who is Lewis Hamilton?" \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra query ^
        --index-id kendra-index-id ^
        --query-text "Who is Lewis Hamilton?" ^
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

AWS CLI 显示您的查询结果。

2. 要运行示例描述性查询，请使用 query 命令：

Linux

```
aws kendra query \
        --index-id kendra-index-id \
        --query-text "How does Formula One work?" \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，

- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra query \
        --index-id kendra-index-id \
        --query-text "How does Formula One work?" \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra query ^
        --index-id kendra-index-id ^
        --query-text "How does Formula One work?" ^
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

AWS CLI 显示您的查询结果。

3. 要运行示例关键字搜索，请使用 query 命令：

Linux

```
aws kendra query \
        --index-id kendra-index-id \
        --query-text "Formula One" \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra query \
        --index-id kendra-index-id \
        --query-text "Formula One" \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra query ^
        --index-id kendra-index-id ^
        --query-text "Formula One" ^
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

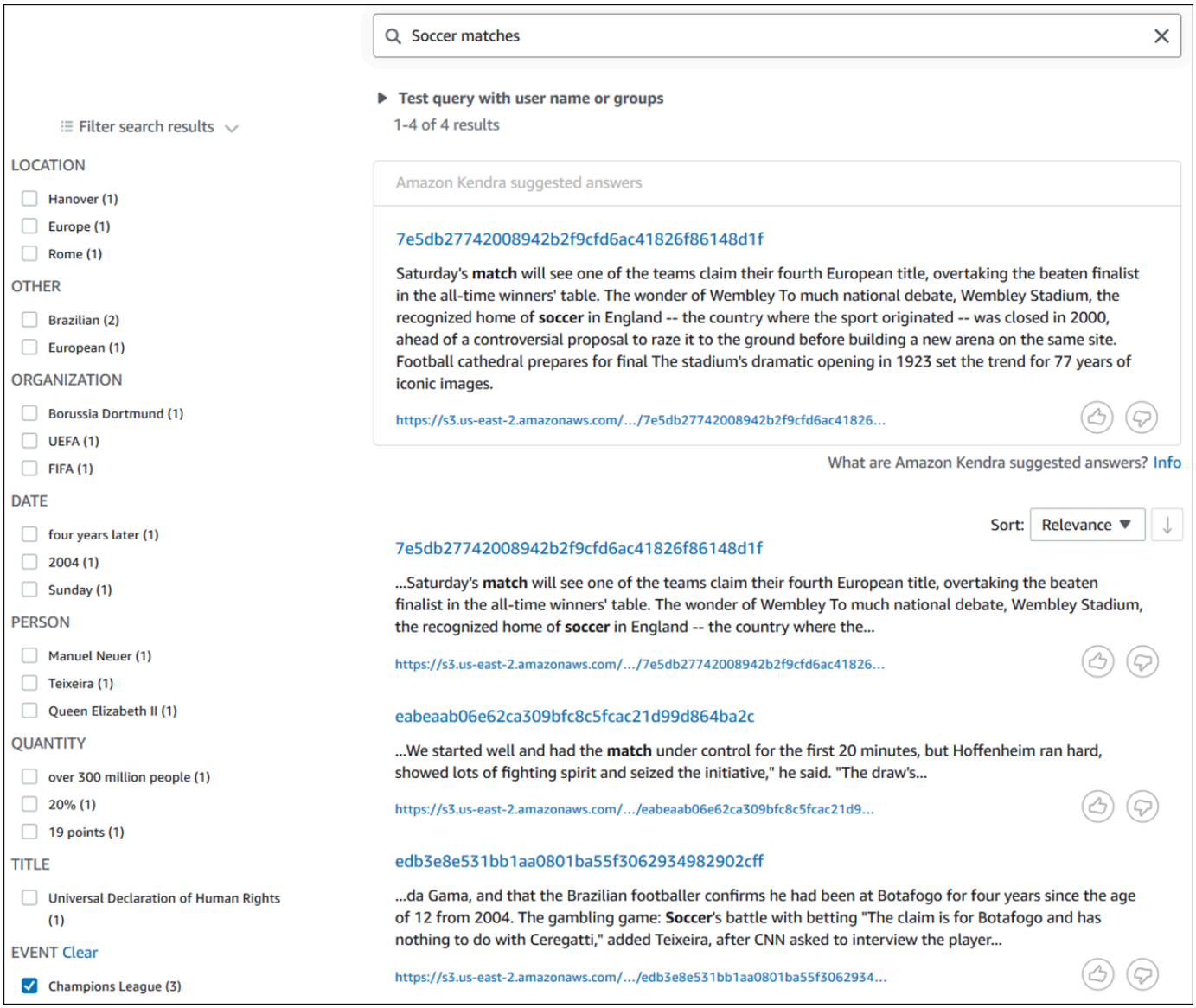
AWS CLI 显示您查询的返回答案。

筛选您的搜索结果

您可以在 Amazon Kendra 控制台中使用自定义文档属性对搜索结果进行筛选和排序。有关 Amazon Kendra 如何处理查询的更多信息，请参阅筛选查询。

筛选搜索结果（控制台）

1. 打开 Amazon SES 控制台，URL 为：https://console.aws.amazon.com/sesv2/。
2. 从索引列表中，单击kendra-index。
3. 从左侧导航菜单中，选择搜索索引的选项。
4. 在搜索框中，输入 **Soccer matches** 作为查询，然后按 Enter。
5. 从左侧导航菜单中，选择筛选搜索结果，查看可用于筛选搜索的分面列表。
6. 选中 EVENT 副标题下的“Champions League”复选框，即可查看仅按包含“Champions League”的结果筛选的搜索结果。



筛选搜索结果（AWS CLI）

1. 要查看可供搜索的特定类型（例如，EVENT）的实体，请使用 query 命令：

   Linux

   ```
   aws kendra query \
           --index-id kendra-index-id \
           --query-text "Soccer matches" \
           --facets '[{"DocumentAttributeKey":"EVENT"}]' \
           --region aws-region
   ```

   其中：

   - *kendra-index-id*你得救了吗kendra-index-id，
   - *aws-re* gion 就是您的区域。 AWS

   macOS

   ```
   aws kendra query \
           --index-id kendra-index-id \
           --query-text "Soccer matches" \
           --facets '[{"DocumentAttributeKey":"EVENT"}]' \
           --region aws-region
   ```

   其中：

   - *kendra-index-id*你得救了吗kendra-index-id，
   - *aws-re* gion 就是您的区域。 AWS

   Windows

   ```
   aws kendra query ^
           --index-id kendra-index-id ^
           --query-text "Soccer matches" ^
           --facets '[{"DocumentAttributeKey":"EVENT"}]' ^
           --region aws-region
   ```

   其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

AWS CLI 显示搜索结果。要获取类型分面的列表EVENT，请导航到 AWS CLI 输出的“FacetResults” 部分，查看可过滤的刻面列表及其计数。例如，其中一个方面是“Champions League”。

> **Note**
>
> 与 EVENT 不同，您可以选择在 the section called “创建 Amazon Kendra 索引” 中为 DocumentAttributeKey 值创建的任何索引字段。

2. 要运行相同的搜索但仅按包含“Champions League”的结果进行筛选，请使用 query 命令：

Linux

```
aws kendra query \
        --index-id kendra-index-id \
        --query-text "Soccer matches" \
        --attribute-filter '{"ContainsAny":{"Key":"EVENT","Value":
{"StringListValue":["Champions League"]}}}' \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

macOS

```
aws kendra query \
        --index-id kendra-index-id \
        --query-text "Soccer matches" \
        --attribute-filter '{"ContainsAny":{"Key":"EVENT","Value":
{"StringListValue":["Champions League"]}}}' \
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

Windows

```
aws kendra query ^
        --index-id kendra-index-id ^
        --query-text "Soccer matches" ^
        --attribute-filter '{"ContainsAny":{"Key":"EVENT","Value":
{"StringListValue":["Champions League"]}}}' ^
        --region aws-region
```

其中：

- *kendra-index-id*你得救了吗kendra-index-id，
- *aws-re* gion 就是您的区域。 AWS

AWS CLI 显示筛选后的搜索结果。

步骤 5：清理

清理文件

要在完成本教程后停止在 AWS 账户中产生费用，您可以采取以下步骤：

1. 删除您的 Amazon S3 存储桶

   有关删除存储桶的信息，请参阅删除存储桶。
2. 删除 Amazon Kendra 索引

   有关删除 Amazon Kendra 存储桶的信息，请参阅删除索引。
3. 删除 **converter.py**
   - 对于控制台：前往 AWS CloudShell，并确保将该区域设置为您的 AWS 区域。加载 bash shell 后，在环境中键入以下命令并按 Enter。

     ```
     rm converter.py
     ```

- For AWS CLI：在终端窗口上运行以下命令。

  Linux

  ```
  rm file/converter.py
  ```

  其中：

  - *file/* 是本地设备上 converter.py 的文件路径。

  macOS

  ```
  rm file/converter.py
  ```

  其中：

  - *file/* 是本地设备上 converter.py 的文件路径。

  Windows

  ```
  rm file/converter.py
  ```

  其中：

  - *file/* 是本地设备上 converter.py 的文件路径。

了解更多信息

要详细了解如何将 Amazon Kendra 集成到您的工作流程中，您可以查看以下博客文章：

- 用于增强搜索效果的内容元数据标记
- 使用自动内容充实功能构建智能搜索解决方案

要了解有关 Amazon Comprehend 的更多信息，请参阅《Amazon Comprehend 开发人员指南》。

Amazon Kendra 的监控和日志记录

主题

监控索引（控制台）

使用 Amazon Kendra 控制台可监控索引和数据来源的状态。您可以使用这些信息来跟踪索引的大小和存储需求，并监控索引和数据来源之间同步的进度和成功状况。

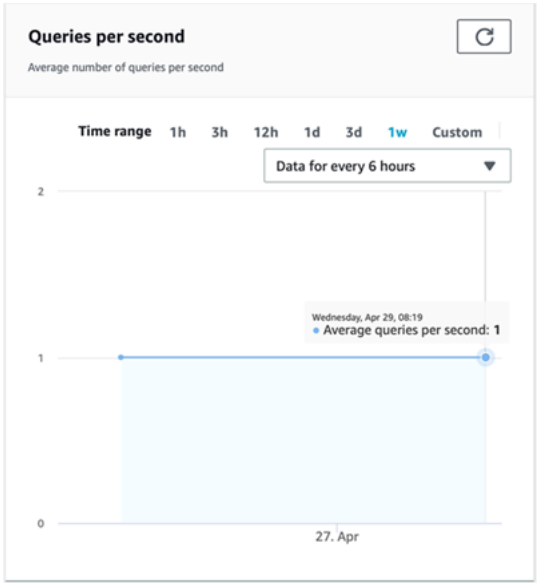
查看索引指标（控制台）

1. 登录 AWS Management Console 并打开 Amazon Kendra 控制台，网址为：https://console.aws.amazon.com/kendra/home。
2. 从索引列表中选择要查看的索引。
3. 滚动屏幕以查看索引指标。

您可以看到有关索引的以下指标。

- 文档数量 – 已编入索引的文档总数量。其中包括来自所有数据来源的所有文档。使用此指标可确定是否需要为索引调整存储单位数量。

- 每秒查询数 – 每秒请求的索引查询数量。使用此指标可确定是否需要为索引调整查询单位数量。



要监控索引与数据来源之间的同步进度和成功状况，请使用 Amazon Kendra 控制台。使用此信息可帮助确定数据来源的运行状况。

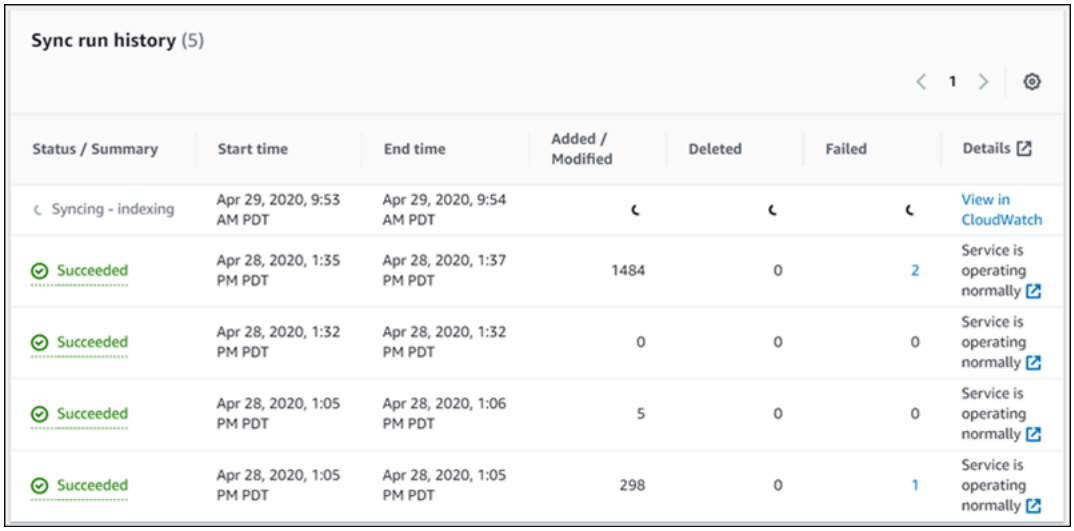
查看同步指标（控制台）

1. 登录 AWS Management Console 并打开 Amazon Kendra 控制台，网址为：https://console.aws.amazon.com/kendra/home。
2. 从索引列表中选择要查看同步指标的索引。

3. 从左侧菜单中选择数据来源。
4. 从数据来源列表中，选择要查看的数据来源。
5. 滚动屏幕以查看同步运行指标。

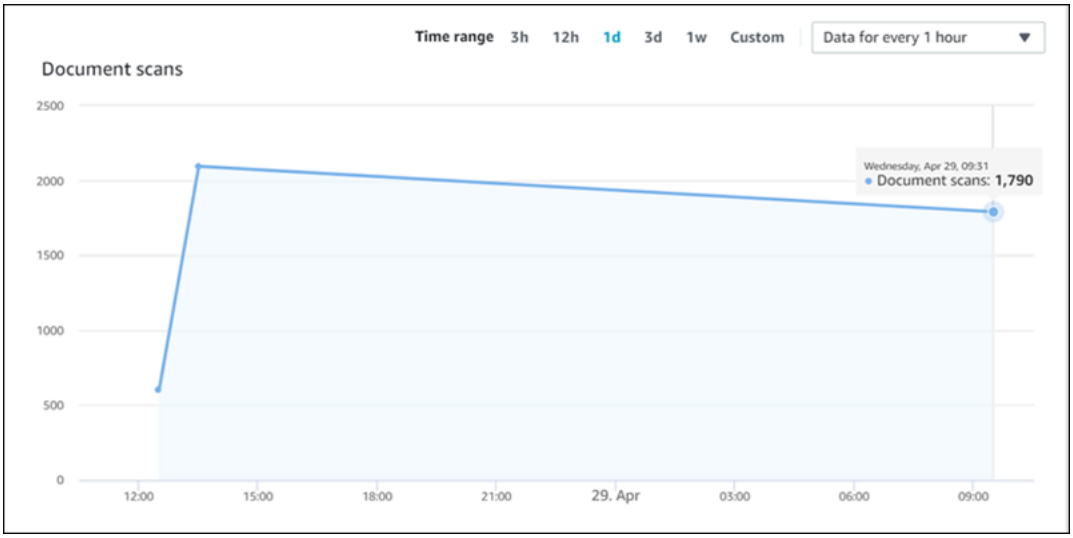
您可以查看以下信息。

- 同步运行历史记录 - 有关同步运行的统计信息，包括开始和结束时间、添加、删除和失败的文档数量。如果同步运行失败，则会有一个指向 CloudWatch Logs 的链接，其中包含更多信息。选择左上角的设置图标可更改历史记录中显示的列。使用此信息可确定数据来源的总体运行状况。

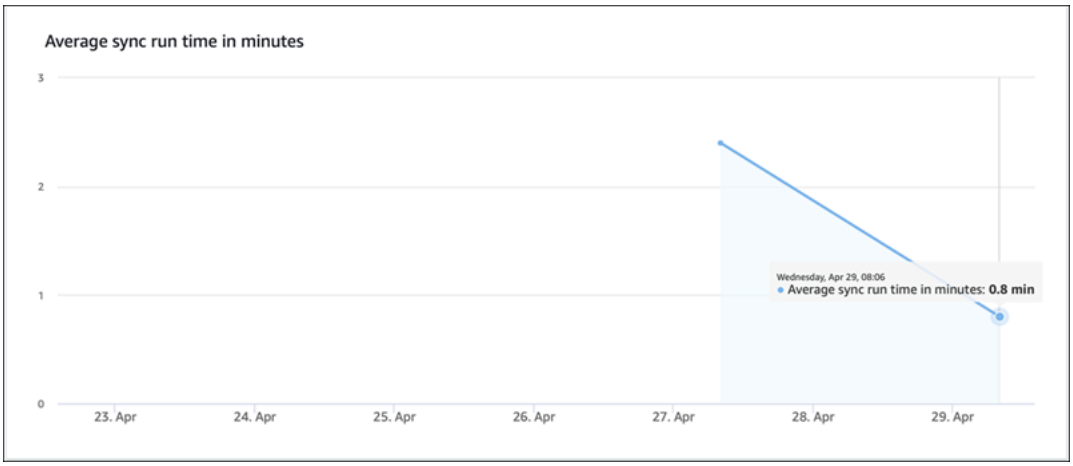


- 文档数量 - 从此数据来源编制索引的文档总数量。这是添加到数据来源的所有文档的总数减去从数据来源中删除的所有文档的总数。使用此信息可确定该数据来源中有多少文档包含在索引中。

- 文档扫描 - 同步运行期间扫描的文档总数量。这包括数据来源中的所有文档，即添加、更新、删除或未更改的所有文档。使用此信息可确定 Amazon Kendra 是否正在扫描数据来源中的所有文档。扫描的文档数量会影响服务收取的费用金额。



- 平均同步运行时间（以分钟为单位） - 完成同步运行所需的平均时间。同步数据来源所需的时间会影响服务收取的费用金额。



使用 AWS CloudTrail 日志记录 Amazon Kendra API 调用

Amazon Kendra 与 AWS CloudTrail 集成，后者是在 Amazon Kendra 中提供用户、角色或 AWS 服务所执行操作的记录的服务。CloudTrail 将对 Amazon Kendra 的所有 API 调用作为事件捕获，包括来自 Amazon Kendra 控制台的调用和对 Amazon Kendra API 的代码调用。如果您创建跟踪，则可以将 CloudTrail 事件持续传输到 Amazon S3 存储桶，包括 Amazon Kendra 的事件。如果您不配置跟踪，则仍可在 CloudTrail 控制台中的 Event history（事件历史记录）中查看最新事件。使用 CloudTrail 收

集的信息，您可以确定向 Amazon Kendra 发出了什么请求、发出请求的 IP 地址、何人发出的请求、请求的发出时间以及其他详细信息。

要了解有关 CloudTrail 的更多信息（包括如何对其进行配置和激活），请参阅《AWS CloudTrail 用户指南》。

CloudTrail 中的 Amazon Kendra 信息

在您创建账户时，将在您的 AWS 账户上激活 CloudTrail。当 Amazon Kendra 中发生活动时，该活动将记录在 CloudTrail 事件中，并与其他 AWS 服务事件一同保存在 CloudTrail 事件历史记录中。您可以在 AWS 账户中查看、搜索和下载最新事件。有关更多信息，请参阅使用 CloudTrail 事件历史记录查看事件。

要持续记录 AWS 账户中的事件（包括 Amazon Kendra 的事件），请创建跟踪。跟踪是一种配置，可让 CloudTrail 将事件作为日志文件传送到指定的 S3 存储桶。默认情况下，在控制台中创建跟踪时，此跟踪应用于所有 AWS 区域。此跟踪在 AWS 分区中记录所有区域中的事件，并将日志文件传送至您指定的 S3 存储桶。此外，您可以配置其他AWS服务，进一步分析在 CloudTrail 日志中收集的事件数据并采取行动。有关更多信息，请参阅：

- 创建跟踪概览
- CloudTrail 支持的服务和集成
- 为 CloudTrail 配置 Amazon SNS 通知
- 从多个区域接收 CloudTrail 日志文件和从多个账户接收 CloudTrail 日志文件

CloudTrail 会记录在 API 参考中记载的所有 Amazon Kendra 操作。例如，对 `CreateIndex`、`CreateDataSource` 和 `Query` 操作的调用将在 CloudTrail 日志文件中生成条目。

每个事件或日志条目都包含有关生成请求的人员信息。有关更多信息，请参阅 CloudTrail userIdentity 元素。

示例：Amazon Kendra 日志文件条目

跟踪是一种配置，可用于将事件作为日志文件传送到指定的 S3 存储桶。CloudTrail 日志文件包含一个或多个日志条目。一个事件表示来自任何源的一个请求，包括有关所请求的操作、操作的日期和时间、请求参数等方面的信息。CloudTrail 日志文件不是公用 API 调用的有序堆栈跟踪，因此它们不会按任何特定顺序显示。

调用 `Query` 操作将创建以下条目。

```
        {
            "eventVersion": "1.05",
            "userIdentity": {
                "type": "AssumedRole | FederatedUser | IAMUser | Root | SAMLUser |
WebIdentityUser",
                "principalId": "principal ID",
                "arn": "ARN",
                "accountId": "account ID",
                "accessKeyId": "access key ID",
                "sessionContext": {
                    "sessionIssuer": {
                        "type": "Role",
                        "principalId": "principal Id",
                        "arn": "ARN",
                        "accountId": "account ID",
                        "userName": "user name"
                    },
                    "webIdFederationData": {

                    },
                    "attributes": {
                        "mfaAuthenticated": false,
                        "creationDate": "timestamp"
                    }
                }
            },
            "eventTime": "timestamp",
            "eventSource": "kendra.amazonaws.com",
            "eventName": "Query",
            "awsRegion": "region",
            "sourceIPAddress": "source IP address",
            "userAgent": "user agent",
            "requestParameters": {
                "indexId": "index ID"
            },
            "responseElements": null,
            "requestID": "request ID",
            "eventID": "event ID",
            "eventType": "AwsApiCall",
            "recipientAccountId": "account ID"
        },
```

使用 AWS CloudTrail 日志记录 Amazon Kendra Intelligent Ranking API 调用

Amazon Kendra Intelligent Ranking 与 AWS CloudTrail 集成，后者是在 Amazon Kendra Intelligent Ranking 中提供用户、角色或 AWS 服务所执行操作的记录的服务。CloudTrail 将来自 Amazon Kendra Intelligent Ranking 的所有 API 调用均作为事件捕获，包括对 Amazon Kendra Intelligent Ranking API 的代码调用。如果您创建跟踪，则可以将 CloudTrail 事件持续传输到 Amazon S3 存储桶，包括 Amazon Kendra Intelligent Ranking 的事件。如果您不配置跟踪，则仍可在 CloudTrail 控制台中的 Event history（事件历史记录）中查看最新事件。使用 CloudTrail 收集的信息，您可以确定向 Amazon Kendra Intelligent Ranking 发出了什么请求、发出请求的 IP 地址、何人发出的请求、请求的发出时间以及其他详细信息。

要了解有关 CloudTrail 的更多信息（包括如何对其进行配置和激活），请参阅《AWS CloudTrail 用户指南》。

CloudTrail 中的 Amazon Kendra Intelligent Ranking 信息

在您创建账户时，将在您的 AWS 账户上激活 CloudTrail。当 Amazon Kendra Intelligent Ranking 中发生活动时，该活动将记录在 CloudTrail 事件中，并与其他 AWS 服务事件一同保存在 CloudTrail 事件历史记录中。您可以在 AWS 账户中查看、搜索和下载最新事件。有关更多信息，请参阅使用 CloudTrail 事件历史记录查看事件。

要持续记录 AWS 账户中的事件（包括 Amazon Kendra Intelligent Ranking 的事件），请创建跟踪。跟踪是一种配置，可让 CloudTrail 将事件作为日志文件传送到指定的 S3 存储桶。默认情况下，在控制台中创建跟踪时，此跟踪应用于所有 AWS 区域。此跟踪在 AWS 分区中记录所有区域中的事件，并将日志文件传送至您指定的 S3 存储桶。此外，您可以配置其他AWS服务，进一步分析在 CloudTrail 日志中收集的事件数据并采取行动。有关更多信息，请参阅：

- 创建跟踪概览
- CloudTrail 支持的服务和集成
- 为 CloudTrail 配置 Amazon SNS 通知
- 从多个区域接收 CloudTrail 日志文件和从多个账户接收 CloudTrail 日志文件

CloudTrail 会记录在 API 参考中记载的所有 Amazon Kendra Intelligent Ranking 操作。例如，对 `CreateRescoreExecutionPlan` 的调用将在 CloudTrail 日志文件中生成条目。

每个事件或日志条目都包含有关生成请求的人员信息。有关更多信息，请参阅 CloudTrail userIdentity 元素。

示例：Amazon Kendra Intelligent Ranking 日志文件条目

跟踪是一种配置，可用于将事件作为日志文件传送到指定的 S3 存储桶。CloudTrail 日志文件包含一个或多个日志条目。一个事件表示来自任何源的一个请求，包括有关所请求的操作、操作的日期和时间、请求参数等方面的信息。CloudTrail 日志文件不是公用 API 调用的有序堆栈跟踪，因此它们不会按任何特定顺序显示。

调用 `CreateRescoreExecutionPlan` 操作将创建以下条目。

```
{
            "eventVersion": "1.08",
            "userIdentity": {
                "type": "AssumedRole",
                "principalId": "principal ID",
                "arn": "ARN",
                "accountId": "account ID",
                "accessKeyId": "access key ID",
                "sessionContext": {
                    "sessionIssuer": {
                        "type": "Role",
                        "principalId": "principal ID",
                        "arn": "ARN",
                        "accountId": "account ID",
                        "userName": "user name"
                    },
                    "webIdFederationData": {},
                    "attributes": {
                        "creationDate": "yyyy-mm-ddThh:mm:ssZ",
                        "mfaAuthenticated": "false"
                    }
                }
            },
            "eventTime": "yyyy-mm-ddThh:mm:ssZ",
            "eventSource": "kendra-ranking.amazonaws.com",
            "eventName": "CreateRescoreExecutionPlan",
            "awsRegion": "region",
            "sourceIPAddress": "source IP address",
            "userAgent": "user agent",
            "requestParameters": {
                "name": "name",
```

```
            "description": "description",
            "clientToken": "client token"
        },
        "responseElements": {
            "id": "rescore execution plan ID",
            "arn": "rescore execution plan ARN"
        },
        "requestID": "request ID",
        "eventID": "event ID",
        "readOnly": false,
        "eventType": "AwsApiCall",
        "managementEvent": true,
        "recipientAccountId": "account ID",
        "eventCategory": "Management",
        "tlsDetails": {
            "tlsVersion": "TLS version",
            "cipherSuite": "cipher suite",
            "clientProvidedHostHeader": "kendra-ranking.[region].api.aws"
        }
    }
```

使用 Amazon CloudWatch 监控 Amazon Kendra

要跟踪索引的运行状况，请使用 Amazon CloudWatch。借助 CloudWatch，您可以获取索引的文档同步指标。您还可以设置 CloudWatch 警报，以便在一个或多个指标超出定义的阈值时收到通知。例如，您可以监控已提交要编制索引的文档数量或无法编制索引的文档数量。

您必须具有适当的 CloudWatch 权限才能使用 CloudWatch 监控 Amazon Kendra。有关更多信息，请参阅《Amazon CloudWatch 用户指南》中的 Amazon CloudWatch 的身份验证和访问控制。

查看 Amazon Kendra 指标

使用 CloudWatch 控制台查看 Amazon Kendra 指标。

要查看指标（CloudWatch 控制台）

1. 登录AWS Management Console并打开 CloudWatch 控制台，网址为 https://console.aws.amazon.com/cloudwatch/。
2. 依次选择指标、所有指标，然后选择 Kendra。
3. 选择维度、指标名称，然后选择 Add to graph (添加到图表)。

4. 选择日期范围的值。所选日期范围的指标计数将显示在该图表中。

创建警报

一个 CloudWatch 警报在一个指定的时间段内监控一个指标，并执行一项或多项操作：向 Amazon Simple Notification Service (Amazon SNS) 主题或自动扩缩策略发送通知。具体执行什么操作取决于在您指定的一系列时间段内指标相对于给定阈值的值。CloudWatch 警报也可以在警报状态发生变化时向您发送 Amazon SNS 消息。

CloudWatch 警报仅当状态发生变化并且已持续了您指定的时间段时才会触发操作。

设置警报

1. 登录AWS Management Console并打开 CloudWatch 控制台，网址为 https://console.aws.amazon.com/cloudwatch/。
2. 选择警报，然后选择创建警报。
3. 选择一个指标。为您的索引和数据来源选择一个 Kendra 指标。您还可以将时间设置为设定的小时数、天数、周数或自定义值。
4. 选择统计数据。例如，平均值。还可以选择警报触发时间段，例如，设定的分钟数、小时数、天数或自定义值。
5. 选择触发警报的阈值，是使用静态值还是范围值，以及达到阈值的条件。
6. 选择触发器的警报状态，指标是否必须超出设定的阈值，或其他状态。选择向谁/哪个电子邮件地址发送警报通知。
7. 如果对警报满意，请选择创建警报。

> **Note**
>
> 您必须为 CloudWatch 警报提供一个名称。

索引同步作业的 CloudWatch 指标

下表介绍了数据来源同步任务的 Amazon Kendra 指标。

如果您使用 API 或 CLI，则在使用 GetMetricStatistics API 时，除了您选择的 `MetricName` 之外，还必须将 `Namespace` 指定为“AWS/kendra”。

指标	描述
DocumentsCrawled	同步作业在运行期间扫描或发现的文档数量。 维度： • IndexId • DataSourceId 单位：计数
DocumentsSubmittedForIndexing	同步作业提交到索引的文档数量。 维度： • IndexId • DataSourceId 单位：计数
DocumentsSubmittedForIndexingFailed	创建索引失败的文档数量。有关详细信息，请查看同步作业的 CloudWatch 日志内容。 维度： • IndexId • DataSourceId 单位：计数
DocumentsSubmittedForDeletion	要求从索引中移除同步作业的文档数量。 维度： • IndexId • DataSourceId

指标	描述
	单位：计数
DocumentsSubmittedForDeletionFailed	删除失败的文档数量。有关详细信息，请查看同步作业的 CloudWatch 日志内容。 维度： • IndexId • DataSourceId 单位：计数

Amazon Kendra 数据来源的指标

下表介绍了数据来源同步任务的 Amazon Kendra 指标。标有星号 (*) 的指标仅适用于 Amazon S3 数据来源。

如果您使用 API 或 CLI，则在使用 GetMetricStatistics API 时，除了您选择的 MetricName 之外，还必须将 Namespace 指定为“AWS/kendra”。

指标	描述
DocumentsSkippedNoChange *	已检查但发现未更改，因而未提交以供创建索引的文档数量。 维度： • IndexId • DataSourceId 单位：计数
DocumentsSkippedInvalidMetadata *	由于关联的元数据文件有问题而跳过的文档数量。有关详细信息，请查看同步运行的 CloudWatch 日志内容。

指标	描述
	维度： • IndexId • DataSourceId 单位：计数
DocumentsCrawled	已检查的文档文件数量。 维度： • IndexId • DataSourceId 单位：计数
DocumentsSubmittedForDeletion	已从数据来源中删除并提交以供删除的已检查文档的数量。 维度： • IndexId • DataSourceId 单位：计数
DocumentsSubmittedForDeletionFailed	从数据来源中删除失败的文档数量。 维度： • IndexId • DataSourceId 单位：计数

指标	描述
DocumentsSubmittedForIndexing	已审查并提交索引的文件数量。 维度： • IndexId • DataSourceId 单位：计数
DocumentsSubmittedForIndexingFailed	已提交以供创建索引但无法创建索引的文档数量。 维度： • IndexId • DataSourceId 单位：计数

已创建索引的文档的指标

下表介绍了已创建索引的文档的 Amazon Kendra 指标。对于使用 BatchPutDocument 操作创建索引的文档，仅支持 IndexId 维度。

如果您使用 API 或 CLI，则在使用 GetMetricStatistics API 时，除了您选择的 MetricName 之外，还必须将 Namespace 指定为“AWS/kendra”。

指标	描述
DocumentsIndexed	已创建索引的文档数量。 维度： • IndexId • DataSourceId

指标	描述
	单位：计数
DocumentsFailedToIndex	无法创建索引的文档数量。有关详细信息，请查看 CloudWatch 日志内容。 维度： • IndexId • DataSourceId 单位：计数
IndexQueryCount	每分钟的索引查询数量。 维度： • IndexId 单位：计数

使用 Amazon CloudWatch Logs 监控 Amazon Kendra

Amazon Kendra 使用 Amazon CloudWatch Logs 来让您深入了解数据来源的运行情况。Amazon Kendra 在创建索引时会记录文档的处理详细信息。为文档创建索引时，它会记录数据来源中发生的错误。您可以使用 CloudWatch Logs 监控、存储和访问日志文件。

CloudWatch Logs 将日志事件存储在日志组中的日志流中。Amazon Kendra 按以下方式使用这些功能：

- 日志组 - Amazon Kendra 将所有日志流存储在每个索引的单个日志组中。在创建索引时，Amazon Kendra 会创建日志组。日志组标识符始终以“aws/kendra/”开头。
- 日志流 - Amazon Kendra 在日志组中为运行的每个索引同步任务创建一个新的数据来源日志流。当一个流达到大约 500 个条目时，它还会创建一个新的文档日志流。
- 日志条目 - 为文档创建索引时，Amazon Kendra 会在日志流中创建日志条目。每个条目都提供有关处理文档或遇到的任何错误的信息。

有关使用 CloudWatch Logs 的更多信息，请参阅《Amazon CloudWatch Logs 用户指南》中的什么是 Amazon CloudWatch Logs。

Amazon Kendra 会创建两种类型的日志流：

- 数据来源日志流
- 文档日志流

数据来源日志流

数据来源日志流发布有关索引同步作业的条目。每个同步作业都会创建一个用于发布条目的新日志流。日志流名称为：

```
data source id/YYYY-MM-DD-HH/data source sync job ID
```

为每个同步作业运行创建一个新的日志流。

发布到数据来源日志流的日志消息有三种类型：

- 无法发送以供创建索引的文档的日志消息。下面是 S3 数据来源中文档的此消息的示例：

```
{
    "DocumentId": "document ID",
    "S3Path": "s3://bucket/prefix/object",
    "Message": "Failed to ingest document via BatchPutDocument.",
    "ErrorCode": "InvalidRequest",
    "ErrorMessage": "No document metadata configuration found for document attribute
 key  city."
}
```

- 无法发送以供删除的文档的日志消息。下面是此消息的示例：

```
{
    "DocumentId": "document ID",
    "Message": "Failed to delete document via BatchDeleteDocument.",
    "ErrorCode": "InvalidRequest",
    "ErrorMessage": "Document can't be deleted because it doesn't exist."
}
```

- 在 Amazon S3 存储桶中发现文档的无效元数据文件时显示的日志消息。下面是此消息的示例。

```
{
    "Message": "Found invalid metadata
 file bucket/prefix/filename.extension.metadata.json."
}
```

- 对于 SharePoint 和数据库连接器，只有在无法为文档创建索引的情况下，Amazon Kendra 才会将消息写入日志流。下面是 Amazon Kendra 记录的错误消息的示例。

```
{
    "DocumentID": "document ID",
    "IndexID": "index ID",
    "SourceURI": "",
    "CrawlStatus": "FAILED",
    "ErrorCode": "403",
    "ErrorMessage": "Access Denied",
    "DataSourceErrorCode": "403"
}
```

文档日志流

Amazon Kendra 在为文档创建索引时记录的有关处理文档的信息。它为存储在 Amazon S3 数据来源中的文档记录一组消息。它仅为存储在 Microsoft SharePoint 或数据库数据来源中的文档记录错误。

如果使用 BatchPutDocument 操作将文档添加到索引中，则日志流的命名如下：

```
YYYY-MM-DD-HH/UUID
```

如果使用数据来源将文档添加到索引中，则日志流的命名如下：

```
dataSourceId/YYYY-MM-DD-HH/UUID
```

每个日志流最多可包含 500 条消息。

如果为文档创建索引失败，则会将以下消息输出到日志流：

```
{
    "DocumentId": "document ID",
    "IndexName": "index name",
    "IndexId": "index ID"
```

```
    "SourceURI": "source URI"
    "IndexingStatus": "DocumentFailedToIndex",
    "ErrorCode": "400 | 500",
    "ErrorMessage": "message"
}
```

Amazon Kendra 中的安全

云安全 AWS 是重中之重。作为 AWS 客户，您可以受益于专为满足大多数安全敏感型组织的要求而构建的数据中心和网络架构。

安全是双方共同承担 AWS 的责任。责任共担模式将其描述为云的安全性和云中的安全性：

- 云安全 —AWS 负责保护在 AWS 云中运行 AWS 服务的基础架构。 AWS 还为您提供可以安全使用的服务。作为AWS 合规计划合规计划合规计划合的一部分，第三方审计师定期测试和验证我们安全的有效性。要了解适用于 Amazon Kendra 的合规计划，请参阅按合规计划提供的范围内的AWS 服务按合规计划划。
- 云端安全-您的责任由您使用的 AWS 服务决定。您还需要对其他因素负责，包括您的数据的敏感性、您公司的要求以及适用的法律法规。

该文档帮助您了解如何在使用 Amazon Kendra 时应用责任共担模式。以下主题说明如何配置 Amazon Kendra 以实现您的安全性和合规性目标。您还将学习如何使用其他 AWS 服务来帮助您监控和保护您的 Amazon Kendra 资源。

主题

Amazon Kendra 中的数据保护

AWS 分担责任模型分担责任模型适用于 Amazon Kendra 中的数据保护。如本模型所述 AWS ，负责保护运行所有内容的全球基础架构 AWS Cloud。您负责维护对托管在此基础架构上的内容的控制。您还负责您所使用的 AWS 服务 的安全配置和管理任务。有关数据隐私的更多信息，请参阅数据隐私

常见问题。有关欧洲数据保护的信息，请参阅 AWS Security Blog 上的 AWS Shared Responsibility Model and GDPR 博客文章。

出于数据保护目的，我们建议您保护 AWS 账户 凭证并使用 AWS IAM Identity Center 或 AWS Identity and Access Management (IAM) 设置个人用户。这样，每个用户只获得履行其工作职责所需的权限。我们还建议您通过以下方式保护数据：

- 对每个账户使用多重身份验证（MFA）。
- 使用 SSL/TLS 与资源通信。 AWS 我们要求使用 TLS 1.2，建议使用 TLS 1.3。
- 使用设置 API 和用户活动日志 AWS CloudTrail。
- 使用 AWS 加密解决方案以及其中的所有默认安全控件 AWS 服务。
- 使用高级托管安全服务（例如 Amazon Macie），它有助于发现和保护存储在 Amazon S3 中的敏感数据。
- 如果您在 AWS 通过命令行界面或 API 进行访问时需要经过 FIPS 140-2 验证的加密模块，请使用 FIPS 端点。有关可用的 FIPS 端点的更多信息，请参阅《美国联邦信息处理标准 (FIPS) 第 140-2 版》。

我们强烈建议您切勿将机密信息或敏感信息（如您客户的电子邮件地址）放入标签或自由格式文本字段（如名称字段）。这包括您 AWS 服务 使用控制台、API 或软件开发工具包与 Amazon Kendra 或其他人合作 AWS CLI的情况。 AWS 在用于名称的标签或自由格式文本字段中输入的任何数据都可能会用于计费或诊断日志。如果您向外部服务器提供网址，强烈建议您不要在网址中包含凭证信息来验证对该服务器的请求。

静态加密

Amazon Kendra 使用所选加密密钥对您的静态数据进行加密。您可以选择以下任一种密钥：

- AWS拥有的 AWS KMS 密钥。如果您未指定加密密钥，则默认使用此密钥对您的数据进行加密。
- 您账户 AWS中由托管的 KMS 密钥。Amazon Kendra 代表您创建、管理和使用密钥。键名称为 `aws/kendra`。
- 客户管理的密钥。您可以提供在账户中创建的加密密钥 ARN。当您使用客户管理的 KMS 密钥时，必须为该密钥提供允许 Amazon Kendra 使用该密钥的密钥策略。选择对称加密客户管理的 KMS 密钥，Amazon Kendra 不支持非对称 KMS 密钥。有关更多信息，请参阅 密钥管理。

传输中加密

Amazon Kendra 使用 HTTPS 协议与客户端应用程序进行通信。它使用 HTTPS 和 AWS 签名代表您的应用程序与其他服务进行通信。如果您使用 VPC，则可以使用在您的 VPC 和 Amazon Kendra 之间建立私有连接。 AWS PrivateLink

密钥管理

Amazon Kendra 使用三种密钥之一对您的索引内容进行加密。您可以选择以下任一种密钥：

- AWS拥有的 AWS KMS。这是默认模式。
- AWS由托管的 KMS 密钥。此密钥在您的账户中创建，由 Amazon Kendra 代表您管理和使用。
- 客户托管的 KMS 密钥。您可以在创建 Amazon Kendra 索引或数据来源时创建密钥，也可以使用 AWS KMS 控制台创建密钥。选择对称加密客户管理的 KMS 密钥。Amazon Kendra 不支持非对称 KMS 密钥。有关更多信息，请参阅《AWS Key Management Service 开发人员指南》中的使用对称和非对称密钥。

亚马逊 Kendra 亚马逊 Kendra 智能排名和接口 VPC 终端节点 ()AWS PrivateLink

您可以通过创建接口 VPC 端点在 VPC 与 Amazon Kendra 之间建立私有连接。接口终端节点由一项技术提供支持 AWS PrivateLink，该技术允许您在没有互联网网关、NAT 设备、VPN 连接或 Direct AWS Connect 连接的情况下私密访问 Amazon Kendra API。VPC 中的实例即使没有公有 IP 地址也可与 Amazon Kendra API 进行通信。您的 VPC 和 Amazon Kendra 之间的流量不会离开亚马逊网络。

每个接口端点均由子网中的一个或多个弹性网络接口表示。

亚马逊 Kendra 和亚马逊 Kendra 智能排名 VPC 终端节点的注意事项

在为 Amazon Kendra 或 Amazon Kendra 智能排名设置接口 VPC 终端节点之前，请务必查看亚马逊 VPC 用户指南中的先决条件。

亚马逊 Kendra 和亚马逊 Kendra 智能排名支持从你的 VPC 调用其所有 API 操作。

为亚马逊 Kendra 和亚马逊 Kendra 智能排名创建接口 VPC 终端节点

您可以使用亚马逊 VPC 控制台或 () 为 Amazon Kendra 或 Amazon Kendra 智能排名服务创建 VPC 终端节点。 AWS Command Line Interface AWS CLI

使用以下服务名称为 Amazon Kendra 创建 VPC 端点：

- com.amazonaws.*region*.kendra

使用以下服务名称为 Amazon Kendra 智能排名创建 VPC 终端节点：

- aws.api.*region*.kendra-ranking

创建 VPC 终端节点后，您可以使用以下示例 AWS CLI 命令，该命令使用`endpoint-url`参数指定 Amazon Kendra API 的接口终端节点：

```
aws kendra list-indices --endpoint-url https://VPC endpoint
```

*VPC ####*是创建接口终端节点时生成的 DNS 名称。此名称包括 VPC 终端节点 ID 和 Amazon Kendra 服务名称（包括该区域）。例如，`vpce-1234-abcdef.kendra.us-west-2.vpce.amazonaws.com`。

如果您为终端节点激活私有 DNS，则可以使用该区域的默认 DNS 名称向 Amazon Kendra 发出 API 请求。例如，`kendra.us-east-1.amazonaws.com`。

有关更多信息，请参阅《Amazon VPC 用户指南》中的创建接口端点。

为亚马逊 Kendra 和亚马逊 Kendra 智能排名创建 VPC 终端节点策略

您可以将终端节点策略附加到您的 VPC 终端节点，以控制对 Amazon Kendra 或 Amazon Kendra 智能排名的访问权限。

亚马逊 Kendra 或亚马逊 Kendra 智能排名政策规定了以下信息：

- 可以执行操作的委托人/授权用户。
- 可执行的操作。
- 可对其执行操作的资源。

示例：Amazon Kendra 操作的 VPC 端点策略

下面是用于 Amazon Kendra 的端点策略示例。当连接到终端节点时，该策略允许所有委托人/授权用户访问所有资源上所有可用的 Amazon Kendra 操作。

```
{
```

```
   "Statement":[
      {
         "Principal":"*",
         "Effect":"Allow",
         "Action":[
            "kendra:*"
         ],
         "Resource":"*"
      }
   ]
}
```

示例：亚马逊 Kendra 智能排名操作的 VPC 终端节点策略

以下是 Amazon Kendra 智能排名的终端节点策略示例。当连接到终端节点时，该策略允许所有委托人/授权用户访问所有资源上所有可用的 Amazon Kendra 智能排名操作。

```
{
   "Statement":[
      {
         "Principal":"*",
         "Effect":"Allow",
         "Action":[
            "kendra-ranking:*"
         ],
         "Resource":"*"
      }
   ]
}
```

有关更多信息，请参阅 Amazon VPC 用户指南中的使用终端节点策略控制 VPC 终端节点的访问权限。

Amazon Kendra 的身份和访问管理

AWS Identity and Access Management (IAM) AWS 服务 可帮助管理员安全地控制对 AWS 资源的访问权限。IAM 管理员控制谁可以通过身份验证（登录）和获得授权（具有权限）来使用 Amazon Kendra 资源。您可以使用 IAM AWS 服务 ，无需支付额外费用。

主题

- 受众

- 使用身份进行身份验证
- 使用策略管理访问
- Amazon Kendra 如何与 IAM 协同工作
- Amazon Kendra 基于身份的策略示例
- AWS 亚马逊 Kendra 的托管政策
- Amazon Kendra 身份和访问问题排查

受众

您的使用方式 AWS Identity and Access Management (IAM) 会有所不同，具体取决于您在 Amazon Kendra 中所做的工作。

服务用户 - 如果您使用 Amazon Kendra 服务来完成任务，管理员会为您提供所需的凭证和权限。随着您使用更多 Amazon Kendra 特征来完成任务，您可能需要额外权限。了解如何管理访问权限有助于您向管理员请求适合的权限。如果您无法访问 Amazon Kendra 中的功能，请参阅 Amazon Kendra 身份和访问问题排查。

服务管理员 – 如果您在公司负责管理 Amazon Kendra 资源，您可能对 Amazon Kendra 具有完全访问权限。您有责任确定您的服务用户应访问哪些 Amazon Kendra 功能和资源。然后，您必须向 IAM 管理员提交请求以更改服务用户的权限。请查看该页面上的信息以了解 IAM 的基本概念。要了解有关您的公司如何将 IAM 与 Amazon Kendra 搭配使用的更多信息，请参阅Amazon Kendra 如何与 IAM 协同工作。

IAM 管理员 - 如果您是 IAM 管理员，您可能希望了解有关如何编写策略以管理对 Amazon Kendra 的访问权限的详细信息。要查看您可在 IAM 中使用的 Amazon Kendra 基于身份的策略示例，请参阅 Amazon Kendra 基于身份的策略示例。

使用身份进行身份验证

身份验证是您 AWS 使用身份凭证登录的方式。您必须以 IAM 用户身份或通过担 AWS 账户根用户任 IAM 角色进行身份验证（登录 AWS）。

您可以使用通过身份源提供的凭据以 AWS 联合身份登录。 AWS IAM Identity Center （IAM Identity Center）用户、贵公司的单点登录身份验证以及您的 Google 或 Facebook 凭据就是联合身份的示例。当您以联合身份登录时，您的管理员以前使用 IAM 角色设置了身份联合验证。当你使用联合访问 AWS 时，你就是在间接扮演一个角色。

根据您的用户类型，您可以登录 AWS Management Console 或 AWS 访问门户。有关登录的更多信息 AWS，请参阅《AWS 登录 用户指南》中的如何登录到您 AWS 账户的。

如果您 AWS 以编程方式访问，则会 AWS 提供软件开发套件 (SDK) 和命令行接口 (CLI)，以便使用您的凭据对请求进行加密签名。如果您不使用 AWS 工具，则必须自己签署请求。有关使用推荐的方法自行签署请求的更多信息，请参阅 IAM 用户指南中的签署 AWS API 请求。

无论使用何种身份验证方法，您可能需要提供其他安全信息。例如， AWS 建议您使用多重身份验证 (MFA) 来提高账户的安全性。要了解更多信息，请参阅《AWS IAM Identity Center 用户指南》中的 多重身份验证和《IAM 用户指南》中的 在 AWS中使用多重身份验证（MFA）。

AWS 账户 root 用户

创建时 AWS 账户，首先要有一个登录身份，该身份可以完全访问账户中的所有资源 AWS 服务 和资源。此身份被称为 AWS 账户 root 用户，使用您创建帐户时使用的电子邮件地址和密码登录即可访问该身份。强烈建议您不要使用根用户执行日常任务。保护好根用户凭证，并使用这些凭证来执行仅根用户可以执行的任务。有关需要您以根用户身份登录的任务的完整列表，请参阅《IAM 用户指南》中的需要根用户凭证的任务。

IAM 用户和组

I AM 用户是您 AWS 账户 内部对个人或应用程序具有特定权限的身份。在可能的情况下，我们建议使用临时凭证，而不是创建具有长期凭证（如密码和访问密钥）的 IAM 用户。但是，如果您有一些特定的使用场景需要长期凭证以及 IAM 用户，建议您轮换访问密钥。有关更多信息，请参阅《IAM 用户指南》中的 对于需要长期凭证的使用场景定期轮换访问密钥。

IAM 组是一个指定一组 IAM 用户的身份。您不能使用组的身份登录。您可以使用组来一次性为多个用户指定权限。如果有大量用户，使用组可以更轻松地管理用户权限。例如，您可能具有一个名为 IAMAdmins 的组，并为该组授予权限以管理 IAM 资源。

用户与角色不同。用户唯一地与某个人员或应用程序关联，而角色旨在让需要它的任何人代入。用户具有永久的长期凭证，而角色提供临时凭证。要了解更多信息，请参阅《IAM 用户指南》中的何时创建 IAM 用户（而不是角色）。

IAM 角色

I AM 角色是您内部具有特定权限 AWS 账户 的身份。它类似于 IAM 用户，但与特定人员不关联。您可以 AWS Management Console 通过切换角色在中临时担任 IAM 角色。您可以通过调用 AWS CLI 或 AWS API 操作或使用自定义 URL 来代入角色。有关使用角色的方法的更多信息，请参阅《IAM 用户指南》中的使用 IAM 角色。

具有临时凭证的 IAM 角色在以下情况下很有用：

- 联合用户访问 – 要向联合身份分配权限，请创建角色并为角色定义权限。当联合身份进行身份验证时，该身份将与角色相关联并被授予由此角色定义的权限。有关联合身份验证的角色的信息，请参阅《IAM 用户指南》中的为第三方身份提供商创建角色。如果您使用 IAM Identity Center，则需要配置权限集。为控制您的身份在进行身份验证后可以访问的内容，IAM Identity Center 将权限集与 IAM 中的角色相关联。有关权限集的信息，请参阅《AWS IAM Identity Center 用户指南》中的 权限集。
- 临时 IAM 用户权限 – IAM 用户可代入 IAM 用户或角色，以暂时获得针对特定任务的不同权限。
- 跨账户存取 – 您可以使用 IAM 角色以允许不同账户中的某个人（可信主体）访问您的账户中的资源。角色是授予跨账户访问权限的主要方式。但是，对于某些资源 AWS 服务，您可以将策略直接附加到资源（而不是使用角色作为代理）。要了解角色和基于资源的跨账户访问策略之间的区别，请参阅 IA M 用户指南中的跨账户资源访问。
- 跨服务访问 — 有些 AWS 服务 使用其他 AWS 服务服务中的功能。例如，当您在某个服务中进行调用时，该服务通常会在 Amazon EC2 中运行应用程序或在 Simple Storage Service (Amazon S3) 中存储对象。服务可能会使用发出调用的主体的权限、使用服务角色或使用服务相关角色来执行此操作。
  - 转发访问会话 (FAS) — 当您使用 IAM 用户或角色在中执行操作时 AWS，您被视为委托人。使用某些服务时，您可能会执行一个操作，然后此操作在其他服务中启动另一个操作。FAS 使用调用委托人的权限以及 AWS 服务 向下游服务发出请求的请求。 AWS 服务只有当服务收到需要与其他 AWS 服务 或资源交互才能完成的请求时，才会发出 FAS 请求。在这种情况下，您必须具有执行这两个操作的权限。有关发出 FAS 请求时的策略详情，请参阅转发访问会话。
  - 服务角色 - 服务角色是服务代表您在您的账户中执行操作而分派的 IAM 角色。IAM 管理员可以在 IAM 中创建、修改和删除服务角色。有关更多信息，请参阅《IAM 用户指南》中的创建向 AWS 服务委派权限的角色。
  - 服务相关角色-服务相关角色是一种与服务相关联的服务角色。 AWS 服务服务可以代入代表您执行操作的角色。服务相关角色出现在您的中 AWS 账户 ，并且归服务所有。IAM 管理员可以查看但不能编辑服务相关角色的权限。
- 在 A@@ mazon EC2 上运行的应用程序 — 您可以使用 IAM 角色管理在 EC2 实例上运行并发出 AWS CLI 或 AWS API 请求的应用程序的临时证书。这优先于在 EC2 实例中存储访问密钥。要向 EC2 实例分配 AWS 角色并使其可供其所有应用程序使用，您需要创建附加到该实例的实例配置文件。实例配置文件包含角色，并使 EC2 实例上运行的程序能够获得临时凭证。有关更多信息，请参阅《IAM 用户指南》中的使用 IAM 角色为 Amazon EC2 实例上运行的应用程序授予权限。

要了解是使用 IAM 角色还是 IAM 用户，请参阅IAM 用户指南中的何时创建 IAM 角色（而不是用户）。

使用策略管理访问

您可以 AWS 通过创建策略并将其附加到 AWS 身份或资源来控制中的访问权限。策略是其中的一个对象 AWS ，当与身份或资源关联时，它会定义其权限。 AWS 在委托人（用户、root 用户或角色会话）发出请求时评估这些策略。策略中的权限确定是允许还是拒绝请求。大多数策略都以 JSON 文档的 AWS 形式存储在中。有关 JSON 策略文档的结构和内容的更多信息，请参阅 IAM 用户指南中的JSON 策略概览。

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体可以对什么资源执行操作，以及在什么条件下执行。

默认情况下，用户和角色没有权限。要授予用户对所需资源执行操作的权限，IAM 管理员可以创建 IAM 策略。管理员随后可以向角色添加 IAM policy，用户可以代入角色。

IAM 策略定义操作的权限，无关乎您使用哪种方法执行操作。例如，假设您有一个允许 `iam:GetRole` 操作的策略。拥有该策略的用户可以从 AWS Management Console AWS CLI、或 AWS API 获取角色信息。

基于身份的策略

基于身份的策略是可附加到身份（如 IAM 用户、用户组或角色）的 JSON 权限策略文档。这些策略控制用户和角色可在何种条件下对哪些资源执行哪些操作。要了解如何创建基于身份的策略，请参阅《IAM 用户指南》中的创建 IAM policy。

基于身份的策略可以进一步归类为内联策略或托管策略。内联策略直接嵌入单个用户、组或角色中。托管策略是独立的策略，您可以将其附加到中的多个用户、群组和角色 AWS 账户。托管策略包括 AWS 托管策略和客户托管策略。要了解如何在托管式策略和内联策略之间进行选择，请参阅 IAM 用户指南中的在托管式策略与内联策略之间进行选择。

基于资源的策略

基于资源的策略是附加到资源的 JSON 策略文档。基于资源的策略的示例包括 IAM 角色信任策略和 Amazon S3 存储桶策略。在支持基于资源的策略的服务中，服务管理员可以使用它们来控制对特定资源的访问。对于在其中附加策略的资源，策略定义指定主体可以对该资源执行哪些操作以及在什么条件下执行。您必须在基于资源的策略中指定主体。委托人可以包括账户、用户、角色、联合用户或 AWS 服务。

基于资源的策略是位于该服务中的内联策略。您不能在基于资源的策略中使用 IAM 中的 AWS 托管策略。

访问控制列表 (ACL)

访问控制列表（ACL）控制哪些主体（账户成员、用户或角色）有权访问资源。ACL 与基于资源的策略类似，尽管它们不使用 JSON 策略文档格式。

Amazon S3 和 Amazon VPC 就是支持 ACL 的服务示例。 AWS WAF要了解有关 ACL 的更多信息，请参阅《Amazon Simple Storage Service 开发人员指南》中的访问控制列表（ACL）概览。

其他策略类型

AWS 支持其他不太常见的策略类型。这些策略类型可以设置更常用的策略类型向您授予的最大权限。

- 权限边界：权限边界是一个高级特征，用于设置基于身份的策略可以为 IAM 实体（IAM 用户或角色）授予的最大权限。您可为实体设置权限边界。这些结果权限是实体基于身份的策略及其权限边界的交集。在 `Principal` 中指定用户或角色的基于资源的策略不受权限边界限制。任一项策略中的显式拒绝将覆盖允许。有关权限边界的更多信息，请参阅《IAM 用户指南》中的 IAM 实体的权限边界。
- 服务控制策略 (SCP)-SCP 是 JSON 策略，用于指定组织或组织单位 (OU) 的最大权限。 AWS Organizations AWS Organizations 是一项用于对您的企业拥有的多 AWS 账户 项进行分组和集中管理的服务。如果在组织内启用了所有功能，则可对任意或全部账户应用服务控制策略 (SCP)。SCP 限制成员账户中的实体（包括每个 AWS 账户根用户实体）的权限。有关 Organizations 和 SCP 的更多信息，请参阅《AWS Organizations 用户指南》中的 SCP 的工作原理。
- 会话策略 – 会话策略是当您以编程方式为角色或联合用户创建临时会话时作为参数传递的高级策略。结果会话的权限是用户或角色的基于身份的策略和会话策略的交集。权限也可以来自基于资源的策略。任一项策略中的显式拒绝将覆盖允许。有关更多信息，请参阅《IAM 用户指南》中的会话策略。

多个策略类型

当多个类型的策略应用于一个请求时，生成的权限更加复杂和难以理解。要了解在涉及多种策略类型时如何 AWS 确定是否允许请求，请参阅 IAM 用户指南中的策略评估逻辑。

Amazon Kendra 如何与 IAM 协同工作

在使用 IAM 管理对 Amazon Kendra 的访问权限之前，您应该了解哪些 IAM 功能可用于 Amazon Kendra。要全面了解 Amazon Kendra 和其他 AWS 服务如何与 IAM 配合使用，请参阅 IAM 用户指南中的与 IAM 配合使用的AWS 服务。

主题

- Amazon Kendra 基于身份的策略
- Amazon Kendra 基于资源的策略
- 访问控制列表 (ACL)
- 基于 Amazon Kendra 标签的授权
- Amazon Kendra IAM 角色

Amazon Kendra 基于身份的策略

通过使用 IAM 基于身份的策略，您可以指定允许或拒绝的操作和资源以及允许或拒绝操作的条件。Amazon Kendra 支持特定的操作、资源和条件键。要了解在 JSON 策略中使用的所有元素，请参阅《IAM 用户指南》 中的 IAM JSON 策略元素参考。

操作

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体可以对什么资源执行操作，以及在什么条件下执行。

JSON 策略的 `Action` 元素描述可用于在策略中允许或拒绝访问的操作。策略操作通常与关联的 AWS API 操作同名。有一些例外情况，例如没有匹配 API 操作的仅限权限 操作。还有一些操作需要在策略中执行多个操作。这些附加操作称为相关操作。

在策略中包含操作以授予执行关联操作的权限。

Amazon Kendra 中的策略操作在操作前面使用以下前缀：`kendra:`。例如，要授予某人ListIndices通过 API 操作列出 Amazon Kendra 索引的权限，您需要将该`kendra:ListIndices`操作包含在他们的策略中。策略语句必须包含 `Action` 或 `NotAction` 元素。Amazon Kendra 定义了一组自己的操作，以描述您可以使用该服务执行的任务。

要在单个语句中指定多项操作，请使用逗号将它们隔开，如下所示：

```
"Action": [
      "kendra:action1",
      "kendra:action2"
```

您也可以使用通配符（*) 指定多个操作。例如，要指定以单词 `Describe` 开头的所有操作，包括以下操作：

```
"Action": "kendra:Describe*"
```

要查看 Amazon Kendra 操作的列表，请参阅《IAM 用户指南》中的 Amazon Kendra 定义的操作。

资源

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体 可以对什么资源执行操作，以及在什么条件下执行。

Resource JSON 策略元素指定要向其应用操作的一个或多个对象。语句必须包含 Resource 或 NotResource 元素。作为最佳实践，请使用其 Amazon 资源名称 (ARN) 指定资源。对于支持特定资源类型（称为资源级权限）的操作，您可以执行此操作。

对于不支持资源级权限的操作（如列出操作），请使用通配符 (*) 指示语句应用于所有资源。

```
"Resource": "*"
```

Amazon Kendra 索引资源具有以下 ARN：

```
arn:${Partition}:kendra:${Region}:${Account}:index/${IndexId}
```

有关 ARN 格式的更多信息，请参阅 A mazon 资源名称 (ARN) 和 AWS 服务命名空间。

例如，要在语句中指定索引，请使用以下 ARN 中索引的 GUID：

```
"Resource": "arn:aws:kendra:${Region}:${Account}:index/${GUID}"
```

要指定属于特定账户的所有索引，请使用通配符（*）：

```
"Resource": "arn:aws:${Region}:${Account}:index/*"
```

无法对特定资源执行某些 Amazon Kendra 操作，例如，用于创建资源的操作。在这些情况下，您必须使用通配符 （*)。

```
"Resource": "*"
```

要查看 Amazon Kendra 资源类型及其 ARN 的列表，请参阅《IAM 用户指南》中的 Amazon Kendra 定义的资源。要了解您可以使用哪些操作指定每个资源的 ARN，请参阅 Amazon Kendra 定义的操作。

条件键

管理员可以使用 AWS JSON 策略来指定谁有权访问什么。也就是说，哪个主体可以对什么资源执行操作，以及在什么条件下执行。

在 Condition 元素（或 Condition 块）中，可以指定语句生效的条件。Condition 元素是可选的。您可以创建使用条件运算符（例如，等于或小于）的条件表达式，以使策略中的条件与请求中的值相匹配。

如果您在一个语句中指定多个 Condition 元素，或在单个 Condition 元素中指定多个键，则 AWS 使用逻辑 AND 运算评估它们。如果您为单个条件键指定多个值，则使用逻辑OR运算来 AWS 评估条件。在授予语句的权限之前必须满足所有的条件。

在指定条件时，您也可以使用占位符变量。例如，只有在使用 IAM 用户名标记 IAM 用户时，您才能为其授予访问资源的权限。有关更多信息，请参阅《IAM 用户指南》中的 IAM policy 元素：变量和标签。

AWS 支持全局条件密钥和特定于服务的条件密钥。要查看所有 AWS 全局条件键，请参阅 IAM 用户指南中的AWS 全局条件上下文密钥。

Amazon Kendra 不提供任何特定于服务的条件键，但支持使用某些全局条件键。要查看所有 AWS 全局条件键，请参阅 IAM 用户指南中的AWS 全局条件上下文密钥。

示例

要查看 Amazon Kendra 基于身份的策略的示例，请参阅Amazon Kendra 基于身份的策略示例。

Amazon Kendra 基于资源的策略

Amazon Kendra 不支持基于资源的策略。

访问控制列表 (ACL)

Amazon Kendra 不支持访问 AWS 服务和资源的访问控制列表（ACL）。

基于 Amazon Kendra 标签的授权

您可以将标签与某些类型的 Amazon Kendra 资源关联以授权访问这些资源。要基于标签控制访问，请使用 `aws:RequestTag/key-name` 或 `aws:TagKeys` 条件键在策略的条件元素中提供标签信息。

下表列出了基于标签的访问控制的操作、相应的资源类型和条件键。根据与相应资源类型关联的标签对每个操作进行授权。

操作	资源类型	条件键
CreateData来源		aws:RequestTag ，aws:TagKeys
CreateFaq		aws:RequestTag ，aws:TagKeys
CreateIndex		aws:RequestTag ，aws:TagKeys
API_ ListTags ForResource	数据来源、常见问题、索引	
TagResource	数据来源、常见问题、索引	aws:RequestTag ，aws:TagKeys
UntagResource	数据来源、常见问题、索引	aws:TagKeys

有关标记 Amazon Kendra 资源的信息，请参阅标签。有关基于资源标签限制对资源的访问的基于身份的策略示例，请参阅基于标签的策略示例。有关使用标签限制对资源的访问的信息，请参阅《IAM 用户指南》中的使用标签控制访问。

Amazon Kendra IAM 角色

I AM 角色是您的 AWS 账户中具有特定权限的实体。

将临时凭证用于 Amazon Kendra

可以使用临时凭证进行联合身份验证登录，分派 IAM 角色或分派跨账户角色。您可以通过调用 AWS STS API 操作（例如AssumeRole或GetFederation令牌）来获取临时安全证书。

Amazon Kendra 支持使用临时凭证。

服务角色

此功能允许服务代表您担任服务角色。此角色允许服务访问其他服务中的资源以代表您完成操作。服务角色显示在 IAM 账户中，并归该账户所有。这意味着，IAM 管理员可以更改该角色的权限。但是，这样做可能会中断服务的功能。

Amazon Kendra 支持服务角色。

在 Amazon Kendra 中选择 IAM 角色

在创建索引、调用 `BatchPutDocument` 操作、创建数据来源或创建常见问题解答时，必须提供访问角色 Amazon 资源名称（ARN），Amazon Kendra 使用该角色代表您访问所需资源。如果您之前创建了一个角色，Amazon Kendra 控制台会为您提供一个角色列表供您选择。选择允许访问所需资源的角色非常重要。有关更多信息，请参阅 IAM 的访问角色 Amazon Kendra。

Amazon Kendra 基于身份的策略示例

原定设置情况下，用户和角色没有创建或修改 Amazon Kendra 资源的权限。他们也无法使用 AWS Management Console AWS CLI、或 AWS API 执行任务。IAM 管理员必须创建 IAM 策略，以便为用户和角色授予权限以对所需的指定资源执行特定的 API 操作。然后，管理员必须将这些策略附加到需要这些权限的用户或组。

要了解如何使用这些示例 JSON 策略文档创建 IAM 基于身份的策略，请参阅《IAM 用户指南》中的在 JSON 选项卡上创建策略。

主题

策略最佳实践

基于身份的策略确定某个人是否可以创建、访问或删除您账户中的 Amazon Kendra 资源。这些操作可能会使 AWS 账户产生成本。创建或编辑基于身份的策略时，请遵循以下指南和建议：

- 开始使用 AWS 托管策略并转向最低权限权限 — 要开始向用户和工作负载授予权限，请使用为许多常见用例授予权限的AWS 托管策略。它们在你的版本中可用 AWS 账户。我们建议您通过定义针对您的用例的 AWS 客户托管策略来进一步减少权限。有关更多信息，请参阅《IAM 用户指南》中的 AWS 托管策略 或 工作职能的AWS 托管策略。
- 应用最低权限 – 在使用 IAM policy 设置权限时，请仅授予执行任务所需的权限。为此，您可以定义在特定条件下可以对特定资源执行的操作，也称为最低权限许可。有关使用 IAM 应用权限的更多信息，请参阅《IAM 用户指南》中的 IAM 中的策略和权限。

- 使用 IAM policy 中的条件进一步限制访问权限 – 您可以向策略添加条件来限制对操作和资源的访问。例如，您可以编写策略条件来指定必须使用 SSL 发送所有请求。如果服务操作是通过特定的方式使用的，则也可以使用条件来授予对服务操作的访问权限 AWS 服务，例如 AWS CloudFormation。有关更多信息，请参阅《IAM 用户指南》中的 IAM JSON 策略元素：条件 。
- 使用 IAM Access Analyzer 验证您的 IAM policy，以确保权限的安全性和功能性 – IAM Access Analyzer 会验证新策略和现有策略，以确保策略符合 IAM policy语言 (JSON) 和 IAM 最佳实践。IAM Access Analyzer 提供 100 多项策略检查和可操作的建议，以帮助您制定安全且功能性强的策略。有关更多信息，请参阅《IAM 用户指南》中的 IAM Acess Analyzer 策略验证。
- 需要多重身份验证 (MFA)-如果 AWS 账户您的场景需要 IAM 用户或根用户，请启用 MFA 以提高安全性。若要在调用 API 操作时需要 MFA，请将 MFA 条件添加到您的策略中。有关更多信息，请参阅《IAM 用户指南》中的 配置受 MFA 保护的 API 访问。

有关 IAM 中的最佳实操的更多信息，请参阅《IAM 用户指南》中的 IAM 中的安全最佳实操。

适用于 Amazon Kendra 的AWS 托管（预定义）策略

AWS 通过提供由创建和管理的独立 IAM 策略来解决许多常见用例 AWS。这些策略称为 AWS 托管策略。 AWS 托管策略使您可以更轻松地向用户、组和角色分配权限，而不必自己编写策略。有关更多信息，请参阅《IAM 用户指南》中的为用户添加权限。

以下 AWS 托管策略仅适用于 Amazon Kendra，您可以将其附加到账户中的群组和角色：

- AmazonKendraReadOnly— 授予对 Amazon Kendra 资源的只读访问权限。
- AmazonKendraFullAccess— 授予创建、读取、更新、删除、标记和运行所有 Amazon Kendra 资源的完全访问权限。

对于控制台，您的角色还必须具有 `iam:CreateRole`、`iam:CreatePolicy`、`iam:AttachRolePolicy` 和 `s3:ListBucket` 权限。

> **Note**
>
> 通过登录 IAM 控制台并搜索特定策略，可以查看这些权限。

您还可以创建自己的自定义策略，以授予执行 Amazon Kendra API 操作的相关权限。您可以将这些自定义策略附加到需要这些权限的 IAM 角色或组。有关 Amazon Kendra 的 IAM 策略的示例，请参阅Amazon Kendra 基于身份的策略示例。

允许用户查看他们自己的权限

该示例说明了您如何创建策略，以允许 IAM 用户查看附加到其用户身份的内联和托管策略。此策略包括在控制台上或使用 AWS CLI 或 AWS API 以编程方式完成此操作的权限。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ViewOwnUserInfo",
            "Effect": "Allow",
            "Action": [
                "iam:GetUserPolicy",
                "iam:ListGroupsForUser",
                "iam:ListAttachedUserPolicies",
                "iam:ListUserPolicies",
                "iam:GetUser"
            ],
            "Resource": ["arn:aws:iam::*:user/${aws:username}"]
        },
        {
            "Sid": "NavigateInConsole",
            "Effect": "Allow",
            "Action": [
                "iam:GetGroupPolicy",
                "iam:GetPolicyVersion",
                "iam:GetPolicy",
                "iam:ListAttachedGroupPolicies",
                "iam:ListGroupPolicies",
                "iam:ListPolicyVersions",
                "iam:ListPolicies",
                "iam:ListUsers"
            ],
            "Resource": "*"
        }
    ]
}
```

访问一个 Amazon Kendra 索引

在此示例中，您想向 AWS 账户中的用户授予查询索引的权限。

```
{
```

```
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "QueryIndex",
            "Effect": "Allow",
            "Action": [
                "kendra:Query"
            ],
            "Resource": "arn:aws:kendra:${Region}:${Account}:index/${Index ID}"
        }
    ]
}
```

基于标签的策略示例

基于标签的策略是 JSON 策略文档，其中指定主体可对所标记的资源执行哪些操作。

示例：使用标签访问资源

此示例策略授予您 AWS 账户中的用户或角色对标有密钥**department**和值的任何资源使用Query操作的权限**finance**。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "kendra:Query"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "aws:ResourceTag/department": "finance"
                }
            }
        }
   ]
}
```

示例：使用标签激活 Amazon Kendra 操作

此示例策略授予您 AWS 账户中的用户或角色使用任何 Amazon Kendra 操作的权限，但TagResource操作任何标有密钥**department**和值的资源除外。**finance**

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "kendra:*",
            "Resource": "*"
        },
        {
            "Effect": "Deny",
            "Action": [
                "kendra:TagResource"
            ],
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "aws:ResourceTag/department": "finance"
                }
            }
        }
    ]
}
```

示例：使用标签限制对操作的访问

此示例策略限制您 AWS 账户中的用户或角色使用该CreateIndex操作的访问权限，除非该用户提供了**department**标签并且标签具有允许的值**finance**和**IT**。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "kendra:CreateIndex",
            "Resource": "*"
        },
        {
            "Effect": "Deny",
```

```
            "Action": "kendra:CreateIndex",
            "Resource": "*",
            "Condition": {
                "Null": {
                    "aws:RequestTag/department": "true"
                }
            }
        },
        {
            "Effect": "Deny",
            "Action": "kendra:CreateIndex",
            "Resource": "*",
            "Condition": {
                "ForAnyValue:StringNotEquals": {
                    "aws:RequestTag/department": [
                        "finance",
                        "IT"
                    ]
                }
            }
        }
    ]
}
```

AWS 亚马逊 Kendra 的托管政策

要向用户、群组和角色添加权限，使用 AWS 托管策略比自己编写策略要容易得多。创建仅为团队提供所需权限的 IAM 客户管理型策略需要时间和专业知识。要快速入门，您可以使用我们的 AWS 托管策略。这些政策涵盖常见用例，可在您的 AWS 账户中使用。有关 AWS 托管策略的更多信息，请参阅 IAM 用户指南中的AWS 托管策略。

AWS 服务维护和更新 AWS 托管策略。您无法更改 AWS 托管策略中的权限。服务偶尔会向 AWS 管理型策略添加额外权限以支持新特征。此类更新会影响附加策略的所有身份（用户、组和角色）。当启动新特征或新操作可用时，服务最有可能会更新 AWS 管理型策略。服务不会从 AWS 托管策略中移除权限，因此策略更新不会破坏您的现有权限。

此外，还 AWS 支持跨多个服务的工作职能的托管策略。例如，ReadOnly访问 AWS 管理策略提供对所有 AWS 服务和资源的只读访问权限。当服务启动一项新功能时， AWS 会为新操作和资源添加只读权限。有关工作职能策略的列表和说明，请参阅 IAM 用户指南中的适用于工作职能的AWS 管理型策略。

AWS 托管策略： AmazonKendraReadOnly

授予 Amazon Kendra 资源只读访问权限。该策略包含以下权限。

- kendra - 允许用户执行返回项目列表或项目详细信息的操作。这包括以 Describe、List、Query、BatchGetDocumentStatus、GetQuerySuggestions 或 GetSnapshots 开头的 API 操作。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Action": [
                "kendra:Describe*",
                "kendra:List*",
                "kendra:Query",
                "kendra:BatchGetDocumentStatus",
                "kendra:GetQuerySuggestions",
                "kendra:GetSnapshots"
            ],
            "Effect": "Allow",
            "Resource": "*"
        }
    ]
}
```

AWS 托管策略： AmazonKendraFullAccess

授予创建、读取、更新、删除、标记和运行全部 Amazon Kendra 资源的完全访问权限。该策略包含以下权限。

- kendra - 允许主体读写访问 Amazon Kendra 中的所有操作。
- s3 - 允许主体获取 Amazon S3 存储桶的位置并列出存储桶。
- iam - 允许主体传递和列出角色。
- kms— 允许委托人描述和列出 AWS KMS 密钥和别名。
- secretsmanager - 允许主体创建、描述和列出密钥。

- ec2 - 允许主体描述安全组、VCP（虚拟私有云）和子网。
- cloudwatch - 允许主体查看 Cloud Watch 指标。

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "iam:PassRole",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "iam:PassedToService": "kendra.amazonaws.com"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:ListRoles"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeVpcs",
                "ec2:DescribeSubnets"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:ListKeys",
                "kms:ListAliases",
                "kms:DescribeKey"
            ],
            "Resource": "*"
        },
        {
```

```
            "Effect": "Allow",
            "Action": [
                "s3:ListAllMyBuckets",
                "s3:GetBucketLocation"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "secretsmanager:ListSecrets"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "cloudwatch:GetMetricData"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "secretsmanager:CreateSecret",
                "secretsmanager:DescribeSecret"
            ],
            "Resource": "arn:aws:secretsmanager:*:*:secret:AmazonKendra-*"
        },
        {
            "Effect": "Allow",
            "Action": "kendra:*",
            "Resource": "*"
        }
    ]
}
```

亚马逊 Kendra 更新了托管政策 AWS

查看有关自 Amazon Kendra AWS 托管政策开始跟踪这些变更以来该服务更新的详细信息。有关此页面更改的自动提示，请订阅 Amazon Kendra 文档历史记录页面上的 RSS 源。

更改	描述	日期
AmazonKendraReadOnly—添加支持 GetSnapshots、BatchGetDocumentStatus API 的权限	Amazon Kendra 添加了新的 API `GetSnapshots` 和 `BatchGetDocumentStatus` 。`GetSnapshots` 提供的数据显示您的用户如何与您的搜索应用程序进行交互。`BatchGetDocumentStatus` 监控为文档编制索引的进度。	2022 年 1 月 3 日
AmazonKendraReadOnly—添加支持 GetQuerySuggestions 操作的权限	Amazon Kendra 添加了一个新的 API `GetQuerySuggestions` ，允许访问获取热门搜索查询的查询建议，从而帮助指导用户的搜索。当用户键入搜索查询时，建议的查询有助于自动完成搜索。	2021 年 5 月 27 日
Amazon Kendra 已开始跟踪更改	亚马逊 Kendra 开始跟踪其 AWS 托管政策的变更。	2021 年 5 月 27 日

Amazon Kendra 身份和访问问题排查

您可以使用以下信息，帮助诊断和修复在使用 Amazon Kendra 和 IAM 时可能遇到的常见问题。

主题

- 我无权在 Amazon Kendra 中执行操作
- 我无权执行 iam：PassRole
- 我是管理员并希望允许其他人访问 Amazon Kendra。
- 我想允许 AWS 账户以外的人访问我的 Amazon Kendra 资源

我无权在 Amazon Kendra 中执行操作

如果 AWS Management Console 告诉您您无权执行某项操作，则必须联系管理员寻求帮助。管理员是向您提供登录凭证的人。

当 `mateojackson` 用户尝试使用控制台查看有关资产的详细信息，但不具有 `kendra:DescribeIndex` 权限时，会发生以下示例错误。

```
User: arn:aws:iam::123456789012:user/mateojackson is not authorized to perform:
 kendra:DescribeIndex on resource: index ARN
```

在这种情况下，Mateo 请求他的管理员更新其策略，以允许他使用 `kendra:DescribeIndex` 操作访问 `index` 资源。

我无权执行 iam：PassRole

如果您收到一个错误，指明您无权执行 `iam:PassRole` 操作，则必须更新策略以允许您将角色传递给 Amazon Kendra。

有些 AWS 服务 允许您将现有角色传递给该服务，而不是创建新的服务角色或服务相关角色。为此，您必须具有将角色传递到服务的权限。

当名为 `marymajor` 的 IAM 用户尝试使用控制台在 Amazon Kendra 中执行操作时，会发生以下示例错误。但是，服务必须具有服务角色所授予的权限才可执行此操作。Mary 不具有将角色传递到服务的权限。

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

在这种情况下，必须更新 Mary 的策略以允许她执行 `iam:PassRole` 操作。

如果您需要帮助，请联系您的 AWS 管理员。您的管理员是提供登录凭证的人。

我是管理员并希望允许其他人访问 Amazon Kendra。

要允许其他人访问 Amazon Kendra，您必须为需要访问权限的人员或应用程序创建一个 IAM 实体（用户或角色）。它们将使用该实体的凭证访问 AWS。然后，您必须将策略附加到实体，以便在 Amazon Kendra 中为它们授予正确的权限。

要立即开始使用，请参阅《IAM 用户指南》中的创建您的第一个 IAM 委派用户和组。

我想允许 AWS 账户以外的人访问我的 Amazon Kendra 资源

您可以创建一个角色，以便其他账户中的用户或您组织外的人员可以使用该角色来访问您的资源。您可以指定谁值得信赖，可以担任角色。对于支持基于资源的策略或访问控制列表（ACL）的服务，您可以使用这些策略向人员授予对您的资源的访问权。

要了解更多信息，请参阅以下内容：

- 要了解 Amazon Kendra 是否支持这些功能，请参阅Amazon Kendra 如何与 IAM 协同工作。
- 要了解如何提供对您拥有的资源的访问权限 AWS 账户 ，请参阅 IAM 用户指南中的向您拥有 AWS 账户 的另一个 IAM 用户提供访问权限。
- 要了解如何向第三方提供对您的资源的访问权限 AWS 账户，请参阅 IAM 用户指南中的向第三方提供访问权限。 AWS 账户
- 要了解如何通过联合身份验证提供访问权限，请参阅《IAM 用户指南》中的为经过外部身份验证的用户（联合身份验证）提供访问权限。
- 要了解使用角色和基于资源的策略进行跨账户访问的区别，请参阅 IA M 用户指南中的跨账户资源访问。

安全最佳实操

Amazon Kendra 提供了在您开发和实施自己的安全策略时需要考虑的大量安全特征。以下最佳实操是一般准则，并不代表完整的安全解决方案。这些最佳实操可能不适合您的环境或不满足您的环境要求，请将其视为有用的考虑因素而不是惯例。

采用最低权限原则

Amazon Kendra 为使用角色的应用程序提供了精细的访问策略。 IAM 我们建议向该角色仅授予任务所需的最低权限集，例如覆盖应用程序和对日志目标的访问权限。我们还建议定期审核作业权限，并在应用程序发生任何更改时进行审核。

基于角色的访问控制 (RBAC) 权限

管理员应针对 Amazon Kendra 应用程序来严格控制基于角色的访问控制（RBAC）权限。

Amazon Kendra 中的日志记录和监控

要保持 Amazon Kendra 应用程序的可靠性、可用性和性能，监控是一个重要环节。要监控 Amazon Kendra API 调用，可以使用。 AWS CloudTrail要监控您的任务状态，请使用 Amazon CloudWatch 日志。

- Amazon CloudWatch Alarms — 使用 CloudWatch 警报，您可以监控您指定的时间段内的单个指标。该指标是否超过了策略。 CloudWatch 当指标处于特定状态时，警报不会调用操作。而是必须在状态已改变并在指定的若干个时间段内保持不变后才调用。有关更多信息，请参阅 使用 Amazon CloudWatch 监控 Amazon Kendra。
- AWS CloudTrail 日志 — CloudTrail 记录用户、角色或 AWS 服务在 Amazon Kendra 或 Amazon Kendra 智能排名中采取的操作。通过收集的信息 CloudTrail，您可以确定向 Amazon Kendra 发出的请求、发出请求的 IP 地址、谁提出了请求、何时提出请求以及其他详细信息。有关更多信息，请参阅 使用 AWS CloudTrail 日志记录 Amazon Kendra API 调用 和 使用 AWS CloudTrail 日志记录 Amazon Kendra Intelligent Ranking API 调用。

Amazon Kendra 的合规性验证

作为多项合 Amazon Kendra 规性计划的一部分，第三方审计员将评估 Amazon Kendra 的安全性和合规性。Amazon Kendra 符合以下规定：

- 健康保险流通与责任法案 (HIPAA)
- 系统和组织控制（SOC）2
- 信息安全注册评估员计划（IRAP）
- 联邦风险与授权管理项目（FedRAMP）在美国东部/西部地区适中
- 联邦风险和授权管理计划 (FedRAMP) 在 AW GovCloud S（美国西部）地区名列前茅

有关特定合规计划范围内的 AWS 服务列表，请参阅按合规计划划分的范围内的AWSAWS 服务按合规计划。有关一般信息，请参阅AWS 合规计划AWS。

您可以使用下载第三方审计报告 AWS Artifact。有关更多信息，请参阅在 Artifact 中下载报告在 AWS Ar 。

您使用 Amazon Kendra 的合规性责任取决于您数据的敏感度、贵公司的合规性目标以及适用的法律法规。 AWS 提供以下资源来帮助满足合规性：

- 安全与合规性快速入门指南 — 这些部署指南讨论了架构注意事项，并提供了在上部署以安全性和合规性为重点的基准环境的步骤。 AWS
- HIPAA 安全与合规架构白皮书 — 本白皮书描述了公司如何使用来创建符合 HIPAA 标准的应用程序。 AWS
- AWS 合规资源AWS-此工作簿和指南集合可能适用于您所在的行业和所在地区。
- 使用AWS Config 开发人员指南中的规则评估资源 — 该 AWS Config 服务评估您的资源配置在多大程度上符合内部实践、行业指导方针和法规。
- AWS Security Hub—此 AWS 服务可全面了解您的安全状态 AWS ，帮助您检查是否符合安全行业标准和最佳实践。

Amazon Kendra 中的恢复能力

AWS 全球基础设施是围绕 AWS 区域和可用区构建的。 AWS 区域提供多个物理隔离和隔离的可用区，这些可用区通过低延迟、高吞吐量和高度冗余的网络相连。利用可用区，您可以设计和操作在可用区之间无中断地自动实现失效转移的应用程序和数据库。与传统的单个或多个数据中心基础设施相比，可用区具有更高的可用性、容错性和可扩展性。

有关 AWS 区域和可用区的更多信息，请参阅AWS 全球基础设施。

借助 AWS 全球基础设施，Amazon Kendra 企业版具有容错能力、可扩展性和高可用性。目前不支持回滚到索引的先前版本，但是您可以通过删除现有数据来源并将其重新添加到索引中来刷新或重新创建索引的某些部分。

Amazon Kendra 中的基础设施安全性

作为一项托管服务，Amazon Kendra 受到 AWS 全球网络安全的保护。有关 AWS 安全服务以及如何 AWS 保护基础设施的信息，请参阅AWS 云安全。要使用基础设施安全的最佳实践来设计您的 AWS 环境，请参阅 S AWS ecurity Pillar Well-Architected Fram ework 中的基础设施保护。

您可以使用 AWS 已发布的 API 调用通过网络访问 Amazon Kendra。客户端必须支持以下内容：

- 传输层安全性协议 (TLS) 我们要求使用 TLS 1.2，建议使用 TLS 1.3。
- 具有完全向前保密 (PFS) 的密码套件，例如 DHE（临时 Diffie-Hellman）或 ECDHE（临时椭圆曲线 Diffie-Hellman）。大多数现代系统（如 Java 7 及更高版本）都支持这些模式。

此外，必须使用访问密钥 ID 和与 IAM 主体关联的秘密访问密钥来对请求进行签名。或者，您可以使用 AWS Security Token Service（AWS STS）生成临时安全凭证来对请求进行签名。

中的配置和漏洞分析 AWS Identity and Access Management

AWS 处理基本的安全任务，例如客户机操作系统 (OS) 和数据库修补、防火墙配置和灾难恢复。这些流程已通过相应第三方审核和认证。有关更多详细信息，请参阅以下资源：

- 责任共担模式
- AWS：安全流程概述（白皮书）

以下资源还涉及在 AWS Identity and Access Management (IAM) 中进行配置和漏洞分析：

- 合规性验证 AWS Identity and Access Management
- 中的@@ 安全最佳实践和用例 AWS Identity and Access Management。

的配额 Amazon Kendra

支持的 区域

有关可用 AWS 区域的列表，请参阅 Amazon Web Services 一般参考中的Amazon Kendra 区域和终端节点。 Amazon Kendra

配额

服务配额，也称为限制，是您 AWS 账户的最大服务资源数量。有关更多信息，请参阅《AWS 一般参考》中的 Amazon Kendra 服务限额。

索引配额

描述	默认	版本	可调整
每个账户的最大索引数	10	开发人员、企业	是
在单个单元中为索引提取的文本量（开发人员）。您无法为开发人员版本添加额外的单位来提取文本。	3 GB	开发人员	否
在单个单元中为索引提取的文本量（企业）。对于企业版本，您最多可以额外添加 100 个单位来提取文本，或者直接联系支持人员。	30 GB	企业	是

数据源连接器配额

描述	默认	版本	可调整
每个索引的最大数据源连接器数量（开发人员）	5	开发人员	否
每个索引的最大数据源连接器数量（企业）	50	企业	是
使用数据源连接器时单个文档或原始文件的最大大小	50 MB	开发人员、企业	是
Amazon S3 数据源连接器中包含的访问控制列表配置文件中的 S3 前缀的最大数量	100	开发人员、企业	否
Amazon S3 数据源连接器中包含的访问控制列表配置文件的最大大小	50 MB	开发人员、企业	是

常见问题配额

描述	默认	版本	可调整
每个索引的最大常见问题解答数	30	开发人员、企业	是
1 个常见问题解答的最大大小	5MB	开发人员、企业	是

描述	默认	版本	可调整
返回的常见问题解答的最大结果数量	4	开发人员、企业	是
常见问题解答问题允许的最大字符数	300	开发人员、企业	否
常见问题答案中的最大字符数	2000	开发人员、企业	否

同义词库配额

描述	默认	版本	可调整
每个索引的最大同义词库数	1	开发人员、企业	否
同义词库文件的最大大小	5MB	开发人员、企业	是
每个同义词库的最大同义词规则数	10000	开发人员、企业	是
索引中所有同义词库中每个术语的最大同义词数	10	开发人员、企业	否

Amazon Kendra 经验配额

描述	默认	版本	可调整
每个索引的最大 Amazon Kendra 体验数量	50	开发人员、企业	是

查询和搜索结果配额

描述	默认	版本	可调整
单个单元中每秒对索引的查询量（开发人员）。您无法为开发人员版本添加额外的单位来进行查询。	0.05	开发人员	否
单个单元中每秒对索引的查询量（企业）。对于企业版本，您最多可以额外添加 100 个单位来进行查询，或者直接联系支持人员。	0.1	企业	是
每个查询文本的最大字符数	1000	开发人员、企业	是
每个查询的最大搜索结果数量。默认值为 100。要获得超过 100 个结果，只需联系支持人员。	100	开发人员、企业	是
每个页面的最大搜索结果数量。	100	开发人员、企业	是
在截断之前，每个查询文本的最大令牌化单词数量。默认值为 30。要允许超过 30 个单词，只需联系支持人员。	30	开发人员、企业	是

描述	默认	版本	可调整
每个查询属性的最大用户组列表大小	1000	开发人员、企业	是
每个查询属性的最大字符串列表大小	10	开发人员、企业	是

查询建议配额

描述	默认	版本	可调整
每次 “建议” 调用返回的最大查询GetQuery建议数	10	开发人员、企业	是
每次 “建议” 调用时，查询建议的最大字段/属性数 GetQuery	10	开发人员、企业	是
每GetQuery次 “建议” 调用时，查询建议的附加字段/属性的最大数量	5	开发人员、企业	是
每个索引的最大阻止列表数	1	开发人员、企业	否
阻止列表文本文件的最大大小	2 MB	开发人员、企业	是
阻止列表中的最大项目（单词或短语）数量	20000	开发人员、企业	是

描述	默认	版本	可调整
Query API 调用返回的拼写更正查询建议的最大数量。	1	开发人员、企业	是

文件配额

描述	默认	版本	可调整
在单个单元中为索引提取的文本量（开发人员）。您无法为开发人员版本添加额外的单位来提取文本。	3 GB	开发人员	否
在单个单元中为索引提取的文本量（企业）。对于企业版本，您最多可以额外添加 100 个单位来提取文本，或者直接联系支持人员。	30 GB	企业	是
使用数据源连接器时单个文档或原始文件的最大大小	50 MB	开发人员、企业	是
使用 BatchPutDocument API 时单个文档或原始文件的最大大小	5MB	开发人员、企业	是
从单个文档中提取的最大文本量	5MB	开发人员、企业	否

描述	默认	版本	可调整
每个索引的最大自定义字段/属性数	500	开发人员、企业	否

精选搜索结果配额

描述	默认	版本	可调整
每个精选结果集的最大精选文档数量	4	企业	是
每个精选结果集的最大查询文本数量	49	企业	否
精选结果集中每个查询文本的最大字符数量	1000	企业	是
每个索引的最大精选结果集数量	50	企业	是

重新评分/重新排名搜索结果配额

描述	默认	版本	可调整
重新评分执行计划或单个容量单位每秒的最大 `Rescore` 请求数。您最多可以添加 1000 个额外单位。	0.01	企业	否
每个账户的最大重新评分执行计划数。	50	企业	是

描述	默认	版本	可调整
Rescore 请求中一个文档的 Title 中的最大令牌数。	100	企业	否
Rescore 请求中一个文档的 Body 中的最大令牌数。	200	企业	否
Rescore 请求中的最大文档数量。	25	企业	否
Rescore 请求中每个组的最大文档数量。	3	企业	否

有关 Amazon Kendra 服务配额的更多信息以及申请增加配额，请参阅 S ervice Quotas。

故障排除

本节可以帮助您解决在使用时可能遇到的常见问题 Amazon Kendra。

主题

数据来源故障排除

本节可以帮助您解决配置和使用 Amazon Kendra 数据源连接器时的常见问题。

我的文档没有编入索引

将 Amazon Kendra 索引与数据源同步时，可能会遇到导致无法对文档编制索引的问题。编制索引包括两个步骤。首先，检查数据来源中是否有要编制索引的新文档和更新文档，并查找要从索引中移除的文档。其次，确保可在文档级别访问每个文档和编制索引。

这两个步骤中都可能出现错误。数据来源级别的错误将在控制台的数据来源详细信息页面的同步运行历史记录部分报告。同步作业的状态可能是成功、未完成或失败。您还可以看到在作业执行期间已编制索引和删除的文档数量。如果状态为失败，则详细信息列中会显示一条消息。

中报告了文档级别的错误 Amazon CloudWatch Logs。您可以使用 CloudWatch 控制台查看错误。

要生成文档同步状态报告，请参阅我要为文档生成同步状态报告。

我的同步作业失败了

当索引或数据来源中出现配置错误时，同步作业通常会失败。在控制台中，您可以在数据来源详细信息页面的同步运行历史记录部分的详细信息列下找到错误消息。文档级别的错误在 Amazon CloudWatch Logs中报告。错误消息会提供有关问题的信息。问题通常是索引或数据源没有适当的 IAM 权限。错误消息会描述缺少的权限。以下是您可能会收到的一些错误消息：

```
Failed to create log group for job. Please make sure that the IAM role
provided has sufficient permissions.
```

如果您的索引角色没有使用权限 CloudWatch，则数据源将无法创建 CloudWatch 日志。如果出现此错误，则必须为索引角色添加 CloudWatch 权限。

`Failed to access Amazon S3 file prefix (bucket name) while trying to crawl your metadata files. Please make sure the IAM role (ARN) provided has sufficient permissions.`

使用 Amazon S3 数据源时， Amazon Kendra 必须具有访问包含文档的存储桶的权限。您需要向数据源 IAM 角色添加读取存储桶的权限。 Amazon Kendra

`The provided IAM role (ARN) could not be assumed. Please make sure Amazon Kendra is a trusted entity that is allowed to assume the role.`

Amazon Kendra 需要权限才能担任索引和数据源 IAM 角色。您需要向具有 `sts:AssumeRole` 操作的权限的角色添加信任策略。

有关 Amazon Kendra 需要为数据源编制索引的 IAM 策略，请参阅IAM 角色。

要生成文档同步状态报告，请参阅我要为文档生成同步状态报告。

我的同步任务未完成

通常，如果完成了数据来源级别的流程，但在文档级别的处理过程中出现了一些错误，则不会完成作业。如果作业未完成，则某些文档可能无法成功编制索引。对于 Amazon S3 数据来源，作业未完成通常是由以下原因造成的：

- 一个或多个文档的元数据无效。
- 当提交文档以编制索引时，至少有一个文档未提交。
- 当提交要从索引中删除的文档时，至少有一个文档未提交。

要对未完成的同步作业进行故障排除，请先查看您的 CloudWatch 日志。

1. 在详细信息列中，选择查看详细信息 CloudWatch。
2. 查看错误消息，以便了解导致文档失败的原因。

要生成文档同步状态报告，请参阅我要为文档生成同步状态报告。

我的同步作业执行成功了，但没有编制了索引的文档

有时，运行的索引同步作业会被标记为成功，但没有按预期为新文档或更新的文档编制索引。可能的原因包括：

- 检查 CloudWatch `DocumentsSubmittedForIndexingFailed`指标以查看是否有任何文档无法同步。请查看您的 CloudWatch 日志以了解详细信息。
- 对于 Amazon S3 数据源，您可能给出了错误 Amazon Kendra 的存储桶名称或前缀。确保 Amazon Kendra 正在使用的存储桶是包含要索引的文档的存储桶。
- 在为之前的作业中未能编制索引的文档重新编制索引时，除非您更改了该文档或与其关联的元数据文件，否则 Amazon Kendra 不会为其编制索引。

要生成文档同步状态报告，请参阅我要为文档生成同步状态报告。

我在同步数据来源时遇到了文件格式问题

如果您在向数据来源添加文件或同步数据来源时遇到文件格式问题，请确保 Amazon Kendra 支持您的文档类型。有关支持的文档类型的列表， Amazon Kendra 请参阅文档类型或格式。

如果您将 `BatchPutDocument` API 用于纯文本文件，请将 `PLAIN_TEXT` 指定为内容类型。

我要为文档生成同步历史记录报告

同步 Amazon Kendra 数据源连接器时， Amazon Kendra 可以为数据源中的每个文档生成同步状态报告并将其复制到 Amazon S3 存储桶。在此过程中，您的数据将使用 AWS KMS 密钥进行加密，并且只能由您查看。报告的文档状态包括：失败、已完成或成功但有错误。

您必须先执行以下操作，然后才能生成同步状态报告：

- 将以下 Amazon Kendra 服务主体添加到您的 Amazon S3 访问策略中

```
{
    “Version”: “2012-10-17",
    “Statement”: [
        {
            “Sid”: “KendraS3Access”,
            “Effect”: “Allow”,
            “Principal”: {
                “Service”: “kendra.amazonaws.com”
            },
```

```
            “Action”: “s3:PutObject”,
            “Resource”: “arn:aws:s3:::your-manifest-bucket-name/*”
        }
    ]
}
```

- 创建具有访问权限的 Amazon S3 存储桶 Amazon Kendra

如果您使用控制台，要生成同步状态报告，请从数据来源详细信息页面进行选择以激活同步历史记录生成选项。然后，输入 Amazon S3 存储桶位置并从可用的配置选项中进行选择。激活报告的生成后，下次同步时即可生成报告。

如果您删除 Amazon S3 存储桶，您将丢失日志数据，并且必须设置新的存储桶来存储新的同步报告。

目前只有 Amazon S3 连接器支持生成同步报告状态。

同步数据来源需要多长时间？

如果文档没有更新，则 Amazon Kendra 索引的同步时间与文档数量成线性增加。例如，1,000 个没有任何更新的文档大约需要五分钟完成同步，而 2,000 份没有任何更新的文档则大约需要 10 分钟。如果文档有更新，则同步时间将根据更新的文档数量而增加。

同步数据来源的费用是多少？

同步索引时，需要两分钟的时间进行预热和激活 Amazon EC2 以建立必要的连接。在此过程中，您无需支付任何费用。在同步作业开始后，您的使用量表才会启动。有关 Amazon Kendra 定价的更多信息，请参阅Amazon Kendra 定价。

我收到 Amazon EC2 授权错误

如果在虚拟私有云 (VPC) 数据源的同步过程中出现 Amazon EC2 未经授权的操作错误，则可能是您的 VPC IAM 角色缺少所需的权限。请检查您用于数据源的 IAM 角色是否具有附加权限。有关更多信息，请参阅虚拟私有云 IAM 角色。

我无法使用搜索索引链接来打开我的 Amazon S3 对象

您的 Amazon Kendra 索引只能访问 Amazon S3 数据源授予其访问权限的文件。例如， Amazon Kendra 无法修改决定对象是公开还是加密的 Amazon S3 权限。 Amazon Kendra 也没有为 Amazon S3 对象创建或返回签名链接的默认权限。如果要为 Amazon Kendra 索引中的 Amazon S3 对象激活签名链接，则有两种选择：

- 在将索引查询结果返回搜索页面之前，您可以使用源 URI 对象对索引查询结果进行签名。有关此过程的 step-by-step演练，请参阅使用预签名 URL 共享对象。
- 您可以覆盖 Amazon S3 对象元数据源 uri，并通过连接到 Amazon S3 存储桶的 CloudFront 内容分发网络 (CDN) 提供服务。或者，您可以使用返回预签名 URL 并重定向到该地址的 API Gateway 代理端点。

我收到 “使用 SSL 证书文件AccessDenied 时” 错误消息

如果在数据源中使用 SSL 证书时遇到拒绝访问错误，请确保您的 IAM 角色有权访问指定位置的 SSL 证书文件。如果证书使用 AWS KMS 密钥加密，则您的 IAM 角色还应有权使用该 AWS KMS 密钥进行解密。有关更多信息，请参阅 AWS KMS的身份验证和访问控制。

使用 SharePoint 数据源时出现授权错误

如果您在将索引与 SharePoint 数据源同步时遇到授权错误，请确认已在中 SharePoint为您分配了站点管理员角色。

索引无法从我的 Confluence 数据来源中爬取文档

如果在同步过程中，您的 Amazon Kendra 索引没有从 Confluence 数据源中搜寻文档，请确认您是 Confluence 管理员群组的成员。

文档搜索结果故障排除

本部分可以帮助您修复 Amazon Kendra 搜索结果中的问题。

搜索结果与我的搜索查询无关

如果您的搜索结果似乎不相关，可能是出于以下原因：

- 结果中包含 LOW 置信度结果。您可以LOW放心地筛选出结果，方法是使用 QueryResultItem's ScoreAttributes 字段排除任何值为的结果LOW。 Amazon Kendra 为每个结果分配一个置信区值VERY_HIGH，即HIGH、MEDIUM和。LOW这些值表示结果与查询相关的可信度。此外，无论置信度区间如何，都会按以下顺序 Amazon Kendra 返回三种类型的结果：ANSWER（建议答案摘录）、QUESTION_ANSWER（常见问题解答）和DOCUMENT（文档摘录）。因此，可以将 LOW 置信度 QUESTION_ANSWER 结果置于 VERY_HIGH 置信度 DOCUMENT 结果之上。但是，LOW 置信度 QUESTION_ANSWER 更好的结果并不一定总是比 VERY_HIGH 置信度 DOCUMENT 结果更好。

- 某些元数据字段或属性会被提升到非常高的值，从而影响结果的排名。 Amazon Kendra 使用多个参数（例如文档标题、文本、日期和自定义文本字段或属性）搜索您的索引。您可以尝试使用不同的提升值，以便在所有查询中获得最佳结果。您还可以在查询级别使用动态相关性调整，为每个查询使用不同的提升值。
- 您的用户在查询信息时使用的是专门的术语，并且没有为索引设置自定义同义词来处理这些专业术语。有关如何以及何时使用同义词的更多详细信息，请参阅向索引添加自定义同义词。

为什么我只能看到 100 个结果？

Amazon Kendra 返回相关文档的总数。默认情况下，每次查询都会返回前 100 个结果。结果将进行分页。您可以使用 `PageNumber` 来访问不同的页面。

您可以配置 Amazon Kendra 为每次查询最多返回 1,000 个文档或搜索结果，每页最多返回 100 个结果。要返回 100 个以上的结果，您可以联系 配额支持团队来申请提高配额。增加搜索结果的数量可能会影响延迟。

为什么没有我预计会看到的文档？

Amazon Kendra 支持基于用户和组的访问控制列表 (ACL)。 Amazon Kendra 通过连接器提取 ACL 策略。如果索引未配置 ACL，则只会显示与用户和组的属性筛选条件相匹配的文档。如果提供了用户或组属性筛选条件，则不会显示没有 ACL 的文档。

如果您使用的是基于令牌的访问控制，则会显示没有 ACL 策略的文档以及与用户和组匹配的文档。

为什么我会看到具有 ACL 策略的文档？

如果索引未配置访问控制策略，则可以通过筛选条件提供用户和组。如果未应用用户和组筛选条件，则会返回所有相关文档。任何 ACL 策略都将被忽略。

排查一般问题

Amazon Kendra 使用 CloudWatch 指标和日志来提供有关同步数据源的见解。您可以使用指标和日志来确定同步运行时出了什么问题以及如何修复。

要进行一般故障排除，请从您的 CloudWatch 指标开始。

- 查看 `DocumentsCrawled` 指标以查看您的数据来源检查了多少文档。对于 Amazon S3 存储桶，如果该数字小于您的预期，请检查您的数据源是否指向正确的存储桶。

- 查看 DocumentsSkippedNoChange 指标以了解跳过了多少文档，因为自上次同步以来，这些文档没有发生更改。如果数字与预期数字不符，请检查您的存储库是否已正确更新。
- 查看 DocumentsSkippedInvalidMetadata 指标以了解有多少文档包含无效的元数据。查看您的 CloudWatch 日志，查看发生的具体错误。
- 查看 DocumentsSubmittedForIndexingFailed 指标以了解有多少文档已从数据来源发送到索引，但未能编制索引。例如，如果您在尚未定义为自定义索引字段的 Amazon S3 数据来源中使用元数据属性，则不会为该文档编制索引。查看您的 CloudWatch 日志，查看发生的具体错误。
- 查看 DocumentsSubmittedForDeletionFailed 指标以了解有多少数据来源尝试从索引中移除的文档未能从索引中删除。查看您的 CloudWatch 日志，查看发生的具体错误。

您可以查看特定同步运行的 CloudWatch 日志，以获取运行期间发生的错误的详细信息。有关使用 CloudWatch 日志的更多信息 Amazon Kendra，请参阅CloudWatch Logs。

Amazon Kendra 智能分层

Amazon Kendra 智能排名使用 Amazon Kendra 语义搜索功能对搜索服务的结果进行智能重新排名。

主题

Amazon Kendra 自我管理的智能排名 OpenSearch

您可以利用 Amazon Kendra基于 Apache 2.0 许可证的自我管理开源搜索服务的语义搜索功能来改善搜索结果。OpenSearch Amazon Kendra 智能排名插件在语义上使用对结果进行重新排名 OpenSearch。 Amazon Kendra它通过使用默认搜索结果中的特定字段（例如文档正文或标题）来理解 OpenSearch 搜索查询的含义和上下文。

以此查询为例：“main keynote address”。由于 “地址” 有多种含义，因此 Amazon Kendra 可以推断出查询背后的含义以返回与预期含义一致的相关信息。在这里的上下文中，它是指会议主题演讲。例如，更简单的搜索服务可能不会考虑实际意图，而是返回 Main Street 的街道地址作为结果。

的智能排名插件可用 OpenSearch 于 OpenSearch （自我管理）版本 2.4.0 及更高版本。您可以使用快速入门 Bash 脚本安装插件，以构建包含智能排名插件 OpenSearch 的新 Docker 镜像。请参阅 设置智能搜索插件 - 这是一个帮助您快速启动和运行的设置示例。

智能搜索插件的工作原理

OpenSearch （自我管理）智能排名插件的整体流程如下：

1. OpenSearch 用户发出查询，并 OpenSearch 提供查询响应或与查询相关的文档列表。
2. 智能排名插件获取查询响应并从文档中提取信息。
3. 智能排名插件调用 Amazon Kendra 智能排名的 Rescore API。
4. `Rescore` API 使用从文档中提取的信息并从语义上对搜索结果进行重新排名。
5. `Rescore` API 将重新排名的搜索结果发送回插件。该插件会重新排列搜索响应中的 OpenSearch 搜索结果，以反映新的语义排名。

智能排名插件使用“body”和“title”字段对结果进行重新排名。这些插件字段可以映射到 OpenSearch 索引中最符合文档正文和标题定义的字段。例如，如果您的索引包含一本书的章节，其中包

含“chapter_heading”和“chapter_contents”之类的字段，则可以将前者映射到“title”，将后者映射到“body”以获得最佳结果。

设置智能搜索插件

以下内容概述了如何使用智能排名插件快速设置 OpenSearch （自我管理）。

使用智能排名插件进行设置 OpenSearch （自我管理）（快速设置）

如果你已经在使用 Docker 镜像`opensearch:2.4.0`，你可以使用这个 Dockerfile 通过智能排名插件构建 OpenSearch 2.4.0 的新镜像。在 docker-compose.yml 文件或 opensearch.yml 文件中包含一个用于存放新映像的容器。您还需要包含在创建重新评分执行计划时生成的重新评分执行计划 ID，以及您的区域和端点信息，请参阅有关创建重新分数执行计划的步骤 2。

如果您之前下载的 `opensearch` Docker 映像版本低于 2.4.0，则必须使用 Docker 映像 `opensearch:2.4.0` 或更高版本，并使用随附的智能排名插件构建新映像。

1. 下载并安装适用于您的操作系统的 Docker 桌面。Docker 桌面包括 Docker Compose 和 Docker 引擎。建议检查您的计算机是否满足 Docker 安装详细信息中所述的系统要求。

   您还可以在 Docker 桌面的设置中提高内存使用要求。超出免费提供的 Docker 服务使用限制后，您应对 Docker 的使用要求负责。请参阅 Docker 订阅。

   检查 Docker 桌面状态是否为“正在运行”。

2. 配置 Amazon Kendra 智能排名和您的容量需求。预配 Amazon Kendra 智能排名后，将根据您设定的容量单位按小时收费。查看免费套餐和价格信息。

   您可以使用 CreateRescoreExecutionPlanAPI 来配置`Rescore API`。如果您不需要比单个单位默认值更多的容量单位，请不要添加更多单位，只需提供重新评分执行计划的名称。您也可以使用 UpdateRescoreExecutionPlanAPI 更新容量需求。有关更多信息，请参阅从语义上对搜索服务的结果进行排名。

   或者，在运行快速入门 Bash 脚本时，您可以转到步骤 3 来创建默认的重新评分执行计划。

   请注意步骤 4，响应中包含的重新评分执行计划 ID。

   CLI

   ```
   aws kendra-ranking create-rescore-execution-plan \
     --name MyRescoreExecutionPlan \
   ```

```
  --capacity-units '{"RescoreCapacityUnits":<integer number of additional
 capacity units>}'

Response:

{
    "Id": "<rescore execution plan ID>",
    "Arn": "arn:aws:kendra-ranking:<region>:<account-id>:rescore-execution-plan/
<rescore-execution-plan-id>"
}
```

Python

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra_ranking = boto3.client("kendra-ranking")

print("Create a rescore execution plan.")

# Provide a name for the rescore execution plan
name = "MyRescoreExecutionPlan"
# Set your required additional capacity units
# Don't set capacity units if you don't require more than 1 unit given by
 default
capacity_units = 1

try:
    rescore_execution_plan_response =
 kendra_ranking.create_rescore_execution_plan(
        Name = name,
        CapacityUnits = {"RescoreCapacityUnits":capacity_units}
    )

    pprint.pprint(rescore_execution_plan_response)

    rescore_execution_plan_id = rescore_execution_plan_response["Id"]

    print("Wait for Amazon Kendra to create the rescore execution plan.")

    while True:
```

```
        # Get the details of the rescore execution plan, such as the status
        rescore_execution_plan_description =
 kendra_ranking.describe_rescore_execution_plan(
            Id = rescore_execution_plan_id
        )
        # When status is not CREATING quit.
        status = rescore_execution_plan_description["Status"]
        print(" Creating rescore execution plan. Status: "+status)
        time.sleep(60)
        if status != "CREATING":
            break

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

3. 从主分支下拉列表中选择版本分支，从中下载 GitHub 适用于您版本 OpenSearch的快速入门 Bash 脚本。

   此脚本使用 Docker 映像 OpenSearch 和 OpenSearch 仪表板，使用您在 GitHub 存储库中为脚本选择的版本。它会下载智能排名插件的 zip 文件，并生成一个`Dockerfile`用于构建包含 OpenSearch 该插件的新 Docker 镜像。它还会创建一个 docker-compose.yml 文件，其中包含 OpenSearch 带有智能排名插件和仪表板的容器。 OpenSearch 此脚本会将您的重新评分执行计划 ID、区域信息和端点（使用该区域）添加到 docker-compose.yml 文件中。然后，脚本运行`docker-compose up`以启动包含智能排名和 OpenSearch 仪表板的容器。 OpenSearch 要停止容器而不将其移除，请运行 `docker-compose stop`。要移除容器，请运行 `docker-compose down`。

4. 打开终端，在 Bash 脚本的目录中运行以下命令。

   ```
   bash search_processing_kendra_quickstart.sh -p <execution-plan-id> -r <region>
   ```

   运行此命令时，您需要提供您在步骤 2 中预置 Amazon Kendra 智能排名时记下的重新评分执行计划 ID 以及您的区域信息。或者，您也可以改为使用 `--create-execution-plan` 选项预配置 Amazon Kendra 智能排名。这会创建一个具有默认名称和默认容量的重新评分执行计划。

   为了在移除默认的临时容器时不丢失索引，您可以使用 `--volume-name` 选项提供数据卷名称，从而使索引在执行期间保持不变。如果您之前创建了索引，则可以在 docker-compose.yml 或 opensearch.yml 文件中指定卷。要保持卷完整，请不要运行 `docker-compose down -v`。

快速入门 Bash 脚本在 OpenSearch 密钥库中配置您的 AWS 凭据以连接到智能排名。 Amazon Kendra 要向脚本提供 AWS 凭据，请使用`--profile`选项指定 AWS 配置文件。如果未指定该`--profile`选项，则快速启动 Bash 脚本将尝试从环境变量和默认配置文件中读取 AWS 凭据（访问/密钥、可选会话令牌）。 AWS 如果未指定该`--profile`选项且未找到凭据，则脚本将不会将凭据传递给 OpenSearch 密钥库。如果 OpenSearch 密钥库中未指定凭据，则该插件仍会检查默认凭证提供程序链中的凭证，包括 Amazon ECS 容器凭据或通过 Amazon EC2 元数据服务提供的实例配置文件凭证。

请确保您已创建具有调用 Amazon Kendra 智能排名的必要权限的 IAM 角色。以下是授予将`Rescore` API 用于特定重新分数执行计划的权限的 IAM 策略示例：

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "kendra-ranking:Rescore",
            "Resource": "arn:aws:kendra-ranking:${Region}:${Account}:rescore-
execution-plan/${RescoreExecutionPlanId}"
        }
    ]
}
```

docker-compose.yml 的示例

使用 OpenSearch 2.4.0 或更高版本以及智能排名插件和仪表板 2.4.0 或更高版本的 docker-compose.yml 文件示例。 OpenSearch

```
version: '3'
networks:
  opensearch-net:
volumes:
  <volume-name>:
services:
  opensearch-node:
    image: <Docker image tag name of OpenSearch with Intelligent Ranking plugin>
    container_name: opensearch-node
    environment:
      - cluster.name=opensearch-cluster
      - node.name=opensearch-node
```

```
      - discovery.type=single-node
      - kendra_intelligent_ranking.service.endpoint=https://kendra-
ranking.<region>.api.aws
      - kendra_intelligent_ranking.service.region=<region>
      - kendra_intelligent_ranking.service.execution_plan_id=<rescore-execution-plan-
id>
    ulimits:
      memlock:
        soft: -1
        hard: -1
      nofile:
        soft: 65536
        hard: 65536
    ports:
      - 9200:9200
      - 9600:9600
    networks:
      - opensearch-net
  volumes:
    <docker-volume-name>:/usr/share/opensearch/data
  opensearch-dashboard:
   image: opensearchproject/opensearch-dashboards:<your-version>
   container_name: opensearch-dashboards
   ports:
     - 5601:5601
   environment:
     OPENSEARCH_HOSTS: '["https://opensearch-node:9200"]'
   networks:
     - opensearch-net
```

Dockerfile 和构建映像的示例

在智能排名插件中使用 OpenSearch 2.4.0 或更高版本的示例。Dockerfile

```
FROM opensearchproject/opensearch:<your-version>
RUN /usr/share/opensearch/bin/opensearch-plugin install --batch  https://github.com/
opensearch-project/search-processor/releases/download/<your-version>/search-
processor.zip
```

OpenSearch 使用智能排名插件构建 Docker 镜像。

```
docker build --tag=<Docker image tag name of OpenSearch with Intelligent Ranking
 plugin>
```

与智能搜索插件交互

使用智能排名插件进行设置 OpenSearch （自我管理）后，即可使用 curl 命令或 OpenSearch 客户端库与该插件进行交互。使用智能排名插件进行访问 OpenSearch 的默认凭据是用户名 “admin” 和密码 “admin”。

要将智能排名插件设置应用于 OpenSearch 索引，请执行以下操作：

Curl

```
curl -XPUT "https://localhost:9200/<your-docs-index>/_settings" -u 'admin:admin' --
insecure -H 'Content-Type: application/json' -d'
{
  "index": {
    "plugin" : {
      "searchrelevance" : {
        "result_transformer" : {
          "kendra_intelligent_ranking": {
              "order": 1,
              "properties": {
                "title_field": "title_field_name_here",
                "body_field": "body_field_name_here"
              }
          }
        }
      }
    }
  }
}
'
```

Python

```
pip install opensearch-py

from opensearchpy import OpenSearch
host = 'localhost'
port = 9200
```

```
auth = ('admin', 'admin')

client = OpenSearch(
    hosts = [{'host': host, 'port': port}],
    http_compress = True, # enables gzip compression for request bodies
    http_auth = auth,
    # client_cert = client_cert_path,
    # client_key = client_key_path,
    use_ssl = True,
    verify_certs = False,
    ssl_assert_hostname = False,
    ssl_show_warn = False,
    ca_certs = ca_certs_path
)

setting_body = {
    "index": {
        "plugin" : {
            "searchrelevance" : {
                "result_transformer" : {
                    "kendra_intelligent_ranking": {
                            "order": 1,
                            "properties": {
                                "title_field": "title_field_name_here",
                                "body_field": "body_field_name_here"
                            }
                    }
                }
            }
        }
    }
}

response = client.indices.put_settings(index_name, body=setting_body)
```

必须包括要用于重新排名的主文本字段的名称，例如，文档正文或文档内容字段。您还可以包括其他文本字段，例如，文档标题或文档摘要。

现在，您可以发出任何查询，并使用智能排名插件对结果进行排名。

Curl

```
curl -XGET "https://localhost:9200/<your-docs-index>/_search?pretty" -u
 'admin:admin' --insecure -H 'Content-Type: application/json' -d'
{
  "query" : {
    "match" : {
      "body_field_name_here": "intelligent systems"
    }
  }
}
'
```

Python

```
from opensearchpy import OpenSearch
host = 'localhost'
port = 9200
auth = ('admin', 'admin')

client = OpenSearch(
    hosts = [{'host': host, 'port': port}],
    http_compress = True, # enables gzip compression for request bodies
    http_auth = auth,
    # client_cert = client_cert_path,
    # client_key = client_key_path,
    use_ssl = True,
    verify_certs = False,
    ssl_assert_hostname = False,
    ssl_show_warn = False,
    ca_certs = ca_certs_path
)

query = {
  'size': 10,
  "query" : {
    "match" : {
      "body_field_name_here": "intelligent systems"
    }
  }
}

response = client.search(
```

```
    body = query,
    index = index_name
)

print('\nSearch results:')
print(response)
```

要移除 OpenSearch 索引的智能排名插件设置，请执行以下操作：

Curl

```
curl -XPUT "http://localhost:9200/<your-docs-index>/_settings" -H 'Content-Type:
 application/json' -d'
{
  "index": {
    "plugin": {
      "searchrelevance": {
        "result_transformer": {
          "kendra_intelligent_ranking.*": null
        }
      }
    }
  }
}
'
```

Python

```
from opensearchpy import OpenSearch
host = 'localhost'
port = 9200
auth = ('admin', 'admin')

client = OpenSearch(
    hosts = [{'host': host, 'port': port}],
    http_compress = True, # enables gzip compression for request bodies
    http_auth = auth,
    # client_cert = client_cert_path,
    # client_key = client_key_path,
    use_ssl = True,
    verify_certs = False,
    ssl_assert_hostname = False,
```

```
    ssl_show_warn = False,
    ca_certs = ca_certs_path
)

setting_body = {
  "index": {
    "plugin": {
      "searchrelevance": {
        "result_transformer": {
          "kendra_intelligent_ranking.*": null
        }
      }
    }
  }
}

response = client.indices.put_settings(index_name, body=setting_body)
```

要对特定查询测试智能排名插件或对某些正文和标题字段进行测试，请执行以下操作：

Curl

```
curl -XGET "https://localhost:9200/<your-docs-index>/_search?pretty" -u
 'admin:admin' --insecure -H 'Content-Type: application/json' -d'
{
  "query": {
    "multi-match": {
      "query": "intelligent systems",
      "fields": ["body_field_name_here", "title_field_name_here"]
    }
  },
  "size": 25,
  "ext": {
    "search_configuration": {
      "result_transformer": {
        "kendra_intelligent_ranking": {
          "order": 1,
          "properties": {
            "title_field": "title_field_name_here",
            "body_field": "body_field_name_here"
          }
        }
```

```
      }
    }
  }
}
'
```

Python

```
from opensearchpy import OpenSearch
host = 'localhost'
port = 9200
auth = ('admin', 'admin')

client = OpenSearch(
    hosts = [{'host': host, 'port': port}],
    http_compress = True, # enables gzip compression for request bodies
    http_auth = auth,
    # client_cert = client_cert_path,
    # client_key = client_key_path,
    use_ssl = True,
    verify_certs = False,
    ssl_assert_hostname = False,
    ssl_show_warn = False,
    ca_certs = ca_certs_path
)

# Index settings null for kendra_intelligent_ranking

query = {
  "query": {
    "multi_match": {
      "query": "intelligent systems",
      "fields": ["body_field_name_here", "title_field_name_here"]
    }
  },
  "size": 25,
  "ext": {
    "search_configuration": {
      "result_transformer": {
        "kendra_intelligent_ranking": {
          "order": 1,
          "properties": {
            "title_field": "title_field_name_here",
```

```
            "body_field": "body_field_name_here"
          }
        }
      }
    }
  }
}

response = client.search(
    body = query,
    index = index_name
)

print('\nSearch results:')
print(response)
```

将 OpenSearch 结果与 Amazon Kendra 结果进行比较

您可以将 side-by-side OpenSearch （自我管理）排名结果与重新排名的结果进行比较。 Amazon Kendra OpenSearch 仪表板版本 2.4.0 及更高版本提供了 side-by-side 结果，因此您可以将文档的 OpenSearch 排名方式 Amazon Kendra 与插件对搜索查询中的文档排名方式进行比较。

在将 OpenSearch 排名结果与 Amazon Kendra 重新排名的结果进行比较之前，请确保您的 OpenSearch 仪表板由带有智能排名插件的 OpenSearch 服务器提供支持。您可以使用 Docker 和快速入门 Bash 脚本进行设置。请参阅 设置智能搜索插件。

以下内容概述了如何在 OpenSearch 仪表板中比较 OpenSearch 和 Amazon Kendra 搜索结果。有关更多信息，请参阅OpenSearch文档。

比较 OpenSearch 仪表板中的搜索结果

1. 打开 http://localhost:5601 并登录 OpenSearch 控制面板。默认凭证是用户名“admin”和密码“admin”。
2. 从导航菜单的 OpenSearch 插件中选择 “搜索相关性”。
3. 在搜索栏中输入搜索文本。
4. 为查询 1 选择您的索引，然后在查询 DSL 中输入 OpenSearch 查询。您可以使用 `%SearchText%` 变量来引用在搜索栏中输入的搜索文本。有关此查询的示例，请参阅OpenSearch 文档。此查询返回的结果是未使用智能排名插件的 OpenSearch 结果。

5. 为查询 2 选择相同的索引，然后在查询 DSL 中输入相同的 OpenSearch 查询。此外，在扩展中包含 `kendra_intelligent_ranking` 并指定作为排名条件的必需 `body_field`。您也可以指定标题字段，但正文字段是必需字段。有关此查询的示例，请参阅OpenSearch 文档。此查询返回的结果是使用智能 Amazon Kendra 排名插件重新排名的结果。该插件最多可对 25 个结果进行排名。
6. 选择搜索以返回和比较结果。

从语义上对搜索服务的结果进行排名

Amazon Kendra 智能排名使用 Amazon Kendra语义搜索功能对搜索服务的结果进行重新排名。它通过考虑搜索查询的上下文以及搜索服务文档中的所有可用信息来做到这一点。 Amazon Kendra 智能排名可以改善简单的关键字匹配。

CreateRescoreExecutionPlanAPI 会创建用于配置 Rescore API 的 Amazon Kendra 智能排名资源。`Rescore`API 会对来自搜索服务（例如 OpenSearch （自我管理）的搜索结果进行重新排名。

调用 `CreateRescoreExecutionPlan` 时，您可以设置所需的容量单位，以便对搜索服务的结果进行重新排名。如果您不需要超过单个单位默认值的容量单位，请不要更改默认值。只需为重新评分执行计划提供一个名称。您最多可以设置 1000 个额外单位。有关单个容量单位中包含的内容的信息，请参阅调整容量。配置 Amazon Kendra 智能排名后，将根据您设定的容量单位按小时收费。查看免费套餐和价格信息。

调用 `CreateRescoreExecutionPlan` 时，系统会生成一个重新评分执行计划 ID，并在响应中返回。`Rescore` API 使用重新评分执行计划 ID，按照您设置的容量对搜索服务的结果进行重新排名。您可以在搜索服务的配置文件中包含重新评分执行计划 ID。例如，如果您使用 OpenSearch （自我管理），则可以在您的 docker-compose.yml 或 opensearch.yml 文件中包含重新分数执行计划 ID，请参阅智能排名（自助服务）结果。 OpenSearch

调用 `CreateRescoreExecutionPlan` 时，还会在响应中生成 Amazon 资源名称（ARN）。您可以使用此 ARN 在 AWS Identity and Access Management (IAM) 中创建权限策略，以限制用户访问特定重新评分执行计划的特定 ARN。有关授予将 `Rescore` API 用于特定重新评分执行计划的权限的 IAM 策略示例，请参阅自我 OpenSearch管理的Amazon Kendra 智能排名。

下面是创建容量单位设置为 1 的重新评分执行计划的示例。

CLI

```
aws kendra-ranking create-rescore-execution-plan \
  --name MyRescoreExecutionPlan \
```

```
  --capacity-units '{"RescoreCapacityUnits":1}'

Response:

{
    "Id": "<rescore execution plan ID>",
    "Arn": "arn:aws:kendra-ranking:<region>:<account-id>:rescore-execution-plan/
<rescore-execution-plan-id>"
}
```

Python

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra_ranking = boto3.client("kendra-ranking")

print("Create a rescore execution plan.")

# Provide a name for the rescore execution plan
name = "MyRescoreExecutionPlan"
# Set your required additional capacity units
# Don't set capacity units if you don't require more than 1 unit given by default
capacity_units = 1

try:
    rescore_execution_plan_response = kendra_ranking.create_rescore_execution_plan(
        Name = name,
        CapacityUnits = {"RescoreCapacityUnits":capacity_units}
    )

    pprint.pprint(rescore_execution_plan_response)

    rescore_execution_plan_id = rescore_execution_plan_response["Id"]

    print("Wait for Amazon Kendra to create the rescore execution plan.")

    while True:
        # Get the details of the rescore execution plan, such as the status
        rescore_execution_plan_description =
 kendra_ranking.describe_rescore_execution_plan(
```

```
            Id = rescore_execution_plan_id
        )
        # When status is not CREATING quit.
        status = rescore_execution_plan_description["Status"]
        print(" Creating rescore execution plan. Status: "+status)
        time.sleep(60)
        if status != "CREATING":
            break

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

```
import java.util.concurrent.TimeUnit;

import software.amazon.awssdk.services.kendraranking.KendraRankingClient;
import
 software.amazon.awssdk.services.kendraranking.model.CapacityUnitsConfiguration;
import
 software.amazon.awssdk.services.kendraranking.model.CreateRescoreExecutionPlanRequest;
import
 software.amazon.awssdk.services.kendraranking.model.CreateRescoreExecutionPlanResponse;
import
 software.amazon.awssdk.services.kendraranking.model.DescribeRescoreExecutionPlanRequest;
import
 software.amazon.awssdk.services.kendraranking.model.DescribeRescoreExecutionPlanResponse;
import
 software.amazon.awssdk.services.kendraranking.model.RescoreExecutionPlanStatus;

public class CreateRescoreExecutionPlanExample {

  public static void main(String[] args) throws InterruptedException {

    String rescoreExecutionPlanName = "MyRescoreExecutionPlan";
    int capacityUnits = 1;

    KendraRankingClient kendraRankingClient = KendraRankingClient.builder().build();

    System.out.println(String.format("Creating a rescore execution plan named %s",
 rescoreExecutionPlanName));
```

```
    CreateRescoreExecutionPlanResponse createResponse =
 kendraRankingClient.createRescoreExecutionPlan(
        CreateRescoreExecutionPlanRequest.builder()
            .name(rescoreExecutionPlanName)
            .capacityUnits(
                CapacityUnitsConfiguration.builder()
                    .rescoreCapacityUnits(capacityUnits)
                    .build()
            )
            .build()
    );

    String rescoreExecutionPlanId = createResponse.id();
    System.out.println(String.format("Waiting for rescore execution plan with id %s
 to finish creating.", rescoreExecutionPlanId));
    while (true) {
      DescribeRescoreExecutionPlanResponse describeResponse =
 kendraRankingClient.describeRescoreExecutionPlan(
          DescribeRescoreExecutionPlanRequest.builder()
              .id(rescoreExecutionPlanId)
              .build()
      );
      RescoreExecutionPlanStatus rescoreExecutionPlanStatus =
 describeResponse.status();
      if (rescoreExecutionPlanStatus != RescoreExecutionPlanStatus.CREATING) {
        break;
      }
      TimeUnit.SECONDS.sleep(60);
    }

    System.out.println("Rescore execution plan creation is complete.");
  }
}
```

下面是更新重新评分执行计划以将容量单位设置为 2 的示例。

CLI

```
aws kendra-ranking update-rescore-execution-plan \
  --id <rescore execution plan ID> \
  --capacity-units '{"RescoreCapacityUnits":2}'
```

Python

```
import boto3
from botocore.exceptions import ClientError
import pprint
import time

kendra_ranking = boto3.client("kendra-ranking")

print("Update a rescore execution plan.")

# Provide the ID of the rescore execution plan
id = <rescore execution plan ID>
# Re-set your required additional capacity units
capacity_units = 2

try:
    kendra_ranking.update_rescore_execution_plan(
        Id = id,
        CapacityUnits = {"RescoreCapacityUnits":capacity_units}
    )

    print("Wait for Amazon Kendra to update the rescore execution plan.")

    while True:
        # Get the details of the rescore execution plan, such as the status
        rescore_execution_plan_description =
 kendra_ranking.describe_rescore_execution_plan(
            Id = id
        )
        # When status is not UPDATING quit.
        status = rescore_execution_plan_description["Status"]
        print(" Updating rescore execution plan. Status: "+status)
        time.sleep(60)
        if status != "UPDATING":
            break

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

```
import java.util.concurrent.TimeUnit;

import software.amazon.awssdk.services.kendraranking.KendraRankingClient;
import
 software.amazon.awssdk.services.kendraranking.model.CapacityUnitsConfiguration;
import
 software.amazon.awssdk.services.kendraranking.model.DescribeRescoreExecutionPlanRequest;
import
 software.amazon.awssdk.services.kendraranking.model.DescribeRescoreExecutionPlanResponse;
import
 software.amazon.awssdk.services.kendraranking.model.RescoreExecutionPlanStatus;
import
 software.amazon.awssdk.services.kendraranking.model.UpdateRescoreExecutionPlanRequest;
import
 software.amazon.awssdk.services.kendraranking.model.UpdateRescoreExecutionPlanResponse;

public class UpdateRescoreExecutionPlanExample {

  public static void main(String[] args) throws InterruptedException {

    String rescoreExecutionPlanId = <rescore execution plan ID>;
    int newCapacityUnits = 2;

    KendraRankingClient kendraRankingClient = KendraRankingClient.builder().build();

    System.out.println(String.format("Updating a rescore execution plan named %s",
 rescoreExecutionPlanId));

    UpdateRescoreExecutionPlanResponse updateResponse =
 kendraRankingClient.updateRescoreExecutionPlan(
        UpdateRescoreExecutionPlanRequest.builder()
            .id(rescoreExecutionPlanId)
            .capacityUnits(
                CapacityUnitsConfiguration.builder()
                    .rescoreCapacityUnits(newCapacityUnits)
                    .build()
            )
            .build()
    );

    System.out.println(String.format("Waiting for rescore execution plan with id %s
 to finish updating.", rescoreExecutionPlanId));
```

```
    while (true) {
      DescribeRescoreExecutionPlanResponse describeResponse =
 kendraRankingClient.describeRescoreExecutionPlan(
          DescribeRescoreExecutionPlanRequest.builder()
              .id(rescoreExecutionPlanId)
              .build()
      );
      RescoreExecutionPlanStatus rescoreExecutionPlanStatus =
 describeResponse.status();
      if (rescoreExecutionPlanStatus != RescoreExecutionPlanStatus.UPDATING) {
        break;
      }
      TimeUnit.SECONDS.sleep(60);
    }

    System.out.println("Rescore execution plan update is complete.");
  }
}
```

下面是使用 Rescore API 的示例。

CLI

```
aws kendra-ranking rescore \
  --rescore-execution-plan-id <rescore execution plan ID> \
  --search-query "intelligent systems" \
  --documents "[{\"Id\": \"DocId1\",\"Title\": \"Smart systems\", \"Body\":
 \"intelligent systems in everyday life\",\"OriginalScore\": 2.0}, {\"Id\":
 \"DocId2\",\"Title\": \"Smarter systems\", \"Body\": \"living with intelligent
 systems\",\"OriginalScore\": 1.0}]"
```

Python

```
import boto3
from botocore.exceptions import ClientError
import pprint

kendra_ranking = boto3.client("kendra-ranking")

print("Use the Rescore API.")

# Provide the ID of the rescore execution plan
```

```
id = <rescore execution plan ID>
# The search query from the search service
query = "intelligent systems"
# The list of documents for Intelligent Ranking to rescore
document_list = [
    {"Id": "DocId1", "Title": "Smart systems", "Body": "intelligent systems in
 everyday life", "OriginalScore": 2.0},
    {"Id": "DocId2", "Title": "Smarter systems", "Body": "living with intelligent
 systems", "OriginalScore": 1.0}
]

try:
    rescore_response = kendra_ranking.rescore(
        rescore_execution_plan_id = id,
        search_query = query,
        documents = document_list
    )

    print(rescore_response["RescoreId"])
    print(rescore_resposne["ResultItems"])

except ClientError as e:
        print("%s" % e)

print("Program ends.")
```

Java

```
import java.util.ArrayList;
import java.util.List;

import software.amazon.awssdk.services.kendraranking.KendraRankingClient;
import software.amazon.awssdk.services.kendraranking.model.RescoreRequest;
import software.amazon.awssdk.services.kendraranking.model.RescoreResponse;
import software.amazon.awssdk.services.kendraranking.model.Document;

public class RescoreExample {

  public static void main(String[] args) {

    String rescoreExecutionPlanId = <rescore execution plan ID>;
    String query = "intelligent systems";
```

```
    List<Document> documentList = new ArrayList<>();
    documentList.add(
        Document.builder()
            .id("DocId1")
            .originalScore(2.0F)
            .body("intelligent systems in everyday life")
            .title("Smart systems")
            .build()
    );
    documentList.add(
        Document.builder()
            .id("DocId2")
            .originalScore(1.0F)
            .body("living with intelligent systems")
            .title("Smarter systems")
            .build()
    );

    KendraRankingClient kendraRankingClient = KendraRankingClient.builder().build();

    RescoreResponse rescoreResponse = kendraRankingClient.rescore(
        RescoreRequest.builder()
            .rescoreExecutionPlanId(rescoreExecutionPlanId)
            .searchQuery(query)
            .documents(documentList)
            .build()
    );

    System.out.println(rescoreResponse.rescoreId());
    System.out.println(rescoreResponse.resultItems());
  }
}
```

的文档历史记录 Amazon Kendra

- 最新文档更新：2024 年 2 月 27 日

下表描述了每个版本中的重要更改 Amazon Kendra。如需对此文档更新的通知，您可以订阅 RSS 源。

变更	说明	日期
新特征	Amazon Kendra 现在支持 GitHub 数据源连接器的更新版本。有关更多信息，请参阅GitHub。	2024年2月27日
新特征	Amazon Kendra 现在支持 Amazon FSx 数据源连接器的更新版本。有关更多信息，请参阅 Amazon FSx (Windows) 和 Amazon FSx (NetAppONTAP)。	2024年2月8日
新特征	Amazon Kendra 现在支持 Slack 数据源连接器的更新版本。有关更多信息，请参阅 Slack。	2024 年 1 月 11 日
新特征	Amazon Kendra 现在支持折叠和展开搜索结果。有关更多信息，请参阅折叠/展开搜索结果。	2023 年 10 月 19 日
新特征	Amazon Kendra 现在支持 Aurora (MySQL) 数据源连接器。有关更多信息，请参阅 Aurora (MySQL)。	2023 年 9 月 28 日

新特征	Amazon Kendra 现在支持 Aurora (PostgreSQL) 数据源连接器。有关更多信息，请参阅 Aurora (PostgreSQL)。	2023 年 9 月 28 日
新特征	Amazon Kendra 现在支持 Amazon RDS (MySQL) 数据源连接器。有关更多信息，请参阅 Amazon RDS (MySQL)。	2023 年 9 月 28 日
新特征	Amazon Kendra 现在支持 Amazon RDS （微软 SQL Server ）数据源连接器。有关更多信息，请参阅 Amazon RDS (Microsoft SQL Server)。	2023 年 9 月 28 日
新特征	Amazon Kendra 现在支持 Amazon RDS (Oracle) 数据源连接器。有关更多信息，请参阅 Amazon RDS (Oracle)。	2023 年 9 月 28 日
新特征	Amazon Kendra 现在支持 Amazon RDS (PostgreSQL) 数据源连接器。有关更多信息，请参阅 Amazon RDS (PostgreSQL)。	2023 年 9 月 28 日
新特征	Amazon Kendra 现在支持 IBM DB2 数据源连接器。有关更多信息，请参阅 IBM DB2。	2023 年 9 月 28 日
新特征	Amazon Kendra 现在支持微软 SQL Server 数据源连接器。有关更多信息，请参阅 Microsoft SQL Server。	2023 年 9 月 28 日

新特征	Amazon Kendra 现在支持 MySQL 数据源连接器。有关更多信息，请参阅 MySQL。	2023 年 9 月 28 日
新特征	Amazon Kendra 现在支持 Oracle 数据库数据源连接器。有关更多信息，请参阅 Oracle Database。	2023 年 9 月 28 日
新特征	Amazon Kendra 现在支持 PostgreSQL 数据源连接器。有关更多信息，请参阅 PostgreSQL。	2023 年 9 月 28 日
新特征	Amazon Kendra 现在为 Drupal 提供了一个数据源连接器。有关更多信息，请参阅 Drupal。	2023 年 9 月 6 日
新特征	使用检索增强生成 (RAG) 系统的 Amazon Kendra Retrieve API 检索语义相关的段落。	2023 年 6 月 22 日
新特征	Amazon Kendra 现在支持 Amazon Kendra Web Crawler 数据源连接器的更新版本。有关更多信息，请参阅 Amazon Kendra Web 爬网程序 v2.0。	2023 年 6 月 21 日
区域扩展	Amazon Kendra 现已在欧洲（伦敦）（eu-west-2）上市。	2023 年 6 月 5 日
新特征	Amazon Kendra 现在支持 Alfresco 数据源连接器的更新版本。有关更多信息，请参阅 Alfresco。	2023 年 5 月 16 日

新特征	Amazon Kendra 现在为 Adobe Experience Manager 提供了数据源连接器。有关更多信息，请参阅 Adobe Experience Manager。	2023 年 5 月 11 日
新特征	Amazon Kendra 现在支持在调用时配置文档字段/属性。GetQuerySuggestions现在，您可以根据文档字段的内容提出查询建议。有关更多信息，请参阅查询建议。	2023 年 5 月 2 日
新特征	Amazon Kendra 现在为 Gmail 提供了数据源连接器。有关更多信息，请参阅 Gmail。	2023 年 4 月 13 日
新特征	Amazon Kendra 现在支持 Microsoft OneDrive 数据源连接器的更新版本。有关更多信息，请参阅微软 OneDrive v2.0。	2023 年 4 月 3 日
新特征	当您的用户使用精选结果键入某些查询时，可以提高新文档的可见性或推广某些文档。	2023 年 3 月 30 日
新特征	Amazon Kendra 现在支持适用于 Microsoft 的更新数据源连接器 SharePoint。有关更多信息，请参阅微软 SharePoint。	2023 年 3 月 2 日
新特征	Amazon Kendra 现在支持 Confluence 数据源连接器的更新版本。有关更多信息，请参阅 Confluence。	2023 年 3 月 1 日

区域扩展	Amazon Kendra 现已在亚太地区（东京）（ap-northeast-1）上市。	2023 年 2 月 7 日
新特征	Amazon Kendra 现在为微软 Exchange 提供了数据源连接器。有关更多信息，请参阅 Microsoft Exchange。	2023 年 1 月 12 日
新特征	Amazon Kendra 现在为微软 Yammer 提供了数据源连接器。有关更多信息，请参阅 Microsoft Yammer。	2023 年 1 月 12 日
新特征	Amazon Kendra 现在支持索引 RTF、XML、XSLT、MS_EXCEL、CSV、JSON 和 MD 文档类型。有关更多信息，请参阅文档类型。	2023 年 1 月 11 日
新特征	Amazon Kendra 现在支持 Amazon S3 数据源连接器的更新版本。有关更多信息，请参阅Amazon S3。	2023 年 1 月 10 日
新特征	OpenSearch（自我管理）搜索结果可以使用Amazon Kendra 智能排名进行语义排名。	2023 年 1 月 9 日
新特征	Amazon Kendra 现在为 Microsoft Teams 提供了数据源连接器。有关更多信息，请参阅 Microsoft Teams。	2023 年 1 月 5 日
新特征	Amazon Kendra 更新了 Google 云端硬盘的数据源连接器。有关更多信息，请参阅 Google Drive。	2023 年 1 月 5 日

新特征	Amazon Kendra 有一个更新的数据源连接器 ServiceNow。有关更多信息，请参阅ServiceNow。	2022 年 12 月 21 日
新特征	Amazon Kendra 更新了 Salesforce 的数据源连接器。有关更多信息，请参阅 Salesforce。	2022 年 12 月 21 日
区域扩展	Amazon Kendra 现已在亚太地区（孟买）（ap-south-1）上市。	2022 年 12 月 14 日
新特征	Amazon Kendra 的表格搜索功能可以从嵌入在 HTML 文档中的表格搜索和提取答案。	2022 年 11 月 27 日
新特征	Amazon Kendra 支持对一组选定语言进行语义搜索。	2022 年 11 月 27 日
新特征	Amazon Kendra 现在为 Dropbox 提供了数据源连接器。有关更多信息，请参阅 Dropbox。	2022 年 9 月 27 日
新特征	Amazon Kendra 现在为 Zendesk 提供了数据源连接器。有关更多信息，请参阅 Zendesk。	2022 年 8 月 17 日
新特征	在为文档编制索引后，现在可以重新配置文档级别的访问控制。有关更多信息，请参阅访问控制配置。	2022 年 7 月 14 日

新特征	Amazon Kendra 现在为 Alfresco 提供了一个数据源连接器。有关更多信息，请参阅 Alfresco。	2022 年 6 月 30 日
新特征	Amazon Kendra 现在为提供了一个数据源连接器 GitHub。有关更多信息，请参阅GitHub。	2022 年 6 月 2 日
新特征	Amazon Kendra 现在为 Jira 提供了数据源连接器。有关更多信息，请参阅 Jira。	2022 年 5 月 12 日
新特征	分面内的嵌套分面可以显示在搜索结果中。有关更多信息，请参阅分面。	2022 年 5 月 5 日
新特征	Amazon Kendra 现在为 Quip 提供了数据源连接器。有关更多信息，请参阅 Quip。	2022 年 4 月 19 日
新特征	Amazon Kendra 现在为 Box 提供了数据源连接器。有关更多信息，请参阅 Box。	2022 年 4 月 6 日
新特征	Amazon Kendra 现在为 Slack 提供了数据源连接器。有关更多信息，请参阅 Slack。	2022 年 3 月 14 日
新特征	Amazon Kendra 现在为提供了一个数据源连接器 Amazon FSx。有关更多信息，请参阅Amazon FSx。	2022 年 2 月 8 日
AWS 托管策略更新-新策略	Amazon Kendra 添加了新的 AWS 托管策略。有关更多信息，请参阅适用于 Amazon Kendra 的 AWS 托管策略。	2022 年 1 月 3 日

新特征	Amazon Kendra 只需点击几下即可部署搜索应用程序，无需任何前端代码。有关更多信息，请参阅部署无代码的搜索应用程序。	2021 年 12 月 1 日
新特征	您可以在文档提取过程中丰富文档元数据和内容。有关更多信息，请参阅在提取过程中自定义文档元数据。	2021 年 12 月 1 日
新特征	Amazon Kendra 提供搜索分析，以获取有关搜索应用程序的有用见解。有关更多信息，请参阅使用搜索分析获取见解。	2021 年 12 月 1 日
区域扩展	Amazon Kendra 现已在 AWS GovCloud （美国西部）(-us-gov-west 1) 中推出。	2021 年 10 月 13 日
新特征	Amazon Kendra 现在可以索引多种语言的文档并按语言筛选搜索结果。请参阅添加非英语语言的文档和搜索语言。	2021 年 10 月 7 日
新特征	Amazon Kendra 现在与 Identity Center 目录集成，可以获取群组和用户的访问级别，以进行用户上下文筛选。请参阅 IAM Identity Center 的用户组配置。	2021 年 10 月 6 日
新教程	Amazon Kendra 现在提供了一个教程，指导你如何构建富含元数据的搜索解决方案。请参阅构建智能搜索解决方案。	2021 年 8 月 13 日

新特征	Amazon Kendra 现在为提供了一个数据源连接器 Amazon WorkDocs。有关更多信息，请参阅Amazon WorkDocs。	2021 年 7 月 20 日
新特征	Amazon Kendra 现在提供了一个 Web 爬虫来抓取和索引网页。有关更多信息，请参阅 Web 爬网程序。	2021 年 6 月 17 日
区域扩展	Amazon Kendra 现已在加拿大（中部）（ca-central-1）上市。	2021 年 6 月 16 日
区域扩展	Amazon Kendra 现已在美国东部（俄亥俄州）（us-east-2）上市。	2021 年 6 月 7 日
新特征	Amazon Kendra 现在支持查询建议，向用户推荐与其搜索相关的热门查询。有关更多信息，请参阅推荐热门搜索查询。	2021 年 5 月 27 日
AWS 托管策略更新-新策略	Amazon Kendra 添加了新的 AWS 托管策略。有关更多信息，请参阅适用于 Amazon Kendra 的 AWS 托管策略。	2021 年 5 月 27 日
区域扩展	Amazon Kendra 现已在亚太地区（新加坡）（ap-southeast-1）上市。	2021 年 5 月 5 日
新特征	Amazon Kendra 现在支持通过覆盖在索引级别设置的调整配置来调整查询中的搜索相关性。有关更多信息，请参阅调整搜索相关性和调整响应。	2021 年 4 月 20 日

新特征	Amazon Kendra 现在支持 OAuth 2.0 身份验证和使用 ServiceNow查询来选择要编制索引的文档。有关更多信息，请参阅ServiceNow。	2021 年 4 月 1 日
新特征	Amazon Kendra 现在支持对常见问题解答文档进行增量学习。有关更多信息，请参阅为增量学习提交反馈。	2021 年 2 月 17 日
新特征	Amazon Kendra 现在支持索引同义词。有关更多信息，请参阅向索引添加同义词。	2020 年 12 月 10 日
新特征	Amazon Kendra 现在为 Google Workspace 云端硬盘提供了数据库连接器。有关更多信息，请参阅使用 Google Workspace Drive 数据来源。	2020 年 12 月 8 日
新特征	Amazon Kendra 现在提供了一个 JavaScript 库，使您可以更轻松地向其提供查询反馈 Amazon Kendra。有关更多信息，请参阅提交反馈。	2020 年 12 月 8 日
新特征	Amazon Kendra 现在支持基于令牌的用户访问控制。有关更多信息，请参阅控制对索引中文档的访问权限。	2020 年 11 月 5 日
新特征	Amazon Kendra Confluence 数据源连接器现在可以与 Confluence 云配合使用。有关更多信息，请参阅使用 Confluence 数据来源。	2020 年 11 月 5 日

区域扩展	Amazon Kendra 现已在亚太地区（悉尼）（ap-southeast-2）上市。	2020 年 11 月 2 日
新特征	Amazon Kendra 现在为 Confluence 服务器提供了数据源连接器。有关更多信息，请参阅使用 Confluence 数据来源。	2020 年 10 月 26 日
新特征	Amazon Kendra 现在提供了一个数据源，您可以使用该数据源为自定义连接器生成统计信息。有关更多信息，请参阅使用自定义数据来源。	2020 年 10 月 21 日
新特征	Amazon Kendra 现在支持常见问题的自定义属性。有关更多信息，请参阅添加问题和答案。	2020 年 9 月 17 日
新特征	Amazon Kendra 现在返回查询结果的置信度分数。有关更多信息，请参阅QueryResultItem。	2020 年 9 月 15 日
新特征	AWS CloudFormation 现在支持 Amazon Kendra。有关更多信息，请参阅Amazon Kendra 资源类型参考- AWS CloudFormation。	2020 年 9 月 10 日
新特征	Amazon Kendra 添加了对的支持 AWS PrivateLink。有关更多信息，请参阅 Amazon Kendra 和接口 VPC 终端节点 (AWS PrivateLink)。	2020 年 7 月 7 日

新指南	这是 Amazon Kendra 开发人员指南的首次发布。	2020 年 5 月 11 日

API 参考

API 参考文档现在是一个单独的指南。

AWS 术语表

有关最新的 AWS 术语，请参阅《AWS 词汇表参考》中的 AWS 词汇表。

本文属于机器翻译版本。若本译文内容与英语原文存在差异，则一律以英文原文为准。