



韧性分析框架

AWS 规范性指导



AWS 规范性指导: 韧性分析框架

Copyright © 2024 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon 的商标和商业外观不得用于任何非 Amazon 的商品或服务，也不得以任何可能引起客户混淆、贬低或诋毁 Amazon 的方式使用。所有非 Amazon 拥有的其他商标均为各自所有者的财产，这些所有者可能附属于 Amazon、与 Amazon 有关联或由 Amazon 赞助，也可能不是如此。

Table of Contents

介绍	1
框架概述	2
了解工作量	5
应用框架	7
缓解潜在故障	9
了解权衡取舍和风险	9
故障模式可观察性	11
常见的缓解策略	11
持续改进	15
结论和资源	16
文档历史记录	17
术语表	18
#	18
A	18
B	21
C	22
D	25
E	28
F	30
G	31
H	32
I	33
L	35
M	36
O	40
P	42
Q	44
R	45
S	47
T	50
U	51
V	52
W	52
Z	53

..... liv

弹性分析框架

约翰·福门托、布鲁诺·埃默尔、史蒂芬·胡珀、杰森·巴托和亚马逊网络服务的迈克尔·哈肯 (AWS)

2023 年 9 月([文档历史记录](#))

一致、可重复的标准和流程是持续改进的重要组成部分。分布式系统的弹性也是如此。本指南的目的是引入一个弹性分析框架，该框架提供了一种一致的方法来分析故障模式以及它们如何影响您的工作负载。在工作负载的整个生命周期（从设计到操作）中使用此框架，可以帮助您以一致且可重复的方式持续提高工作负载在更广泛的潜在故障模式下的弹性。这有助于确保您实现弹性目标，并保持工作负载所需的弹性特性。

该框架是基于 AWS 解决方案架构现场团队与各行各业客户合作的经验开发的。它针对的是可以拥有多种职称的建筑商，包括产品经理、软件开发人员、系统工程师、运营团队和架构师。这些人最了解正在分析的系统、服务或产品。在持续练习中使用该框架可以帮助您取得渐进的进步并实现长期的复原力目标。

该框架的重点是识别潜在的故障模式以及可以用来减轻其影响的预防性和纠正性控制措施。即使故障发生在您无法直接控制的组件中，例如依赖项中的错误率增加，您也需要考虑这些故障会如何影响您的工作负载，以及如何设计工作负载来应对这些故障。归根结底，你应该专注于您可以应对的故障通过使用由您控制的缓解措施。

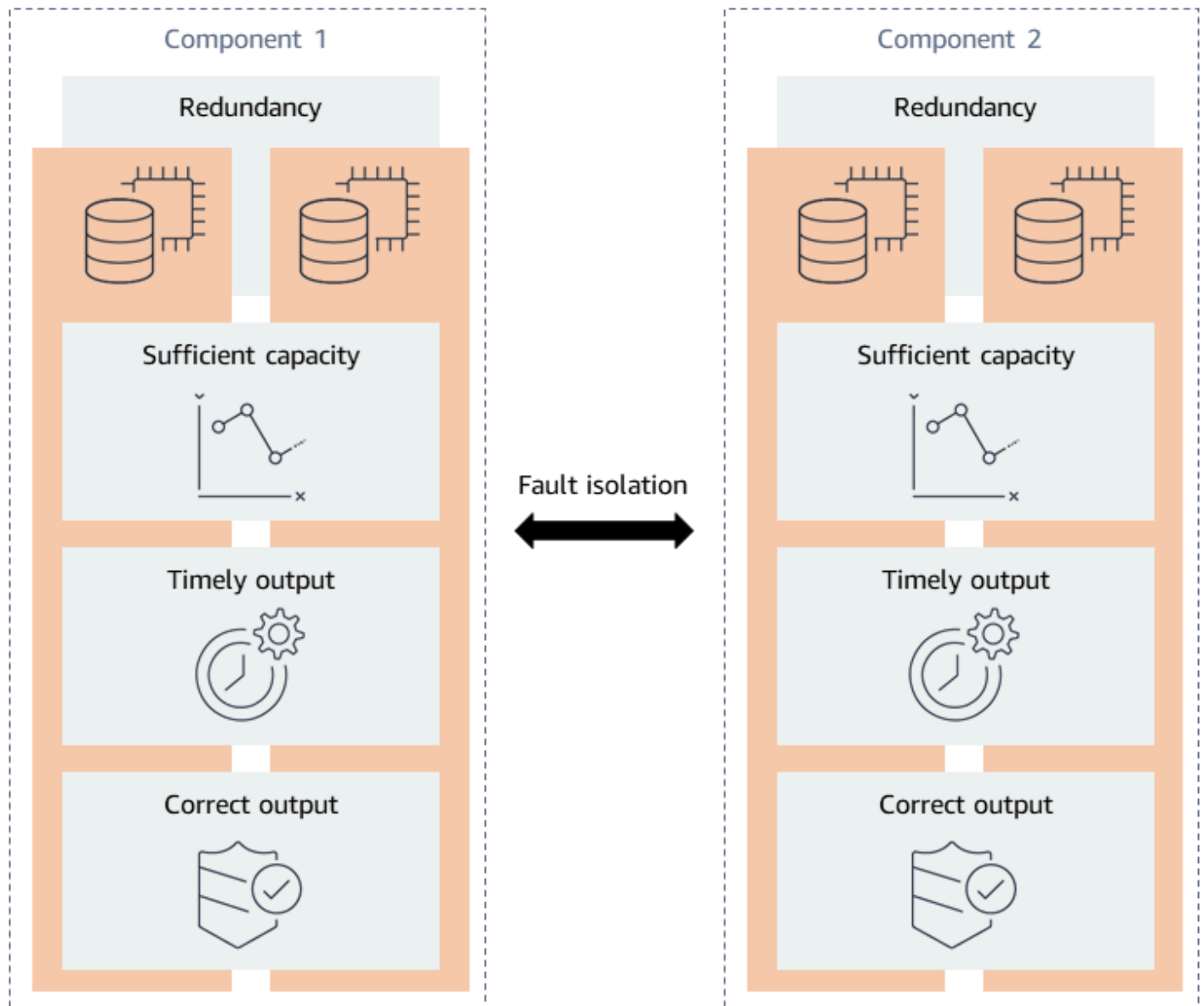
本指南概述了框架，然后讨论了如何识别和记录工作负载、如何将框架应用于该工作负载，以及如何针对发现的任何潜在故障评估缓解策略。

内容

- [框架概述](#)
- [了解工作量](#)
- [应用框架](#)
- [缓解潜在故障](#)
- [结论和资源](#)

框架概述

弹性分析框架是通过确定工作负载所需的弹性特性而开发的。所需的属性是你想要对系统保持真实的东西。弹性通常通过可用性来衡量，因此高可用性分布式系统的五个特征是：冗余、足够的容量、及时的输出、正确的输出和故障隔离。这些属性如下图所示。



- 冗余—容错是通过消除单点故障 (SPOF) 的冗余实现的。冗余可以从工作负载中的备用组件到整个应用程序堆栈的完整副本。在考虑应用程序的冗余时，重要的是要考虑所使用的基础架构、数据存储和依赖项提供的冗余级别。例如，Amazon DynamoDB 和亚马逊简单存储服务 (Amazon S3) 通过跨

区域的多个可用区复制数据来提供冗余，以及AWS Lambda在多个可用区的多个工作节点上运行您的函数。对于您使用的每项服务，请考虑该服务提供的内容以及您需要设计的内容。

- **足够的容量**— 您的工作负载需要足够的资源才能按预期运行。资源包括内存、CPU 周期、线程、存储、吞吐量、服务配额等。
- **及时输出**— 当客户使用您的工作负载时，他们希望它能在合理的时间内执行其预期功能。除非服务提供延迟服务级别协议 (SLA)，否则他们的期望通常基于经验证据，即他们自己的经验。这个客户体验一般通常被视为系统中延迟的中位数 (P50)。如果您的工作负载花费的时间超过预期，则这种延迟可能会影响您的客户体验。
- **正确输出**— 需要工作负载软件的正确输出才能提供预期功能。不正确或不完整的结果可能比根本没有回应更糟糕。
- **故障隔离**— 故障隔离将故障发生时的影响范围限制在预期的故障容器内。它可以确保工作负载的特定组件一起出现故障，同时防止故障级联到其他意想不到的组件。它还有助于限制您的工作负载对客户的影响范围。故障隔离与前四个属性略有不同，因为它接受已经发生故障，但应加以控制。您可以在基础架构、依赖项和软件功能中创建故障隔离。

当违反所需属性时，可能会导致工作负载不可用或被认为不可用。基于这些所需的弹性特性以及我们与许多人合作的经验AWS客户，我们已经确定了五种常见的故障类别：单点故障、负载过大、延迟过长、配置错误和错误以及共同命运，我们将其缩写为 SEES。它们为对潜在的失效模式进行分类提供了一种一致的方法，下表对此进行了介绍。

失败类别	违反	定义
单点故障 (SPOF)	冗余	由于组件缺乏冗余，单个组件出现故障会导致系统中断。
负荷过大	足够的容量	由于需求过大或流量过度消耗资源会使资源无法执行其预期功能。这可能包括达到限制和配额，这会导致请求受到限制和拒绝。
延迟过长	及时输出	系统处理或网络流量延迟超过预期时间、服务级别目标 (SLO) 或服务级别协议 (SLA)。

配置错误和错误

正确输出

软件错误或系统配置错误会导致输出不正确。

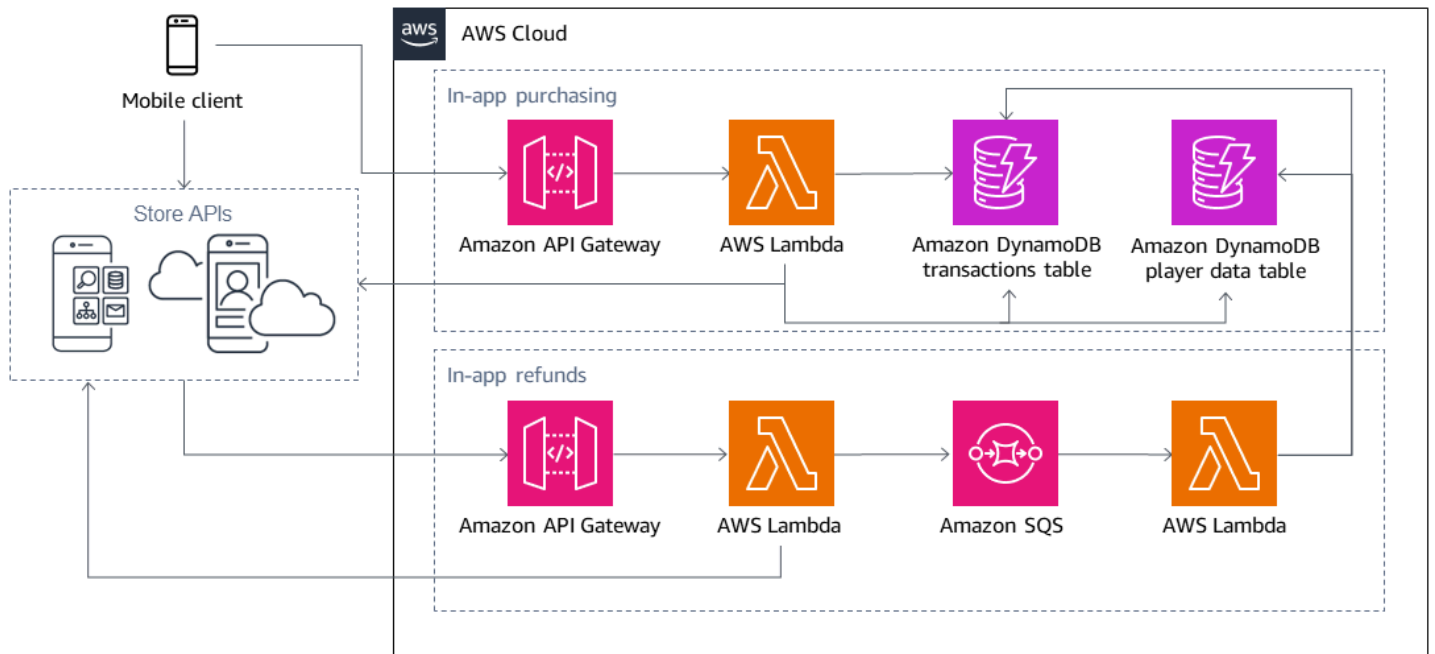
共同的命运

故障隔离

由先前的任何故障类别引起的故障跨越了预期的故障隔离界限，并级联到系统的其他部分或其他客户。

了解工作量

要应用该框架，首先要了解要分析的工作负载。系统架构图为记录系统最相关的细节提供了一个起点。但是，尝试分析整个工作负载可能很复杂，因为许多系统都有许多组件和交互作用。相反，我们建议您集中精力[用户故事](#)，它们是从最终用户的角度对软件功能进行非正式、一般性的解释。他们的目的是阐明软件功能如何为客户提供价值。然后，您可以使用架构图和数据流程图对这些用户故事进行建模，以便更轻松地评估提供所述业务功能的技术组件。例如，应用内手机游戏购买解决方案可能包含两个用户故事，“购买应用内积分”和“获得应用内退款”，如下图所示。（此示例架构重点介绍了如何将系统分解为用户故事；它并不打算代表具有高弹性的应用程序。）



每个用户故事都包含四个常见组件：代码和配置、基础架构、数据存储和外部依赖关系。您的图表应包括所有这些组件，并反映各组件之间的相互作用。例如，如果您的 Amazon API Gateway 终端节点负载过大，请考虑该负载如何级联到系统中的其他组件，例如您的 AWS Lambda 函数或亚马逊 DynamoDB 表。跟踪这些互动有助于你了解故障模式会如何影响用户故事。如上图所示，您可以使用数据流图或在架构图中使用简单的流程箭头直观地捕获此流。对于每个组件，请考虑捕获详细信息，例如正在传输的信息类型、接收的信息、通信是同步还是异步通信以及跨越了哪些故障边界。在示例中，DynamoDB 表在两个用户故事中共享，您可以从箭头中看到，该箭头表示应用内退款故事中的 Lambda 组件访问了应用内购买故事中的 DynamoDB 表。这意味着，由于共同的命运，由应用内购买用户故事引起的失败可能会级联到应用内退款用户故事。

此外，了解每个组件的基准配置也很重要。基准配置确定了约束条件，例如每秒的平均和最大事务数、有效负载的最大大小、客户端超时以及资源的默认或当前服务配额。如果您要对新设计进行建模，我们建议您记录设计的功能要求并考虑限制。这可以帮助你了解故障模式在组件中是如何表现出来的。

最后，你应该根据用户故事提供的商业价值来确定其优先级。这种优先级排序可帮助您首先专注于工作负载中最关键的功能。然后，您可以将分析重点放在作为该功能关键路径一部分的工作负载组件上，并更快地利用该框架实现价值。在迭代该过程时，您可以按不同的优先级检查其他用户故事。

应用框架

应用弹性分析框架的最佳方法是从一组按失败类别组织的标准问题开始，你应该询问你正在分析的用户故事中的每个组成部分。如果有些问题不适用于工作负载中的每个组件，请使用最适用的问题。

你可以从两个角度思考故障模式：

- 故障如何影响组件支持用户故事的能力？
- 故障如何影响组件与其他组件的交互？

例如，当你考虑数据存储和过度负载时，你可能会考虑数据库负载过大而查询超时的故障模式。您可能还会想一想，您的数据库客户端会如何通过重试使数据库不堪重负或无法关闭数据库连接，从而耗尽连接池。另一个例子是身份验证过程，它可能包括几个步骤。您需要仔细考虑多因素身份验证 (MFA) 应用程序或第三方身份提供商 (IdP) 的故障会如何影响此身份验证系统中的用户情景。

在回答以下问题时，应考虑故障的根源。例如，过载是由客户激增引起的，还是由人工操作员在维护活动期间停止使用太多节点造成的？你也许能够在每个问题中找出多种失败来源，这可能需要不同的缓解措施。在提问时，请记录下您发现的潜在故障模式、它们适用于哪些组件以及每个故障的来源。

单点故障

- 该组件是否专为冗余而设计？
- 如果组件出现故障会怎样？
- 您的应用程序能否容忍单个可用区的部分或全部损失？

延迟过长

- 如果此组件的延迟时间增加，或者与之交互的组件延迟增加（或网络中断，例如 TCP 重置），会发生什么？
- 您是否使用重试策略对超时进行了适当的配置？
- 你失败的速度是快还是慢？是否存在级联效应，例如由于受损资源快速失败而无意中将所有流量发送到受损资源？
- 对这个组件提出的最昂贵的请求是什么？

负荷过大

- 什么能压倒这个组件？这个组件怎么能压倒其他组件？
- 如何防止将资源浪费在永远不会成功的工作上？
- 您是否有为该组件配置的断路器？
- 有什么东西会造成无法克服的待办事项吗？
- 这个组件在哪里可以体验双模态行为？
- 可以超过哪些限制或服务配额（包括存储容量）？
- 组件在负载下如何缩放？

配置错误和错误

- 如何防止将错误配置和错误部署到生产环境中？
- 您能否自动回滚错误的部署或将流量从部署更新或更改的故障容器中移开？
- 您有哪些护栏可以防止操作员失误？
- 哪些项目（例如证书或证书）可以过期？

共同的命运

- 您的故障隔离界限是多少？
- 对部署单位所做的更改是否至少与您的预期一样小 [故障隔离边界](#) 但理想情况下更小，例如单机环境（故障隔离边界内的单个实例）？
- 此组件是否在用户故事或其他工作负载之间共享？
- 还有哪些组件与该组件紧密耦合？
- 如果此组件或其依赖项出现部分故障或灰色故障，会发生什么？

在问完这些问题之后，您还可以使用 SEES 来提出针对你的工作负载和每个组件的其他问题。SEES 最适合用作思考故障模式的结构化方式，也可用作执行弹性分析时的灵感来源。它不是一个严格的分类法。不要花时间担心特定的故障模式属于哪个类别，这并不重要。什么是重要的是你想到了失败然后把它写下来。没有错误的答案；发挥创造力和跳出框框思考是有益的。此外，不要假设故障模式已经得到缓解；请包括你能想到的所有潜在故障模式。

在第一次练习中，您不太可能预见到所有潜在的故障模式。框架的多次迭代可以帮助你生成更完整的模型，因此你不必在第一次通过时就尝试解决所有问题。您可以按定期、每周或每两周的节奏进行分析。在每个会话中，重点关注特定的故障模式或组件。这有助于在提高工作负载弹性方面取得稳定、渐进的进展。收集用户情景的潜在故障模式列表后，您可以决定如何处理这些模式。

缓解潜在故障

现在，用户情景中的组件可能出现故障，您可以专注于缓解措施。首先，根据你发现的每一次失败的潜在影响和可能性，回顾一下潜在的权衡取舍。然后确定所需的可观察性级别并选择缓解策略。权衡应包括努力制定适当水平的可观测性和缓解策略。最后，确定正确的节奏来定期进行弹性分析审查。

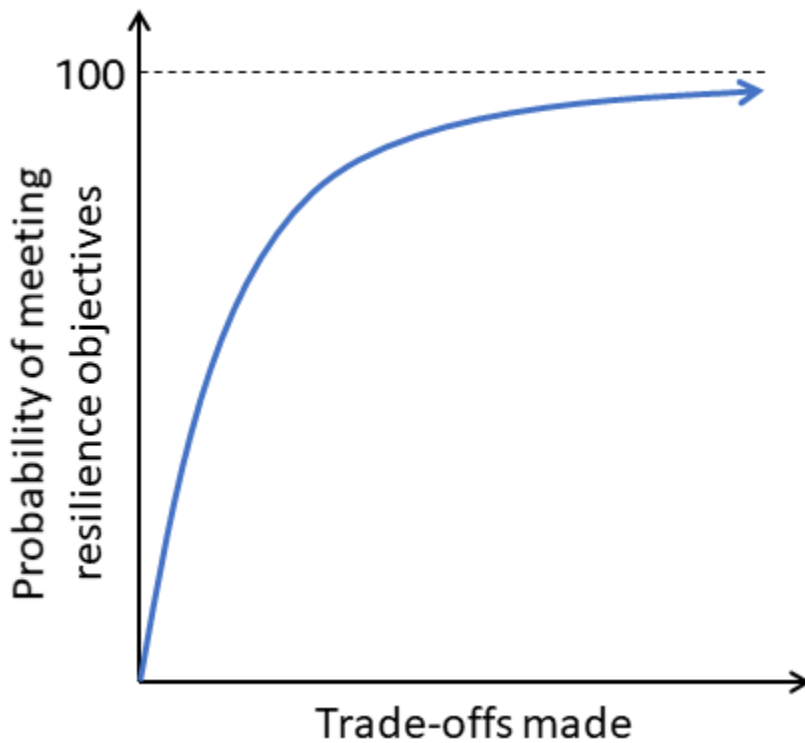
Sections

- [了解权衡取舍和风险](#)
- [故障模式可观察性](#)
- [常见的缓解策略](#)
- [持续改进](#)

了解权衡取舍和风险

弹性架构应使用少量经过充分测试、简单且可靠的机制来应对故障。为了实现最高水平的弹性，工作负载应自动检测尽可能多的故障模式并从中恢复。要做到这一点，就需要在进行弹性分析方面进行大量投资。这意味着，实现更高的弹性水平需要权衡取舍。但是，当你继续权衡取舍时，你已经到了相对于弹性目标的回报递减的地步。以下是最典型的权衡取舍：

- 成本 — 冗余组件、增强的可观察性、额外的工具或更高的资源利用率将导致成本增加。
- 系统复杂性 — 检测和响应故障模式（包括缓解解决方案）以及可能不使用托管服务会增加系统的复杂性。
- 工程工作 — 需要额外的开发人员时间来构建用于检测和响应故障模式的解决方案。
- 运营开销 — 监控和操作能够处理更多故障模式的系统可能会增加运营开销，尤其是在您无法使用托管服务来缓解特定的故障模式时。
- 延迟和一致性 — 如 [PACEL C](#) 定理所述，构建有利于可用性的分布式系统需要在一致性和延迟之间进行权衡。



在考虑用户情景中已识别的故障模式的缓解措施时，请考虑需要做出的权衡取舍。与安全性一样，弹性也是一个优化问题。您必须决定是避免、缓解、转移还是接受已识别故障带来的风险。可能有一些故障模式你可以避免，一组你可以接受，还有一些你可以转移。您可以选择缓解所识别的许多故障模式。要确定要采取哪种方法，请通过问两个问题进行评估：失败发生的可能性有多大？如果确实发生，会对工作负载产生什么影响？

事件@@ 发生的可能性是多么合理。例如，如果用户情景中有一个在单个 Amazon Elastic Compute Cloud (Amazon EC2) 实例上运行的组件，则该组件可能会在系统运行期间的某个时刻中断，这可能是由于修补程序或操作系统错误所致。或者，由 Amazon Relational Database Service (Amazon RDS) 管理的在主实例和辅助实例之间同步数据的数据库完全不可用的可能性很低。

影响是对事件可能造成的伤害的估计。应从财务和声誉的角度对其进行评估，并与其影响的用户故事的价值相关。例如，数据库不堪重负可能会对电子商务系统接受新订单的能力产生重大影响。但是，在负载均衡器后面的 20 个实例队列中丢失一个实例可能影响不大。

您可以将这些问题的答案与为降低风险而必须进行权衡的成本进行比较。当你根据风险阈值和弹性目标考虑这些信息时，它会为你决定计划积极缓解哪些故障模式提供依据。

故障模式可观察性

要缓解故障模式，您首先必须检测它当前正在影响或即将影响您的工作负载。只有当有信号表明必须采取行动时，缓解措施才有效。这意味着，创建任何缓解措施的部分至少包括验证您是否拥有或正在构建检测故障影响所必需的可观察性。

您应该从两个维度考虑故障模式中可观察到的症状：

- 哪些主要指标可以告知您系统正在接近可能很快就会出现影响的状态？
- 有哪些滞后指标可以在故障模式发生后尽快显示其影响？

例如，应用于数据库元素的过度负载故障可以将连接计数作为领先指标。您可以将连接计数的稳步增加视为数据库可能很快就会超过连接限制的先行指标，因此您可以采取措施（例如终止最近最少使用的连接）来减少连接数。滞后指示器会指示何时超过数据库连接限制，何时出现数据库连接错误。除了收集应用程序和基础设施指标外，还可以考虑收集[关键绩效指标 \(KPI\)](#)，以检测故障何时会影响您的客户体验。

如果可能，我们建议您在可观测性策略中加入这两种指标。在某些情况下，您可能无法创建领先指标，但您应该始终计划为要缓解的每个失败设置一个滞后指标。要选择正确的缓解措施，还应考虑领先指标还是滞后指标检测到故障。例如，假设您的网站流量突然激增。你可能只会看到一个滞后指标。在这种情况下，仅靠自动扩展可能不是最好的缓解措施，因为部署新资源需要时间，而限制几乎可以立即防止过载，让您的应用程序有时间扩展或减少负载。相反，对于流量逐渐增加，你会看到一个领先的指标。在这种情况下，限制是不合适的，因为你有时间通过自动扩展系统来做出响应。

常见的缓解策略

首先，请考虑使用预防性缓解措施来防止故障模式影响用户情景。然后你应该考虑纠正性的缓解措施。纠正性缓解措施可帮助系统自我修复或适应不断变化的条件。以下是每个故障类别的常见缓解措施列表，这些缓解措施与弹性属性一致。

失败类别	所需的弹性特性	缓解措施
单点故障 (SPOF)	冗余和容错	<ul style="list-style-type: none">• 实现冗余——例如，通过在 Elastic Load Balancing (ELB) 后面使用多个 EC2 实例。

- 移除对[AWS全局服务控制平面](#)的依赖关系，只依赖于全局服务数据平面。
- 当资源不可用时，使用[优雅降级](#)，这样您的系统就可以静态稳定到单点故障。

负荷过大

足够的容量

- [关键的缓解策略包括速率限制、减负和工作优先级划分、持续工作、指数级退避和重试时抖动或根本不重试、控制较小的服务、管理队列深度、自动扩展、避免冷缓存和断路器。](#)
- 您还应该考虑自己的容量计划，并考虑未来可能达到的容量和扩展限制，这两者都与 AWS 资源和系统内的限制有关。

延迟过长

及时输出

- 实现适当配置的[超时](#)或自适应超时（根据当前和预测的延迟条件更改超时值，以允许慢速依赖项取得进展，而不是放弃缓慢的请求）。
- [通过抖动、对冲、在本地环境连接到云服务时使用多路径 TCP 等技术实现指数级退避和重试，在特定路由上遇到延迟，使用与松散耦合系统的异步交互、缓存和不浪费工作。](#)

配置错误和错误

正确输出

- 捕捉软件中可重复的功能错误的主要方法是通过[静态分析](#)、[单元测试](#)、[集成测试](#)、[回归测试](#)、[负载测试](#)和[弹性测试](#)等机制进行严格测试。
 -
- 实施[基础设施即代码 \(IaC\)](#)、[持续集成和持续交付 \(CI/CD\)](#) [自动化](#)等策略，以帮助缓解配置错误的威胁。
- 使用部署技术，例如[单箱](#)、[金丝雀部署](#)、与故障隔离边界一致的部分部署或[蓝/绿部署](#)，以减少错误配置和错误。

共同的命运

故障隔离

- 在您的系统中实现[容错能力](#)，并使用逻辑和物理故障隔离边界，例如多个计算或容器集群、多个 AWS 账户、多个 AWS Identity and Access Management (IAM) 委托人、多个可用区，可能还有多个 AWS 区域。
- [基于单元的架构](#)和[随机分片](#)等技术也可以改善故障隔离。
- 考虑[松散耦合](#)和[优雅退化等模式](#)，以防止级联故障。当你优先考虑用户故事时，您还可以使用该优先级来区分对主要业务功能至关重要的用户故事和可以优雅降级的用户故事。例如，在电子商务网站中，您不希望网站上的促销小工具受到损害，从而影响处理新订单的能力。

尽管其中一些缓解措施只需很少的精力即可实施，但其他缓解措施（例如采用基于单元的架构以实现可预测的故障隔离和最大限度地减少共同命运故障）可能需要重新设计整个工作负载，而不仅仅是特定用户故事的组成部分。如前所述，重要的是要权衡故障模式的可能性和影响，以及为缓解故障模式所做的权衡取舍。

除了适用于每种故障模式类别的缓解技术外，您还应该考虑恢复用户情景或整个系统所需的缓解措施。例如，故障可能会暂停工作流程并阻止将数据写入预定目的地。在这种情况下，您可能需要操作工具来重新推动工作流程或手动修复数据。您可能还需要在工作负载中构建检查点机制，以帮助防止在发生故障时丢失数据。或者，您可能需要建立一根 andon cord 来暂停工作流程并停止接受新工作以防止进一步的伤害。在这些情况下，您应该考虑所需的操作工具和护栏。

最后，你应该始终假设人类在制定缓解策略时会犯错误。尽管现代 DevOps 实践试图实现操作自动化，但由于各种原因，人类仍然必须与您的工作负载进行交互。不正确的人为操作可能会导致任何 SEES 类别的故障，例如在维护期间移除太多节点并导致过载，或者错误地设置了功能标志。这些情况

确实是预防性护栏的失败。根本原因分析永远不应该以“人类犯了错误”的结论而告终。相反，它应该首先解决可能出错的原因。因此，您的缓解策略应考虑人类操作员如何与工作负载组件进行交互，以及如何通过安全护栏防止或最大限度地减少人为操作员错误造成的影响。

持续改进

复原力是一个[持续的过程](#)。在系统的生命周期中，其运行环境将发生变化。为确保您的系统保持弹性，您应将该框架整合到定期的运营和架构审查中。你可能会发现新的故障模式是你第一次发现的，或者可能有新的或以前没有想过的缓解措施可以到位。复原力分析应该是一个反复的过程，而不是一次性的工作。

您应该通过[混沌工程](#)或[游戏日](#)等流程对缓解策略进行实证测试，以验证它们是否按预期运行。如果你没有严格的测试机制，你就无法确信缓解措施会在你需要时按预期发挥作用。在弹性分析期间，您可能会确定故障模式已经由特定的缓解措施处理，但测试这些假设也很重要。您应该测试现有的缓解措施和使用弹性分析框架创建的新缓解措施。

您还应该通过团队回顾来评估自己在分析方面的表现。在分析过程中，每个人都知道自己在做什么吗？你通过弹性分析发现的故障模式数量是否符合团队的预期？您能否找到针对您发现的所有故障模式的缓解措施？团队觉得这个过程有用吗？您认为它会提高工作负载的弹性吗？

当发生影响工作负载可用性的实际故障事件时，请记录特定的故障模式、故障中包含的组件以及使用的缓解模式。在事后分析工具中设置可搜索此元数据，这样您就可以确定将来要重点关注哪些故障模式和组件。在整个过程中，您可以与您的AWS客户团队和解决方案架构师接触。

结论和资源

本指南提供了一个以持续、一致的方式执行弹性分析的框架。此框架可帮助您确定单点故障、过度负载、延迟过长、配置错误和错误以及共同命运可能如何影响您的工作负载组件。识别这些故障模式有助于您确定适当的缓解策略，作为构建面向恢复的架构的一部分。

有关弹性分析的更多内容，请参阅以下链接：

- [弹性生命周期框架](#) (AWS规范性指南)
- [弹性解决方案](#) (AWS解决方案库)
- [迈向持续的韧性](#) (阿德里安·霍恩斯比，《云架构师》，2021年3月24日)

文档历史记录

下表描述了本指南的重大更改。如果您想收到有关未来更新的通知，可以订阅[RSS 提要](#)。

变更	说明	日期
首次出版	—	2023年9月5日

AWS 规范性指导词汇表

以下是 AWS 规范性指导提供的策略、指南和模式中的常用术语。若要推荐词条，请使用术语表末尾的提供反馈链接。

数字

7 R

将应用程序迁移到云中的 7 种常见迁移策略。这些策略以 Gartner 于 2011 年确定的 5 R 为基础，包括以下内容：

- **重构/重新架构** - 充分利用云原生功能来提高敏捷性、性能和可扩展性，以迁移应用程序并修改其架构。这通常涉及到移植操作系统和数据库。示例：将您的本地 Oracle 数据库迁移到 SQL 兼容 Amazon Aurora PostgreSQL 的版本。
- **更换平台**：将应用程序迁移到云中，并进行一定程度的优化，以利用云功能。示例：将您的本地 Oracle 数据库迁移到适用于 Oracle 的 Amazon Relational Database Service (Amazon RDS) AWS Cloud。
- **重新购买** - 转换到其他产品，通常是从传统许可转向 SaaS 模式。示例：将您的客户关系管理 (CRM) 系统迁移到 Salesforce.com。
- **更换主机 (直接迁移)** - 将应用程序迁移到云中，无需进行任何更改即可利用云功能。示例：在 EC2 实例上将您的本地 Oracle 数据库迁移到 Oracle AWS Cloud。
- **重新定位 (虚拟机监控器级直接迁移)**：将基础设施迁移到云中，无需购买新硬件、重写应用程序或修改现有操作。您可以将服务器从本地平台迁移到同一平台的云服务。示例：迁移 Microsoft Hyper-V 应用到 AWS。
- **保留 (重访)** - 将应用程序保留在源环境中。其中可能包括需要进行重大重构的应用程序，并且您希望将工作推迟到以后，以及您希望保留的遗留应用程序，因为迁移它们没有商业上的理由。
- **停用** - 停用或删除源环境中不再需要的应用程序。

A

ABAC

请参阅[基于属性的访问控制](#)。

抽象服务

参见[托管服务](#)。

ACID

参见[原子性、一致性、隔离性、耐久性](#)。

主动-主动迁移

一种数据库迁移方法，在这种方法中，源数据库和目标数据库保持同步（通过使用双向复制工具或双写操作），两个数据库都在迁移期间处理来自连接应用程序的事务。这种方法支持小批量、可控的迁移，而不需要一次性割接。与[主动-被动迁移](#)相比，它更灵活，但需要更多的工作。

主动-被动迁移

一种数据库迁移方法，在这种方法中，源数据库和目标数据库保持同步，但在将数据复制到目标数据库时，只有源数据库处理来自连接应用程序的事务。目标数据库在迁移期间不接受任何事务。

聚合函数

对一组行进行操作并计算该组的单个返回值的SQL函数。聚合函数的示例包括SUM和MAX。

AI

参见[人工智能](#)。

AIOps

参见[人工智能操作](#)。

匿名化

永久删除数据集中个人信息的过程。匿名化可以帮助保护个人隐私。匿名化数据不再被视为个人数据。

反模式

一种用于解决反复出现的问题的常用解决方案，而在这类问题中，此解决方案适得其反、无效或不如替代方案有效。

应用程序控制

一种安全方法，仅允许使用经批准的应用程序，以帮助保护系统免受恶意软件的侵害。

应用程序组合

有关组织使用的每个应用程序的详细信息的集合，包括构建和维护该应用程序的成本及其业务价值。这些信息是[产品组合发现和分析过程](#)的关键，有助于识别需要进行迁移、现代化和优化的应用程序并确定其优先级。

人工智能 (AI)

计算机科学领域致力于使用计算技术执行通常与人类相关的认知功能，例如学习、解决问题和识别模式。有关更多信息，请参阅[什么是人工智能？](#)

人工智能操作 (AIOps)

使用机器学习技术解决运营问题、减少运营事故和人为干预以及提高服务质量的过程。有关如何在 AIOps AWS 迁移策略中使用的更多信息，请参阅[操作集成指南](#)。

非对称加密

一种加密算法，使用一对密钥，一个公钥用于加密，一个私钥用于解密。您可以共享公钥，因为它不用于解密，但对私钥的访问应受到严格限制。

原子性、一致性、隔离性、耐久性 () ACID

一组软件属性，即使在出现错误、电源故障或其他问题的情况下，也能保证数据库的数据有效性和操作可靠性。

基于属性的访问控制 () ABAC

根据用户属性（如部门、工作角色和团队名称）创建精细访问权限的做法。有关更多信息，[ABAC](#) 请参阅 AWS Identity and Access Management (IAM) 文档 AWS 中的。

权威数据源

存储主要数据版本的位置，被认为是最可靠的信息源。您可以将数据从权威数据源复制到其他位置，以便处理或修改数据，例如对数据进行匿名化、编辑或假名化。

可用区

中的一个不同位置 AWS 区域，不受其他可用区域故障的影响，并向同一区域中的其他可用区提供低成本、低延迟的网络连接。

AWS 云采用框架 (AWS CAF)

该框架包含指导方针和最佳实践 AWS，可帮助组织制定高效且有效的计划，以成功迁移到云端。AWS CAF 将指导分为六个重点领域，称为视角：业务、人员、治理、平台、安全和运营。业务、人员和治理角度侧重于业务技能和流程；平台、安全和运营角度侧重于技术技能和流程。例如，人员角度针对的是负责人力资源 (HR)、人员配置职能和人员管理的利益相关者。从这个角度来看，AWS CAF 为人员发展、培训和沟通提供了指导，以帮助组织为成功采用云做好准备。有关更多信息，请参阅[AWS CAF 网站](#)和[AWS CAF 白皮书](#)。

AWS 工作负载资格框架 (AWS WQF)

一种评估数据库迁移工作负载、推荐迁移策略和提供工作估算的工具。AWS WQF包含在 AWS Schema Conversion Tool (AWS SCT) 中。它用来分析数据库架构和代码对象、应用程序代码、依赖关系和性能特征，并提供评测报告。

B

坏机器人

旨在破坏个人或组织或对其造成伤害的[机器人](#)。

BCP

参见[业务连续性计划](#)。

行为图

一段时间内资源行为和交互的统一交互式视图。您可以在 Amazon Detective 中使用行为图来检查登录尝试失败、可疑API呼叫和类似操作。有关更多信息，请参阅 Detective 文档中的[行为图中的数据](#)。

大端序系统

一个先存储最高有效字节的系统。另请参见[字节顺序](#)。

二进制分类

一种预测二进制结果（两个可能的类别之一）的过程。例如，您的 ML 模型可能需要预测诸如“该电子邮件是否为垃圾邮件？”或“这个产品是书还是汽车？”之类的问题

bloom 筛选条件

一种概率性、内存高效的数据结构，用于测试元素是否为集合的成员。

蓝/绿部署

一种部署策略，您可以创建两个独立但完全相同的环境。在一个环境中运行当前的应用程序版本（蓝色），在另一个环境中运行新的应用程序版本（绿色）。此策略可帮助您在影响最小的情况下快速回滚。

自动程序

一种通过互联网运行自动任务并模拟人类活动或互动的软件应用程序。有些机器人是有用或有益的，例如在互联网上索引信息的网络爬虫。其他一些被称为恶意机器人的机器人旨在破坏个人或组织或对其造成伤害。

僵尸网络

被[恶意软件](#)感染并受单方（称为[机器人](#)牧民或机器人操作员）控制的机器人网络。僵尸网络是最著名的扩展机器人及其影响力的机制。

分支

代码存储库的一个包含区域。在存储库中创建的第一个分支是主分支。您可以从现有分支创建新分支，然后在新分支中开发功能或修复错误。为构建功能而创建的分支通常称为功能分支。当功能可以发布时，将功能分支合并回主分支。有关更多信息，请参阅[关于分支](#)（GitHub 文档）。

破碎的玻璃通道

在特殊情况下，通过批准的流程，用户 AWS 账户 可以快速访问他们通常没有访问权限的内容。有关更多信息，请参阅 [Well -Architected 指南中的“实施破碎玻璃程序”](#) 指示 AWS 器。

棕地策略

您环境中的现有基础设施。在为系统架构采用棕地策略时，您需要围绕当前系统和基础设施的限制来设计架构。如果您正在扩展现有基础设施，则可以将棕地策略和[全新](#)策略混合。

缓冲区缓存

存储最常访问的数据的内存区域。

业务能力

企业如何创造价值（例如，销售、客户服务或营销）。微服务架构和开发决策可以由业务能力驱动。有关更多信息，请参阅[在 AWS 上运行容器化微服务](#)白皮书中的[围绕业务能力进行组织](#)部分。

业务连续性计划 (BCP)

一项计划，旨在应对大规模迁移等破坏性事件对运营的潜在影响，并使企业能够快速恢复运营。

C

CAF

参见[AWS 云采用框架](#)。

金丝雀部署

向最终用户缓慢而渐进地发布版本。当你有信心时，你可以部署新版本并全部替换当前版本。

CCoE

参见[云卓越中心](#)。

CDC

参见[变更数据捕获](#)。

更改数据捕获 (CDC)

跟踪数据来源（如数据库表）的更改并记录有关更改的元数据的过程。您可以CDC用于各种用途，例如审计或复制目标系统中的更改以保持同步。

混沌工程

故意引入故障或破坏性事件来测试系统的弹性。您可以使用 [AWS Fault Injection Service \(AWS FIS\)](#) 来执行实验，对您的 AWS 工作负载施加压力并评估其响应。

CI/CD

查看[持续集成和持续交付](#)。

分类

一种有助于生成预测的分类流程。分类问题的 ML 模型预测离散值。离散值始终彼此不同。例如，一个模型可能需要评估图像中是否有汽车。

客户端加密

在目标 AWS 服务 收到数据之前，对数据进行本地加密。

云卓越中心 (CCoE)

一个多学科团队，负责推动整个组织的云采用工作，包括开发云最佳实践、调动资源、制定迁移时间表、领导组织完成大规模转型。有关更多信息，请参阅 AWS Cloud 企业战略博客上的[CCoE帖子](#)。

云计算

通常用于远程数据存储和 IoT 设备管理的云技术。云计算通常与[边缘计算](#)技术相关。

云运营模型

在 IT 组织中，一种用于构建、完善和优化一个或多个云环境的运营模型。有关更多信息，请参阅[构建您的云运营模型](#)。

云采用阶段

组织迁移到以下阶段时通常会经历四个阶段 AWS Cloud：

- 项目 - 出于概念验证和学习目的，开展一些与云相关的项目
- 基础 — 进行基础投资以扩大云采用率（例如，创建着陆区、定义CCoE、建立运营模型）

- 迁移 - 迁移单个应用程序
- 重塑 - 优化产品和服务，在云中创新

Stephen Orban在 AWS Cloud 企业战略博客的博客文章 [《云优先之旅和采用阶段》](#) 中定义了这些阶段。有关它们与 AWS 迁移策略的关系的信息，请参阅[迁移准备指南](#)。

CMDB

参见[配置管理数据库](#)。

代码存储库

通过版本控制过程存储和更新源代码和其他资产（如文档、示例和脚本）的位置。常见的云存储库包括 GitHub 或者 Bitbucket Cloud。每个版本的代码都称为分支。在微服务结构中，每个存储库都专门用于一个功能。单个 CI/CD 管道可以使用多个存储库。

冷缓存

一种空的、填充不足或包含过时或不相关数据的缓冲区缓存。这会影响性能，因为数据库实例必须从主内存或磁盘读取，这比从缓冲区缓存读取要慢。

冷数据

很少访问的数据，且通常是历史数据。查询此类数据时，通常可以接受慢速查询。将这些数据转移到性能较低且成本更低的存储层或类别可以降低成本。

计算机视觉 (CV)

[人工智能](#)领域，使用机器学习来分析和提取数字图像和视频等视觉格式中的信息。例如，AWS Panorama 提供将 CV 添加到本地摄像机网络的设备，而 Amazon 则为 CV SageMaker 提供图像处理算法。

配置偏差

对于工作负载，配置会从预期状态发生变化。这可能会导致工作负载变得不合规，而且通常是渐进的，不是故意的。

配置管理数据库 (CMDB)

一种存储库，用于存储和管理有关数据库及其 IT 环境的信息，包括硬件和软件组件及其配置。您通常使用来自投资组合CMDB中的迁移发现和分析阶段的数据。

合规性包

一系列 AWS Config 规则和补救措施，您可以汇编这些规则和补救措施，以自定义您的合规性和安全性检查。您可以使用YAML模板将合规包作为单个实体部署在 AWS 账户 和区域或整个组织中。有关更多信息，请参阅 AWS Config 文档中的[一致性包](#)。

持续集成和持续交付 (CI/CD)

自动执行软件发布过程的源代码、构建、测试、暂存和生产阶段的过程。CI/CD is commonly described as a pipeline. CI/CD可以帮助您实现流程自动化、提高生产力、提高代码质量和更快地交付。有关更多信息，请参阅[持续交付的优势](#)。CD 也可以表示持续部署。有关更多信息，请参阅[持续交付与持续部署](#)。

CV

参见[计算机视觉](#)。

D

静态数据

网络中静止的数据，例如存储中的数据。

数据分类

根据网络中数据的关键性和敏感性对其进行识别和分类的过程。它是任何网络安全风险管理策略的关键组成部分，因为它可以帮助您确定对数据的适当保护和保留控制。数据分类是 Well-Architected AWS d Framework 中安全支柱的一个组成部分。有关详细信息，请参阅[数据分类](#)。

数据漂移

生产数据与用来训练机器学习模型的数据之间的有意义差异，或者输入数据随时间推移的有意义变化。数据漂移可能降低机器学习模型预测的整体质量、准确性和公平性。

传输中数据

在网络中主动移动的数据，例如在网络资源之间移动的数据。

数据网格

一种架构框架，可提供分布式、去中心化的数据所有权以及集中式管理和治理。

数据最少化

仅收集并处理绝对必要数据的原则。在中进行数据最小化 AWS Cloud 可以降低隐私风险、成本和分析碳足迹。

数据边界

AWS 环境中的一组预防性防护措施，可帮助确保只有可信身份才能访问来自预期网络的可信资源。有关更多信息，请参阅在[上构建数据边界](#)。AWS

数据预处理

将原始数据转换为 ML 模型易于解析的格式。预处理数据可能意味着删除某些列或行，并处理缺失、不一致或重复的值。

数据溯源

在数据的整个生命周期跟踪其来源和历史的过程，例如数据如何生成、传输和存储。

数据主体

正在收集和处理其数据的个人。

数据仓库

一种支持商业智能（例如分析）的数据管理系统。数据仓库通常包含大量历史数据，通常用于查询和分析。

数据库定义语言 (DDL)

在数据库中创建或修改表和对象结构的语句或命令。

数据库操作语言 (DML)

在数据库中修改（插入、更新和删除）信息的语句或命令。

DDL

参见[数据库定义语言](#)。

深度融合

组合多个深度学习模型进行预测。您可以使用深度融合来获得更准确的预测或估算预测中的不确定性。

深度学习

一个 ML 子字段使用多层神经网络来识别输入数据和感兴趣的目标变量之间的映射。

defense-in-depth

一种信息安全方法，经过深思熟虑，在整个计算机网络中分层实施一系列安全机制和控制措施，以保护网络及其中数据的机密性、完整性和可用性。当你采用这种策略时 AWS，你会在 AWS Organizations 结构的不同层面添加多个控件来帮助保护资源。例如，一种 defense-in-depth 方法可以结合多因素身份验证、网络分段和加密。

委托管理员

在中 AWS Organizations，兼容的服务可以注册 AWS 成员帐户来管理组织的帐户并管理该服务的权限。此帐户被称为该服务的委托管理员。有关更多信息和兼容服务列表，请参阅 AWS Organizations 文档中[使用 AWS Organizations 的服务](#)。

部署

使应用程序、新功能或代码修复在目标环境中可用的过程。部署涉及在代码库中实现更改，然后在应用程序的环境中构建和运行该代码库。

开发环境

参见[环境](#)。

侦测性控制

一种安全控制，在事件发生后进行检测、记录日志和发出警报。这些控制是第二道防线，提醒您注意绕过现有预防性控制的安全事件。有关更多信息，请参阅在 AWS 上实施安全控制中的[侦测性控制](#)。

开发价值流映射 (DVSM)

用于识别对软件开发生命周期中的速度和质量产生不利影响的限制因素并确定其优先级的流程。DVSM 扩展了最初为精益生产实践设计的价值流映射流程。其重点关注在软件开发过程中创造和转移价值所需的步骤和团队。

数字孪生

真实世界系统的虚拟再现，如建筑物、工厂、工业设备或生产线。数字孪生支持预测性维护、远程监控和生产优化。

维度表

在[星型架构](#)中，一种较小的表，其中包含事实表中有关定量数据的数据属性。维度表属性通常是文本字段或行为类似于文本的离散数字。这些属性通常用于查询约束、筛选和结果集标注。

灾难

阻止工作负载或系统在其主要部署位置实现其业务目标的事件。这些事件可能是自然灾害、技术故障或人为操作的结果，例如无意的配置错误或恶意软件攻击。

灾难恢复 (DR)

您用来最大限度地减少[灾难](#)造成的停机时间和数据丢失的策略和流程。有关更多信息，请参阅 Well-Architected Framework AWS work 中的[“工作负载灾难恢复：云端 AWS 恢复”](#)。

DML

参见[数据库操作语言](#)。

领域驱动设计

一种开发复杂软件系统的方法，通过将其组件连接到每个组件所服务的不断发展的领域或核心业务目标。Eric Evans 在其著作领域驱动设计：软件核心复杂性应对之道 (Boston: Addison-Wesley Professional, 2003) 中介绍了这一概念。[有关如何使用带有 strangler fig 模式的域驱动设计的信息，请参阅对旧版 Microsoft 进行现代化改造。ASP NET\(ASMX\) 通过使用容器和 Amazon API Gateway 逐步提供网络服务。](#)

DR

参见[灾难恢复](#)。

漂移检测

跟踪与基线配置的偏差。例如，您可以使用 AWS CloudFormation 来[检测系统资源中的偏差](#)，也可以使用 AWS Control Tower 来[检测着陆区中可能影响监管要求合规性的变化](#)。

DVSM

参见[开发价值流映射](#)。

E

EDA

参见[探索性数据分析](#)。

EDI

参见[电子数据交换](#)。

边缘计算

该技术可提高位于 IoT 网络边缘的智能设备的计算能力。与[云计算](#)相比，边缘计算可以减少通信延迟并缩短响应时间。

电子数据交换 (EDI)

组织之间自动交换业务文档。有关更多信息，请参阅[什么是电子数据交换](#)。

加密

一种将人类可读的纯文本数据转换为密文的计算过程。

加密密钥

由加密算法生成的随机位的加密字符串。密钥的长度可能有所不同，而且每个密钥都设计为不可预测且唯一。

字节顺序

字节在计算机内存中的存储顺序。大端序系统先存储最高有效字节。小端序系统先存储最低有效字节。

端点

参见[服务端点](#)。

端点服务

您可以托管在虚拟私有云 (VPC) 中与其他用户共享的服务。您可以使用其他 AWS 账户 或 AWS Identity and Access Management (IAM) 委托人创建终端节点服务，AWS PrivateLink 并向其授予权限。这些账户或委托人可以通过创建接口终端节点私密连接到您的终VPC端节点服务。有关更多信息，请参阅 Amazon Virtual Private Cloud (AmazonVPC) 文档中的[创建终端节点服务](#)。

企业资源规划 (ERP)

一种自动化和管理企业关键业务流程（例如会计和项目管理）的系统。[MES](#)

信封加密

用另一个加密密钥对加密密钥进行加密的过程。有关更多信息，请参阅 AWS Key Management Service (AWS KMS) 文档中的[信封加密](#)。

environment

正在运行的应用程序的实例。以下是云计算中常见的环境类型：

- 开发环境 — 正在运行的应用程序的实例，只有负责维护应用程序的核心团队才能使用。开发环境用于测试更改，然后再将其提升到上层环境。这类环境有时称为测试环境。
- 下层环境 — 应用程序的所有开发环境，比如用于初始构建和测试的环境。
- 生产环境 — 最终用户可以访问的正在运行的应用程序的实例。在 CI/CD 管道中，生产环境是最后一个部署环境。
- 上层环境 — 除核心开发团队以外的用户可以访问的所有环境。这可能包括生产环境、预生产环境和用户验收测试环境。

epic

在敏捷方法学中，有助于组织工作和确定优先级的功能类别。epics 提供了对需求和实施任务的总体描述。例如，AWS CAF 安全史诗包括身份和访问管理、侦探控制、基础设施安全、数据保护和事件响应。有关 AWS 迁移策略中 epics 的更多信息，请参阅[计划实施指南](#)。

ERP

参见[企业资源规划](#)。

探索性数据分析 () EDA

分析数据集以了解其主要特征的过程。您收集或汇总数据，并进行初步调查，以发现模式、检测异常并检查假定情况。EDA 是通过计算汇总统计数据 and 创建数据可视化来执行的。

F

事实表

[星形架构](#) 中的中心表。它存储有关业务运营的定量数据。通常，事实表包含两种类型的列：包含度量的列和包含维度表外键的列。

失败得很快

一种使用频繁和增量测试来缩短开发生命周期的理念。这是敏捷方法的关键部分。

故障隔离边界

在中 AWS Cloud，诸如可用区 AWS 区域、控制平面或数据平面之类的边界，它限制了故障的影响并有助于提高工作负载的弹性。有关更多信息，请参阅[AWS 故障隔离边界](#)。

功能分支

参见[分支](#)。

特征

您用来进行预测的输入数据。例如，在制造环境中，特征可能是定期从生产线捕获的图像。

特征重要性

特征对于模型预测的重要性。这通常表示为数字分数，可以通过各种技术进行计算，例如 Shapley Additive Explantions (SHAP) 和积分梯度。有关更多信息，请参阅[机器学习模型的可解释性：AWS](#)。

功能转换

为 ML 流程优化数据，包括使用其他来源丰富数据、扩展值或从单个数据字段中提取多组信息。这使得 ML 模型能从数据中获益。例如，如果您将“2021-05-27 00:15:37”日期分解为“2021”、“五月”、“星期四”和“15”，则可以帮助学习与不同数据成分相关的算法学习精细模式。

few-shot 提示

在要求 LLM 其执行类似任务之前，提供少量示例，演示该任务和所需的输出。这种技术是情境学习的应用，模型可以从提示中嵌入的示例（镜头）中学习。对于需要特定格式、推理或领域知识的任务，Few-shot 提示可能非常有效。另请参见[零镜头提示](#)。

FGAC

请参阅[精细的访问控制](#)。

精细的访问控制 () FGAC

使用多个条件允许或拒绝访问请求。

快闪迁移

一种数据库迁移方法，它使用连续的数据复制，通过[更改数据捕获](#)在尽可能短的时间内迁移数据，而不是使用分阶段的方法。目标是将停机时间降至最低。

FM

参见[基础模型](#)。

基础模型 (FM)

一个大型深度学习神经网络，一直在广义和未标记数据的大量数据集上进行训练。FMs 能够执行各种各样的一般任务，例如理解语言、生成文本和图像以及用自然语言进行对话。有关更多信息，请参阅[什么是基础模型](#)。

G

生成式人工智能

[人工智能](#)模型的一个子集，这些模型已经过大量数据训练，可以使用简单的文本提示来创建新的内容和工件，例如图像、视频、文本和音频。有关更多信息，请参阅[什么是生成式 AI](#)。

地理封锁

请参阅[地理限制](#)。

地理限制 (地理阻止)

在 Amazon 中 CloudFront , 一种阻止特定国家/地区的用户访问内容分发的选项。您可以使用允许列表或阻止列表来指定已批准和已禁止的国家/地区。有关更多信息 , 请参阅 CloudFront 文档[中的限制内容的地理分布](#)。

GitFlow 工作流程

一种方法 , 在这种方法中 , 下层和上层环境在源代码存储库中使用不同的分支。Gitflow 工作流程被认为是传统的 , 而[基于主干的工作流程](#)是现代的首选方法。

金色影像

系统或软件的快照 , 用作部署该系统或软件的新实例的模板。例如 , 在制造业中 , 黄金映像可用于在多个设备上配置软件 , 并有助于提高设备制造运营的速度、可扩展性和生产力。

全新策略

在新环境中缺少现有基础设施。在对系统架构采用全新策略时 , 您可以选择所有新技术 , 而不受对现有基础设施 (也称为[棕地](#)) 兼容性的限制。如果您正在扩展现有基础设施 , 则可以将棕地策略和全新策略混合。

防护机制

一项高级规则 , 可帮助管理各组织单位的资源、策略和合规性 (OUs)。预防性防护机制会执行策略以确保符合合规性标准。它们通过使用服务控制策略和IAM权限边界来实现。侦测性防护机制会检测策略违规和合规性问题 , 并生成警报以进行修复。它们通过使用 AWS Config、Amazon、AWS Security Hub GuardDuty AWS Trusted Advisor、Amazon Inspector 和自定义 AWS Lambda 支票来实现。

H

HA

参见[高可用性](#)。

异构数据库迁移

将源数据库迁移到使用不同数据库引擎的目标数据库 (例如 , 从 Oracle 迁移到 Amazon Aurora) 。异构迁移通常是重新架构工作的一部分 , 而转换架构可能是一项复杂的任务。[AWS 提供了 AWS SCT](#) 来帮助实现架构转换。

高可用性 (HA)

在遇到挑战或灾难时，工作负载无需干预即可连续运行的能力。HA 系统旨在自动进行故障转移、持续提供良好性能，并以最小的性能影响处理不同负载和故障。

历史数据库现代化

一种用于实现运营技术 (OT) 系统现代化和升级以更好满足制造业需求的方法。历史数据库是一种用于收集和存储工厂中各种来源数据的数据库。

抵制数据

从用于训练[机器学习](#)模型的数据集中扣留的一部分带有标签的历史数据。通过将模型预测与抵制数据进行比较，您可以使用抵制数据来评估模型性能。

同构数据库迁移

将您的源数据库迁移到共享相同数据库引擎的目标数据库（例如，将 Microsoft SQL Server 迁移到 Amazon RDS 或 SQL Server）。同构迁移通常是更换主机或更换平台工作的一部分。您可以使用本机数据库实用程序来迁移架构。

热数据

经常访问的数据，例如实时数据或近期的转化数据。这些数据通常需要高性能存储层或存储类别才能提供快速的查询响应。

修补程序

针对生产环境中关键问题的紧急修复。由于其紧迫性，修补程序通常是在典型的 DevOps 发布工作流程之外进行的。

hypercure 周期

割接之后，迁移团队立即管理和监控云中迁移的应用程序以解决任何问题的时间段。通常，这个周期持续 1-4 天。在 hypercure 周期结束时，迁移团队通常会将应用程序的责任移交给云运营团队。

I

IaC

参见[基础架构即代码](#)。

基于身份的策略

附加到一个或多个 IAM 委托人的策略，用于定义他们在 AWS Cloud 环境中的权限。

空闲应用程序

一种在 90 天内平均使用率 CPU 和内存使用率介于 5% 到 20% 之间的应用程序。在迁移项目中，通常会停用这些应用程序或将其保留在本地。

IloT

参见 [工业物联网](#)。

不可变的基础架构

一种为生产工作负载部署新基础架构，而不是更新、修补或修改现有基础架构的模型。[不可变基础架构本质上比可变基础架构更一致、更可靠、更可预测](#)。有关更多信息，请参阅 Well-Architected Framework 中的 [使用不可变基础架构 AWS 部署最佳实践](#)。

入站 (入口) VPC

在 AWS 多账户架构中 VPC，接受、检查和路由来自应用程序外部的网络连接。[AWS 安全参考架构](#) 建议设置您的网络帐户，包括入站、出站和检查，VPCs 以保护您的应用程序与更广泛的互联网之间的双向接口。

增量迁移

一种割接策略，在这种策略中，您可以将应用程序分成小部分进行迁移，而不是一次性完整割接。例如，您最初可能只将几个微服务或用户迁移到新系统。在确认一切正常后，您可以逐步迁移其他微服务或用户，直到停用遗留系统。这种策略降低了大规模迁移带来的风险。

工业 4.0

该术语由 [克劳斯·施瓦布 \(Klaus Schwab \)](#) 于 2016 年推出，指的是通过连接、实时数据、自动化、分析和人工智能/机器学习的进步实现制造流程的现代化。

基础设施

应用程序环境中包含的所有资源和资产。

基础设施即代码 (IaC)

通过一组配置文件预置和管理应用程序基础设施的过程。IaC 旨在帮助您集中管理基础设施、实现资源标准化和快速扩展，使新环境具有可重复性、可靠性和一致性。

工业物联网 (IloT)

在工业领域使用联网的传感器和设备，例如制造业、能源、汽车、医疗保健、生命科学和农业。有关更多信息，请参阅 [构建工业物联网 \(IloT\) 数字化转型战略](#)。

检查 VPC

在 AWS 多账户架构中，VPC 一种集中式管理 VPCs (相同或不同 AWS 区域)、互联网和本地网络之间的网络流量的检查。[AWS 安全参考架构](#) 建议设置您的网络帐户，包括入站、出站和检查，VPCs 以保护您的应用程序与更广泛的互联网之间的双向接口。

物联网 (IoT)

由带有嵌入式传感器或处理器的连接物理对象组成的网络，这些传感器或处理器通过互联网或本地通信网络与其他设备和系统进行通信。有关更多信息，请参阅[什么是 IoT ?](#)

可解释性

它是机器学习模型的一种特征，描述了人类可以理解模型的预测如何取决于其输入的程度。有关更多信息，请参阅使用[机器学习模型的可解释性 AWS](#)。

IoT

参见[物联网](#)。

IT 信息库 (ITIL)

提供 IT 服务并使这些服务符合业务要求的一套最佳实践。ITIL 为... 提供了基础 ITSM。

IT 服务管理 (ITSM)

为组织设计、实施、管理和支持 IT 服务的相关活动。有关将云操作与 ITSM 工具集成的信息，请参阅[操作集成指南](#)。

ITIL

请参阅[IT 信息库](#)。

ITSM

请参阅[IT 服务管理](#)。

L

基于标签的访问控制 () LBAC

强制访问控制 (MAC) 的实现，其中明确为用户和数据本身分配了安全标签值。用户安全标签和数据安全标签之间的交集决定了用户可以看到哪些行和列。

登录区

landing zone 是一个架构精良的多账户 AWS 环境，具有可扩展性和安全性。这是一个起点，您的组织可以从这里放心地在安全和基础设施环境中快速启动和部署工作负载和应用程序。有关登录区的更多信息，请参阅[设置安全且可扩展的多账户 AWS 环境](#)。

大型语言模型 (LLM)

一种基于大量数据进行预训练的深度学习 [AI](#) 模型。LLM 可以执行多项任务，例如回答问题、总结文档、将文本翻译成其他语言以及完成句子。有关更多信息，请参阅[什么是 LLMs](#)。

大规模迁移

迁移 300 台或更多服务器。

LBAC

请参阅[基于标签的访问控制](#)。

最低权限

授予执行任务所需的最低权限的最佳安全实践。有关更多信息，请参阅文档中的[应用最低权限权限](#)。IAM

直接迁移

见 [7 R](#)。

小端序系统

一个先存储最低有效字节的系统。另请参见[字节顺序](#)。

LLM

参见[大型语言模型](#)。

下层环境

参见[环境](#)。

M

机器学习 (ML)

一种使用算法和技术进行模式识别和学习的人工智能。ML 对记录的数据 (例如物联网 (IoT) 数据) 进行分析和学习，以生成基于模式的统计模型。有关更多信息，请参阅[机器学习](#)。

主分支

参见[分支](#)。

恶意软件

旨在危害计算机安全或隐私的软件。恶意软件可能会破坏计算机系统、泄露敏感信息或获得未经授权的访问。恶意软件的示例包括病毒、蠕虫、勒索软件、特洛伊木马、间谍软件和键盘记录器。

托管服务

AWS 服务 它 AWS 运行基础设施层、操作系统和平台，您可以访问端点来存储和检索数据。亚马逊简单存储服务 (Amazon S3) Service 和 Amazon DynamoDB 就是托管服务的示例。这些服务也称为抽象服务。

制造执行系统 (MES)

一种软件系统，用于跟踪、监控、记录和控制将原材料转化为成品的生产过程。

MAP

参见[迁移加速计划](#)。

机制

一个完整的过程，在此过程中，您可以创建工具，推动工具的采用，然后检查结果以进行调整。机制是一种在运行过程中自我增强和改进的循环。有关更多信息，请参阅在 Well-Architect AWS ed 框架中[构建机制](#)。

成员账户

AWS 账户 除属于组织中的管理账户之外的所有账户 AWS Organizations。一个账户一次只能是一个组织的成员。

MES

参见[制造执行系统](#)。

消息队列遥测传输 (MQTT)

[一种基于发布/订阅模式的轻量级 machine-to-machine \(M2M\) 通信协议，适用于资源受限的物联网设备。](#)

微服务

一种小型的独立服务，通过明确的定义进行通信APIs，通常由小型的独立团队拥有。例如，保险系统可能包括映射到业务能力（如销售或营销）或子域（如购买、理赔或分析）的微服务。微服务的

好处包括敏捷、灵活扩展、易于部署、可重复使用的代码和恢复能力。有关更多信息，请参阅[使用 AWS 无服务器服务集成微服务](#)。

微服务架构

一种使用独立组件构建应用程序的方法，这些组件将每个应用程序进程作为微服务运行。这些微服务使用轻量级通过定义明确的接口进行通信。APIs 该架构中的每个微服务都可以更新、部署和扩展，以满足对应用程序特定功能的需求。有关更多信息，请参阅[在上实现微服务](#)。AWS

Migration 加速计划 (MAP)

AWS 该计划提供咨询支持、培训和服务，以帮助组织为迁移到云奠定坚实的运营基础，并帮助抵消迁移的初始成本。MAP 包括一种以有条不紊的方式执行遗留迁移的迁移方法，以及一组用于自动化和加速常见迁移场景的工具。

大规模迁移

将大部分应用程序组合分波迁移到云中的过程，在每一波中以更快的速度迁移更多应用程序。本阶段使用从早期阶段获得的最佳实践和经验教训，实施由团队、工具和流程组成的迁移工厂，通过自动化和敏捷交付简化工作负载的迁移。这是 [AWS 迁移策略](#) 的第三阶段。

迁移工厂

跨职能团队，通过自动化、敏捷的方法简化工作负载迁移。迁移工厂团队通常包括运营、业务分析师和所有者、迁移工程师、开发 DevOps 人员和冲刺专业人员。20% 到 50% 的企业应用程序组合由可通过工厂方法优化的重复模式组成。有关更多信息，请参阅本内容集中[有关迁移工厂的讨论](#)和[云迁移工厂指南](#)。

迁移元数据

有关完成迁移所需的应用程序和服务器器的信息。每种迁移模式都需要一套不同的迁移元数据。迁移元数据的示例包括目标子网、安全组和 AWS 账户。

迁移模式

一种可重复的迁移任务，详细列出了迁移策略、迁移目标以及所使用的迁移应用程序或服务。示例：EC2 使用 AWS 应用程序迁移服务重新托管向 Amazon 的迁移。

迁移组合评估 (MPA)

一种在线工具，可提供信息，用于验证迁移到的业务案例。AWS Cloud MPA 提供详细的产品组合评估（服务器规模调整、定价、TCO 比较、迁移成本分析）以及迁移规划（应用程序数据分析和数据收集、应用程序分组、迁移优先级划分和波浪规划）。所有 AWS 顾问和 APN 合作伙伴顾问均可免费使用该 [MPA 工具](#)（需要登录）。

迁移准备情况评估 (MRA)

使用以下方法获取有关组织云就绪状态的见解、确定优势和劣势以及制定行动计划以缩小已发现差距的过程 AWS CAF。有关更多信息，请参阅[迁移准备指南](#)。MRA是[AWS 迁移策略](#)的第一阶段。

迁移策略

用于将工作负载迁移到的方法 AWS Cloud。有关更多信息，请参阅此词汇表中的 [7 R](#) 条目和[动员组织以加快大规模迁移](#)。

ML

参见[机器学习](#)。

现代化

将过时的（原有的或单体）应用程序及其基础设施转变为云中敏捷、弹性和高度可用的系统，以降低成本、提高效率和利用创新。有关更多信息，请参阅[中的应用程序现代化策略](#)。AWS Cloud

现代化准备情况评估

一种评估方式，有助于确定组织应用程序的现代化准备情况；确定收益、风险和依赖关系；确定组织能够在多大程度上支持这些应用程序的未来状态。评估结果是目标架构的蓝图、详细说明现代化进程发展阶段和里程碑的路线图以及解决已发现差距的行动计划。有关更多信息，请参阅[中的评估应用程序的现代化准备情况](#) AWS Cloud。

单体应用程序（单体式）

作为具有紧密耦合进程的单个服务运行的应用程序。单体应用程序有几个缺点。如果某个应用程序功能的需求激增，则必须扩展整个架构。随着代码库的增长，添加或改进单体应用程序的功能也会变得更加复杂。若要解决这些问题，可以使用微服务架构。有关更多信息，请参阅[将单体分解为微服务](#)。

MPA

参见[迁移组合评估](#)。

MQTT

请参阅[消息队列遥测传输](#)。

多分类器

一种帮助为多个类别生成预测（预测两个以上结果之一）的过程。例如，ML 模型可能会询问“这个产品是书、汽车还是手机？”或“此客户最感兴趣什么类别的产品？”

可变基础架构

一种用于更新和修改现有生产工作负载基础架构的模型。为了提高一致性、可靠性和可预测性，Well-Architect AWS ed Framework 建议使用[不可变基础设施](#)作为最佳实践。

O

OAC

请参阅[源站访问控制](#)。

OAI

参见[源访问身份](#)。

OCM

参见[组织变更管理](#)。

离线迁移

一种迁移方法，在这种方法中，源工作负载会在迁移过程中停止运行。这种方法会延长停机时间，通常用于小型非关键工作负载。

OI

参见[运营集成](#)。

OLA

参见[运营层协议](#)。

在线迁移

一种迁移方法，在这种方法中，源工作负载无需离线即可复制到目标系统。在迁移过程中，连接工作负载的应用程序可以继续运行。这种方法的停机时间为零或最短，通常用于关键生产工作负载。

OPC-UA

参见[开放流程通信-统一架构](#)。

开放流程通信-统一架构 (OPC-UA)

一种用于工业自动化的 machine-to-machine (M2M) 通信协议。OPC-UA 提供数据加密、身份验证和授权方案的互操作性标准。

运营层协议 () OLA

一项协议，阐明 IT 职能部门承诺相互提供哪些服务，以支持服务级别协议 () SLA。

操作准备情况审查 (ORR)

一份问题清单和相关的最佳实践，可帮助您理解、评估、预防或缩小事件和可能的故障的范围。有关更多信息，请参阅 Well-Architecte AWS d Frame [ORRwork 中的运营准备情况评估 \(\)](#)。

操作技术 (OT)

与物理环境配合使用以控制工业运营、设备和基础设施的硬件和软件系统。在制造业中，OT 和信息技术 (IT) 系统的集成是[工业 4.0](#) 转型的重点。

运营整合 (OI)

在云中实现运营现代化的过程，包括就绪计划、自动化和集成。有关更多信息，请参阅[运营整合指南](#)。

组织跟踪

由 AWS CloudTrail 此创建的跟踪记录组织 AWS 账户 中所有人的所有事件 AWS Organizations。该跟踪是在每个 AWS 账户 中创建的，属于组织的一部分，并跟踪每个账户的活动。有关更多信息，请参阅 CloudTrail文档中的[为组织创建跟踪](#)。

组织变革管理 (OCM)

一个从人员、文化和领导力角度管理重大、颠覆性业务转型的框架。OCM通过加快变革采用、解决过渡问题以及推动文化和组织变革，帮助组织为新系统和战略做好准备并过渡到新系统和战略。在 AWS 迁移策略中，该框架被称为人员加速，因为云采用项目需要变更的速度。有关更多信息，请参阅[OCM指南](#)。

源站访问控制 (OAC)

在中 CloudFront，一个增强的选项，用于限制访问以保护您的亚马逊简单存储服务 (Amazon S3) 内容。OAC全部支持所有 S3 存储桶 AWS 区域、使用 AWS KMS (SSE-KMS) 的服务器端加密以及对 S3 存储桶的动态PUT和DELETE请求。

源站访问身份 (OAI)

在中 CloudFront，一个用于限制访问权限以保护您的 Amazon S3 内容的选项。使用时OAI，CloudFront 会创建 Amazon S3 可以对其进行身份验证的委托人。经过身份验证的委托人只能通过特定 CloudFront 分配访问 S3 存储桶中的内容。另请参阅 [OAC](#)，它提供了更精细和更增强的访问控制。

ORR

参见[运营准备情况审查](#)。

OT

参见[运营技术](#)。

出站 (出口) VPC

在 AWS 多账户架构中 VPC，用于处理从应用程序内部启动的网络连接。[AWS 安全参考架构](#)建议设置您的网络帐户，包括入站、出站和检查，VPCs 以保护您的应用程序与更广泛的互联网之间的双向接口。

P

权限边界

附加到 IAM 委托人的 IAM 管理策略，用于设置用户或角色可以拥有的最大权限。有关更多信息，请参阅 IAM 文档中的[权限边界](#)。

个人身份信息 (PII)

直接查看其他相关数据或与之配对时可用于合理推断个人身份的信息。的示例 PII 包括姓名、地址和联系信息。

PII

查看[个人身份信息](#)。

playbook

一套预定义的步骤，用于捕获与迁移相关的工作，例如在云中交付核心运营功能。playbook 可以采用脚本、自动化运行手册的形式，也可以是操作现代化环境所需的流程或步骤的摘要。

PLC

参见[可编程逻辑控制器](#)。

PLM

参见[产品生命周期管理](#)。

策略

一个对象，可以在中定义权限（参见[基于身份的策略](#)）、指定访问条件（参见[基于资源的策略](#)）或定义组织中所有账户的最大权限 AWS Organizations（参见[服务控制策略](#)）。

多语言持久性

根据数据访问模式和其他要求，独立选择微服务的数据存储技术。如果您的微服务采用相同的数据存储技术，它们可能会遇到实现难题或性能不佳。如果微服务使用最适合其需求的数据存储，则可以更轻松地实现微服务，并获得更好的性能和可扩展性。有关更多信息，请参阅[在微服务中实现数据持久性](#)。

组合评测

一个发现、分析和确定应用程序组合优先级以规划迁移的过程。有关更多信息，请参阅[评估迁移准备情况](#)。

谓词

返回true或的查询条件false，通常位于子WHERE句中。

谓词下推

一种数据库查询优化技术，可在传输前筛选查询中的数据。这减少了必须从关系数据库检索和处理的数据量，并提高了查询性能。

预防性控制

一种安全控制，旨在防止事件发生。这些控制是第一道防线，帮助防止未经授权的访问或对网络的意外更改。有关更多信息，请参阅在 AWS 上实施安全控制中的[预防性控制](#)。

主体

中 AWS 可以执行操作和访问资源的实体。该实体通常是 AWS 账户、IAM角色或用户的 root 用户。有关更多信息，请参见IAM文档中的[角色承担者术语和概念](#)。

通过设计保护隐私

一种在整个开发过程中考虑隐私的系统工程方法。

私有托管区

一个容器，其中包含有关您希望 Amazon Route 53 如何响应对一个或多个VPCs域名及其子域名的DNS查询的信息。有关更多信息，请参阅 Route 53 文档中的[私有托管区的使用](#)。

主动控制

一种[安全控制](#)措施，旨在防止部署不合规的资源。这些控件会在资源配置之前对其进行扫描。如果资源与控件不兼容，则不会对其进行配置。有关更多信息，请参阅 AWS Control Tower 文档中的[控制参考指南](#)，并参见在上实施安全[控制中的主动](#)控制 AWS。

产品生命周期管理 (PLM)

在产品的整个生命周期中，从设计、开发和上市，到成长和成熟，再到衰落和移除，对产品进行数据和流程的管理。

生产环境

参见[环境](#)。

可编程逻辑控制器 (PLC)

在制造业中，一种高度可靠、适应性强的计算机，用于监控机器并实现制造过程自动化。

提示链接

使用一个[LLM](#)提示的输出作为下一个提示的输入，以生成更好的响应。该技术用于将复杂的任务分解为子任务，或者迭代地完善或扩展初步响应。它有助于提高模型响应的准确性和相关性，并允许获得更精细的个性化结果。

假名化

用占位符值替换数据集中个人标识符的过程。假名化可以帮助保护个人隐私。假名化数据仍被视为个人数据。

publish/subscribe (pub/sub)

一种支持微服务间异步通信的模式，以提高可扩展性和响应能力。例如，在基于微服务的微服务中[MES](#)，微服务可以将事件消息发布到其他微服务可以订阅的频道。系统可以在不更改发布服务的情况下添加新的微服务。

Q

查询计划

一系列步骤，例如指令，用于访问SQL关系数据库系统中的数据。

查询计划回归

当数据库服务优化程序选择的最佳计划不如数据库环境发生特定变化之前时。这可能是由统计数据、约束、环境设置、查询参数绑定更改和数据库引擎更新造成的。

R

RACI矩阵

参见[负责任、负责、咨询、知情 \(RACI \)](#)。

RAG

请参见[检索增强生成](#)。

勒索软件

一种恶意软件，旨在阻止对计算机系统或数据的访问，直到付款为止。

RASCI矩阵

参见[负责任、负责、咨询、知情 \(RACI \)](#)。

RCAC

请参阅[行和列访问控制](#)。

只读副本

用于只读目的的数据库副本。您可以将查询路由到只读副本，以减轻主数据库的负载。

重新架构师

见 [7 R](#)。

恢复点目标 (RPO)

自上一个数据恢复点以来可接受的最长时间。这决定了从上一个恢复点到服务中断之间可接受的数据丢失情况。

恢复时间目标 (RTO)

服务中断和服务恢复之间可接受的最大延迟。

重构

见 [7 R](#)。

区域

地理区域内的 AWS 资源集合。每一个 AWS 区域 都相互隔离，彼此独立，以提供容错、稳定性和弹性。有关更多信息，请参阅[指定 AWS 区域 您的账户可以使用的账户](#)。

回归

一种预测数值的 ML 技术。例如，要解决“这套房子的售价是多少？”的问题 ML 模型可以使用线性回归模型，根据房屋的已知事实（如建筑面积）来预测房屋的销售价格。

重新托管

见 [7 R](#)。

版本

在部署过程中，推动生产环境变更的行为。

搬迁

见 [7 R](#)。

更换平台

见 [7 R](#)。

回购

见 [7 R](#)。

故障恢复能力

应用程序抵御中断或从中断中恢复的能力。在中规划弹性时，[高可用性](#)和[灾难恢复](#)是常见的考虑因素。AWS Cloud有关更多信息，请参阅[AWS Cloud 弹性](#)。

基于资源的策略

一种附加到资源的策略，例如 AmazonS3 存储桶、端点或加密密钥。此类策略指定了允许哪些主体访问、支持的操作以及必须满足的任何其他条件。

负责、负责、咨询、知情 (RACI) 矩阵

定义参与迁移活动和云运营的所有各方的角色和责任的矩阵。矩阵名称源自矩阵中定义的责任类型：负责 (R)、问责 (A)、咨询 (C) 和知情 (I)。支持 (S) 类型是可选的。如果包括支持，则该矩阵称为 RASCI 矩阵；如果将其排除在外，则称为 RACI 矩阵。

响应性控制

一种安全控制，旨在推动对不良事件或偏离安全基线的情况进行修复。有关更多信息，请参阅在 AWS 上实施安全控制中的[响应性控制](#)。

保留

见 [7 R](#)。

退休

见 [7 R](#)。

检索增强生成 (RAG)

一种[生成式 AI](#) 技术，其中在生成响应之前[LLM](#)引用其训练数据源之外的权威数据源。例如，RAG 模型可以对组织的知识库或自定义数据执行语义搜索。有关更多信息，请参阅[什么是RAG](#)。

轮换

定期更新[密钥](#)以使攻击者更难访问凭据的过程。

行和列访问控制 (RCAC)

使用已定义访问规则的基本、灵活的SQL表达式。RCAC由行权限和列掩码组成。

RPO

参见[恢复点目标](#)。

RTO

参见[恢复时间目标](#)。

运行手册

执行特定任务所需的一套手动或自动程序。它们通常是为了简化重复性操作或高错误率的程序而设计的。

S

SAML2.0

许多身份提供商 (IdPs) 使用的开放标准。此功能支持联合单点登录 (SSO)，因此用户无需为组织中的IAM所有人创建用户即可登录 AWS Management Console 或调用 AWS API操作。有关SAML基于 2.0 的联合身份验证的更多信息，请参阅文档中的[关于SAML基于 2.0 的联合](#)。IAM

SCADA

参见[监督控制和数据采集](#)。

SCP

参见[服务控制政策](#)。

secret

在中 AWS Secrets Manager，您以加密形式存储的机密或受限信息，例如密码或用户凭证。它由密钥值及其元数据组成。密钥值可以是二进制、单个字符串或多个字符串。有关更多信息，请参阅 [Secret s Manager 密钥中有什么？](#) 在 Secrets Manager 文档中。

安全性源于设计

一种在整个开发过程中考虑安全性的系统工程方法。

安全控制

一种技术或管理防护机制，可防止、检测或降低威胁行为体利用安全漏洞的能力。安全控制主要有四种类型：[预防性](#)、[侦测](#)、[响应式](#)和[主动式](#)。

安全加固

缩小攻击面，使其更能抵御攻击的过程。这可能包括删除不再需要的资源、实施授予最低权限的最佳安全实践或停用配置文件中不必要的功能等操作。

安全信息和事件管理 (SIEM) 系统

结合了安全信息管理 (SIM) 和安全事件管理 (SEM) 系统的工具和服务。SIEM系统收集、监控和分析来自服务器、网络、设备和其他来源的数据，以检测威胁和安全漏洞，并生成警报。

安全响应自动化

一种预定义和编程的操作，旨在自动响应或修复安全事件。这些自动化可作为[侦探或响应式](#)安全控制措施，帮助您实施 AWS 安全最佳实践。自动响应操作的示例包括修改VPC安全组、修补 Amazon EC2 实例或轮换证书。

服务器端加密

由接收数据的人在目的地对数据 AWS 服务 进行加密。

服务控制策略 (SCP)

一种策略，用于集中控制 AWS Organizations的组织中所有账户的权限。SCPs定义防护措施或限制管理员可以委托给用户或角色的操作。您可以使用SCPs允许列表或拒绝列表来指定允许或禁止哪些服务或操作。有关更多信息，请参阅 AWS Organizations 文档中的[服务控制策略](#)。

服务端点

URL的入口点的 AWS 服务。您可以使用端点，通过编程方式连接到目标服务。有关更多信息，请参阅 AWS 一般参考 中的 [AWS 服务 端点](#)。

服务级别协议 () SLA

一份协议，阐明了 IT 团队承诺向客户交付的内容，比如服务正常运行时间和性能。

服务级别指示器 () SLI

对服务性能方面的衡量，例如其错误率、可用性或吞吐量。

服务级别目标 () SLO

代表服务运行状况的目标指标，由服务[级别指标](#)衡量。

责任共担模式

描述您在云安全与合规方面共同承担 AWS 的责任的模型。AWS 负责云的安全，而您则负责云中的安全。有关更多信息，请参阅[责任共担模式](#)。

SIEM

参见[安全信息和事件管理系统](#)。

单点故障 (SPOF)

应用程序的单个关键组件出现故障，可能会中断系统。

SLA

参见[服务级别协议](#)。

SLI

参见[服务级别指标](#)。

SLO

参见[服务级别目标](#)。

split-and-seed 模型

一种扩展和加速现代化项目的模式。随着新功能和产品发布的定义，核心团队会拆分以创建新的产品团队。这有助于扩展组织的能力和服务，提高开发人员的工作效率，支持快速创新。有关更多信息，请参阅[中的分阶段实现应用程序现代化的方法。AWS Cloud](#)

SPOF

参见[单点故障](#)。

星型架构

一种数据库组织结构，它使用一个大型事实表来存储交易数据或测量数据，并使用一个或多个较小的维度表来存储数据属性。此结构专为在[数据仓库](#)中使用或用于商业智能目的而设计。

strangler fig 模式

一种通过逐步重写和替换系统功能直至可以停用原有的系统来实现单体系统现代化的方法。这种模式用无花果藤作为类比，这种藤蔓成长为一棵树，最终战胜并取代了宿主。该模式是由 [Martin Fowler](#) 提出的，作为重写单体系统时管理风险的一种方法。有关如何应用此模式的示例，请参阅 [对旧版 Microsoft ASP 进行现代化改造。NET\(ASM\) 通过使用容器和 Amazon API Gateway 逐步提供网络服务。](#)

子网

您的 IP 地址范围 VPC。子网必须位于单个可用区中。

监督控制和数据采集 (SCADA)

在制造业中，一种使用硬件和软件来监控有形资产和生产操作的系统。

对称加密

一种加密算法，它使用相同的密钥来加密和解密数据。

综合测试

以模拟用户交互的方式测试系统，以检测潜在问题或监控性能。您可以使用 [Amazon S CloudWatch ynthetic](#) 来创建这些测试。

系统提示符

一种向提供上下文、说明或指导方针 [LLM](#) 以指导其行为的技术。系统提示有助于设置上下文并制定与用户交互的规则。

T

标签

键值对，用作组织资源的元数据。AWS 标签可帮助您管理、识别、组织、搜索和筛选资源。有关更多信息，请参阅 [标记您的 AWS 资源](#)。

目标变量

您在监督式 ML 中尝试预测的值。这也被称为结果变量。例如，在制造环境中，目标变量可能是产品缺陷。

任务列表

一种通过运行手册用于跟踪进度的工具。任务列表包含运行手册的概述和要完成的常规任务列表。对于每项常规任务，它包括预计所需时间、所有者和进度。

测试环境

参见[环境](#)。

训练

为您的 ML 模型提供学习数据。训练数据必须包含正确答案。学习算法在训练数据中查找将输入数据属性映射到目标（您希望预测的答案）的模式。然后输出捕获这些模式的 ML 模型。然后，您可以使用 ML 模型对不知道目标的新数据进行预测。

中转网关

一个网络传输中心，可用于将您的网络VPCs和本地网络互连。有关更多信息，请参阅 AWS Transit Gateway 文档中的[什么是公交网关](#)。

基于中继的工作流程

一种方法，开发人员在功能分支中本地构建和测试功能，然后将这些更改合并到主分支中。然后，按顺序将主分支构建到开发、预生产和生产环境。

可信访问权限

向您指定的服务授予权限，该服务可以代表您在其账户中执行任务。AWS Organizations 当需要服务相关的角色时，受信任的服务会在每个账户中创建一个角色，为您执行管理任务。有关更多信息，请参阅 AWS Organizations 文档中的[AWS Organizations 与其他 AWS 服务一起使用](#)。

优化

更改训练过程的各个方面，以提高 ML 模型的准确性。例如，您可以通过生成标签集、添加标签，并在不同的设置下多次重复这些步骤来优化模型，从而训练 ML 模型。

双披萨团队

一个小 DevOps 团队，你可以用两个披萨来喂食。双披萨团队的规模可确保在软件开发过程中充分协作。

U

不确定性

这一概念指的是不精确、不完整或未知的信息，这些信息可能会破坏预测式 ML 模型的可靠性。不确定性有两种类型：认知不确定性是由有限的、不完整的数据造成的，而偶然不确定性是由数据中固有的噪声和随机性导致的。有关更多信息，请参阅[量化深度学习系统中的不确定性指南](#)。

无差别任务

也称为繁重工作，即创建和运行应用程序所必需的工作，但不能为最终用户提供直接价值或竞争优势。无差别任务的示例包括采购、维护和容量规划。

上层环境

参见[环境](#)。

V

vacuum 操作

一种数据库维护操作，包括在增量更新后进行清理，以回收存储空间并提高性能。

版本控制

跟踪更改的过程和工具，例如存储库中源代码的更改。

VPC凝视

两者之间的连接VPCs，允许您使用私有 IP 地址路由流量。有关更多信息，请参阅 Amazon VPC 文档中的[什么是VPC对等互连](#)。

漏洞

损害系统安全的软件缺陷或硬件缺陷。

W

热缓存

一种包含经常访问的当前相关数据的缓冲区缓存。数据库实例可以从缓冲区缓存读取，这比从主内存或磁盘读取要快。

暖数据

不常访问的数据。查询此类数据时，通常可以接受中速查询。

窗口函数

一种对一组以某种方式与当前记录相关的行进行计算的SQL函数。窗口函数对于处理任务很有用，例如计算移动平均线或根据当前行的相对位置访问行的值。

工作负载

一系列资源和代码，它们可以提供商业价值，如面向客户的应用程序或后端过程。

工作流

迁移项目中负责一组特定任务的职能小组。每个工作流都是独立的，但支持项目中的其他工作流。例如，组合工作流负责确定应用程序的优先级、波次规划和收集迁移元数据。组合工作流将这些资产交付给迁移工作流，然后迁移服务器和应用程序。

WORM

参见[一次写入，多读](#)。

WQF

参见[AWS工作负载资格框架](#)。

写一次，读多次 (WORM)

一种存储模型，它可以一次写入数据并防止数据被删除或修改。授权用户可以根据需要多次读取数据，但他们无法对其进行更改。这种数据存储基础架构被认为是[不可变的](#)。

Z

零日漏洞利用

一种利用未修补[漏洞](#)的攻击，通常是恶意软件。

零日漏洞

生产系统中不可避免的缺陷或漏洞。威胁主体可能利用这种类型的漏洞攻击系统。开发人员经常因攻击而意识到该漏洞。

零镜头提示

提供执行任务[LLM](#)的说明，但没有可以帮助指导任务的示例（镜头）。LLM必须使用其预先训练的知识来处理任务。零镜头提示的有效性取决于任务的复杂性和提示的质量。另请参阅[few-shot 提示](#)。

僵尸应用程序

平均值CPU和内存使用率低于 5% 的应用程序。在迁移项目中，通常会停用这些应用程序。

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。