



Archivado de datos en Amazon RDS for MySQL, Amazon RDS for MariaDB y Aurora My: compatible SQL

AWS Guía prescriptiva



AWS Guía prescriptiva: Archivado de datos en Amazon RDS for MySQL, Amazon RDS for MariaDB y Aurora My: compatible SQL

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

Las marcas comerciales y la imagen comercial de Amazon no se pueden utilizar en relación con ningún producto o servicio que no sea de Amazon, de ninguna manera que pueda causar confusión entre los clientes y que menosprecie o desacredite a Amazon. Todas las demás marcas registradas que no son propiedad de Amazon son propiedad de sus respectivos propietarios, que pueden o no estar afiliados, conectados o patrocinados por Amazon.

Table of Contents

Introducción	1
Información general	1
Resultados específicos	2
Archivar desde tablas particionadas	4
Archivar desde tablas sin particiones	6
Mover datos a Amazon S3	8
Utilice SELECT INTO OUTFILE S3	8
Utilice Data Pipeline	9
Usa AWS Glue	9
Acceder a los datos archivados	13
Clase de almacenamiento estándar	13
Clases de almacenamiento de S3 Glacier	15
Prácticas recomendadas	16
Limpieza	18
Recursos	19
Apéndice I	20
Apéndice II	22
Historial de documentos	25
Glosario	26
#	26
A	27
B	30
C	32
D	35
E	40
F	42
G	43
H	44
I	45
L	48
M	49
O	53
P	56
Q	59

R	59
S	62
T	66
U	67
V	68
W	68
Z	70
.....	lxxi

Archivar datos en Amazon RDS para MySQL, Amazon RDS para MariaDB y Aurora, compatible con MySQL

Shyam Sunder Rakhecha, Abhishek Karmakar, Oliver Francis y Saumya Singh Amazon Web Services (AWS)

abril de 2023([historial de documentos](#))

La necesidad de archivar datos históricos puede deberse a diferentes casos de uso. Es posible que su aplicación se haya diseñado sin capacidad de archivado, y el crecimiento de su empresa a lo largo del tiempo podría generar grandes cantidades de datos históricos. Esto conduce inevitablemente a una degradación del rendimiento. También puede conservar datos históricos debido a los requisitos de cumplimiento de su organización.

En esta guía se explica cómo archivar los datos históricos en Amazon Simple Storage Service (Amazon S3) con un impacto mínimo en la aplicación y cómo recuperar la información archivada cuando la necesite.

Información general

Esta guía describe diferentes enfoques para archivar datos históricos de tablas grandes en Amazon Relational Database Service (Amazon RDS) para MySQL, Amazon RDS para MariaDB y Amazon Aurora MySQL en la nube de Amazon Web Services (AWS). En esta guía, aprenderá a archivar datos de tablas particionadas y datos que no están particionados y residen en tablas grandes. Puede implementar los enfoques presentados en la guía para reducir el tamaño de sus datos en vivo y, al mismo tiempo, conservar los datos históricos importantes para su posterior análisis.

Archivar los datos de las tablas con regularidad da como resultado un conjunto más reducido de datos en tiempo real en las tablas, lo que permite lecturas y escrituras más rápidas y mejora el rendimiento de la aplicación. El archivado regular de datos se incluye en los pilares de excelencia operativa y eficiencia del rendimiento del [Marco bien diseñado](#). Al mover datos más antiguos a Amazon Simple Storage Service (Amazon S3) y limpiar los datos archivados en su instancia de Amazon RDS o en un clúster compatible con Aurora MySQL, puede ahorrar en costes de almacenamiento. Esto se ajusta al pilar de la optimización de costes y le ayuda a evitar costes innecesarios en AWS.

Resultados empresariales específicos

Esta guía se centra en los siguientes resultados empresariales:

- Experiencia de usuario mejorada
- Se cumplieron los requisitos de cumplimiento de datos
- Costes de almacenamiento reducidos
- Datos organizados

Experiencia de usuario mejorada

Las bases de datos que conservan datos históricos pueden tener un rendimiento lento debido al gran tamaño de las tablas e índices. Al archivar los datos históricos, se reducen las tablas y los índices. Esto tiene un impacto positivo directo en las operaciones de API orientadas al cliente que interactúan con la base de datos.

Se cumplieron los requisitos de cumplimiento de datos

Industrias como los servicios financieros, las organizaciones del sector público y la atención médica tienen requisitos de archivo estrictos. Al archivar los datos de las aplicaciones que residen en su base de datos compatible con Amazon RDS para MySQL, Amazon RDS para MariaDB o Aurora MySQL en Amazon S3, puede cumplir con los requisitos de cumplimiento normativo, incluidos los siguientes:

- La norma de seguridad de datos del sector de pagos con tarjeta (PCI DSS)
- Ley de Portabilidad y Responsabilidad del Seguro Médico (HIPAA) y Ley de Tecnología de la Información Sanitaria para la Salud Económica y Clínica (HITECH)
- Programa Federal de Administración de Riesgos y Autorizaciones (Federal Risk and Authorization Management Program, FedRAMP)
- Reglamento General de Protección de Datos (RGPD)
- Estándares federales de procesamiento de información (FIPS) 140-2
- Instituto Nacional de Estándares y Tecnología (NIST) 800—171

Costes de almacenamiento reducidos

Mantener los datos en Amazon RDS aumenta el costo de almacenamiento y requiere mayores IOPS. Si comparas el costo del almacenamiento por GB al mes para [Amazon RDS para MySQL Multi-AZ GP2](#) con la de [Amazon S3 Glacier](#) en el us-east-1 En la región de AWS, el costo de almacenamiento de S3 Glacier es aproximadamente 57 veces menor que el de Amazon RDS.

Datos organizados

Es bueno mantener en la base de datos datos informativos a los que la aplicación accederá con frecuencia. Sin embargo, las aplicaciones generan una gran cantidad de datos que no se necesitan con mucha frecuencia o se vuelven obsoletos. Estos registros se pueden archivar y mantener en su lugar, lo que resulta rentable y no afecta al rendimiento de la aplicación.

Archivar datos en tablas particionadas

MySQL admite [partición](#) para el motor de almacenamiento InnoDB, y puede utilizar esta función para particionar tablas grandes. Las particiones de la tabla se almacenan como tablas físicas independientes, aunque el SQL que opera en la tabla particionada lee toda la tabla. Esto le da la libertad de eliminar particiones innecesarias de la tabla sin tener que realizar ninguna operación row-by-row elimina, para que pueda archivar las filas históricas de la base de datos.

Considera el siguiente código de ejemplo. `TABLE orders` existe dentro de `orderprocessingschema`. Sus datos históricos están presentes en la partición `historical`, que contiene datos de 2021 y anteriores. En la misma tabla, los datos activos a nivel de aplicación están presentes en las particiones activas de cada mes de 2022. Para archivar los datos de la partición `historical`, puedes crear un archivo `table orders_2021_and_older` con la misma estructura en `archiveeschema`. A continuación, puede utilizar el MySQL [PARTICIÓN DE INTERCAMBIO](#) mover la partición `historical` en esa mesa. Tenga en cuenta que la tabla de archivado no está particionada. Tras archivarlos, puede verificar sus datos y moverlos a Amazon S3.

```
CREATE TABLE orders (
  orderid bigint NOT NULL AUTO_INCREMENT,
  customerid bigint DEFAULT NULL,
  .....
  .....
  order_date date NOT NULL,
  PRIMARY KEY (`orderid`, `order_date`))
PARTITION BY RANGE (TO_DAYS(order_date)) (
  PARTITION pstart VALUES LESS THAN (0),
  PARTITION phistorical VALUES LESS THAN (TO_DAYS('2022-01-01')),
  PARTITION p2022JAN VALUES LESS THAN (TO_DAYS('2022-02-01')),
  PARTITION p2022FEB VALUES LESS THAN (TO_DAYS('2022-03-01')),
  PARTITION p2022MAR VALUES LESS THAN (TO_DAYS('2022-04-01')),
  PARTITION p2022APR VALUES LESS THAN (TO_DAYS('2022-05-01')),
  PARTITION p2022MAY VALUES LESS THAN (TO_DAYS('2022-06-01')),
  PARTITION p2022JUN VALUES LESS THAN (TO_DAYS('2022-07-01')),
  PARTITION p2022JUL VALUES LESS THAN (TO_DAYS('2022-08-01')),
  PARTITION p2022AUG VALUES LESS THAN (TO_DAYS('2022-09-01')),
  PARTITION p2022SEP VALUES LESS THAN (TO_DAYS('2022-10-01')),
  PARTITION p2022OCT VALUES LESS THAN (TO_DAYS('2022-11-01')),
  PARTITION p2022NOV VALUES LESS THAN (TO_DAYS('2022-12-01')),
  PARTITION p2022DEC VALUES LESS THAN (TO_DAYS('2023-01-01')),
  PARTITION pfuture VALUES LESS THAN MAXVALUE
```



```
);  
CREATE TABLE orders_2021_and_older (  
    orderid bigint NOT NULL AUTO_INCREMENT,  
    customerid bigint DEFAULT NULL,  
    .....  
    .....  
    order_date date NOT NULL,  
    PRIMARY KEY (`orderid`, `order_date`));  
mysql> alter table orderprocessing.orders exchange partition phistorical with table  
archive.orders_2021_and_older;  
Query OK, 0 rows affected (0.33 sec)
```

Cuando usa el `EXCHANGE PARTITION` función para archivar datos históricos, recomendamos las siguientes prácticas recomendadas:

- Cree un esquema independiente para almacenar los datos archivados en la aplicación. Este esquema contendrá tablas de archivado que albergarán los datos archivados. Una tabla de archivado del esquema de archivado debe tener la misma estructura que la tabla dinámica, incluidos sus índices y su clave principal. Sin embargo, la tabla de archivado de destino no puede ser una tabla particionada. El intercambio de particiones entre dos tablas particionadas no está permitido en MySQL.
- Siga una convención de nomenclatura para la tabla de archivado que le ayude a identificar los datos históricos almacenados en ella. Esto resulta útil cuando realiza tareas de auditoría o trabajos de diseño que transfieren estos datos a Amazon S3.
- Realice el `EXCHANGE PARTITION` sentencia de lenguaje de definición de datos (DDL) en una ventana de tiempo de inactividad cuando no hay tráfico en sus instancias de escritor compatible con Aurora MySQL, Amazon RDS para MySQL o Amazon RDS para MariaDB.

Podría ser posible ejecutar `EXCHANGE PARTITION` durante las ventanas de poco tráfico de su aplicación o microservicio. Sin embargo, no debe haber escrituras ni seleccionar ninguna o muy pocas selecciones en la tabla particionada. Las consultas selectas existentes y de larga duración pueden provocar que `EXCHANGE PARTITION` DDL espere, lo que provoca una contención de recursos en la base de datos. Diseñe scripts que comprueben que se cumplen todas estas condiciones antes de ejecutarlas `EXCHANGE PARTITION` en su sistema.

Si el diseño de su aplicación admite datos particionados y actualmente tiene una tabla sin particiones, considere la posibilidad de mover los datos a tablas particionadas para poder archivarlos. Para obtener más información, consulte la [documentación de MySQL](#).

Archivar datos de tablas sin particiones

En las tablas de bases de datos donde no es posible crear particiones, puede utilizar el kit de herramientas Percona [pt-archiver](#) herramienta para archivar los datos de su tabla en otra tabla de su base de datos MySQL.

La herramienta pt-archiver se utiliza para archivar los registros de tablas grandes en otras tablas o archivos. Es una herramienta de lectura/escritura, lo que significa que elimina los datos de la tabla de origen después de archivarla, por lo que no tiene que gestionar la eliminación de los datos de origen por separado. El objetivo principal de este script es archivar los datos antiguos de la tabla sin afectar a la carga de consultas del procesamiento de transacciones en línea (OLTP) existente (consulte el apéndice I) e insertar los datos en otra tabla del mismo servidor o de uno diferente.

Puede descargar el [Kit de herramientas Percona](#) [instalar](#) en su máquina local o en la instancia de Amazon Elastic Compute Cloud (Amazon EC2) desde donde se conecta a la base de datos. Para ejecutar la herramienta pt-archiver, utilice la siguiente sintaxis.

```
pt-archiver --source h=<HOST>,D=<DATABASE>,t=<TABLE>,u=<USER>,p=<PASSWORD> --dest  
h=<HOST>,D=<DATABASE>,t=<TABLE> --where "'1=1'" --statistics
```

Sustituya el `HOST`, `DATABASE`, `TABLE`, y `USER` con los detalles y credenciales de la base de datos de origen y destino.

También puedes usar [AWS Batch](#) para crear y programar este trabajo para sus tablas.

Cuando utilice la herramienta pt-archiver para archivar los datos de la tabla, tenga en cuenta lo siguiente:

- Tener una clave principal en la tabla de origen mejorará el rendimiento de esta herramienta. Si la tabla no tiene una clave principal, puede [crear un índice en una columna única](#), lo que ayudará a pt-archiver a revisar todas las filas de la tabla y archivarlas.
- De forma predeterminada, pt-archiver elimina los datos después de archivar la tabla. Antes de ejecutarlo en el servidor de producción, asegúrese de probar sus trabajos de archivado con `--dry-run`. Alternativamente, puede utilizar el `--no-delete` opción.
- La herramienta pt-archiver ajusta su velocidad de archivado en función de la carga del sistema (consulte el apéndice II). Con cargas más altas, puede esperar un rendimiento de archivado más lento.

Después de ejecutar `pt-archiver`, los datos archivados deberían estar en la tabla correspondiente del esquema de archivado. Desde allí, puede moverlo a Amazon S3.

Traslado de datos de tablas archivados a Amazon S3

Amazon Simple Storage Service (Amazon S3) es el objetivo natural de los datos archivados. Ofrece una durabilidad del 99,999999999 por ciento y es menos costoso que el almacenamiento de bases de datos.

Además, Amazon S3 tiene clases de almacenamiento integradas cuyo precio se basa en el patrón de recuperación. Tiene la opción de hacer la transición de los objetos de S3 descargados a un nivel de almacenamiento de menor precio en función de la frecuencia de recuperación de los datos. Para obtener más información sobre las clases de almacenamiento y los precios, consulte la [documentación de Amazon S3](#).

En el caso de las aplicaciones que utilizan flotas de SQL instancias My, la descarga a Amazon S3 implicaría ahorrar dinero en datos que cumplan los siguientes criterios:

- Debe archivarse desde la base de datos para aumentar la eficiencia
- No se requiere de forma inmediata o se necesita con moderación para ningún proceso empresarial
- Debe conservarse durante un período prolongado debido a los requisitos de auditoría

Puede archivar Mis SQL datos de las siguientes maneras:

- Exporte datos mediante `SELECT INTO OUTFILE S3`.
- Exporte datos mediante AWS Data Pipeline.
- Exporta datos con AWS Glue.

Exporte datos mediante `SELECT INTO OUTFILE S3`

Para copiar datos de la base de datos Aurora My SQL -Compatible directamente a Amazon S3, puede usar la declaración `SELECT INTO OUTFILE S3`. Esta SQL declaración se puede ejecutar en la tabla y el número necesario de filas se puede descargar como archivos de valores separados por comas (CSV) con un tamaño máximo de 6 GB. Cuando se supera el umbral de 6 GB, se crean varios archivos.csv.

Para que la exportación funcione, configure lo siguiente:

- AWS Funciones y políticas de Identity and Access Management (IAM)

- Permisos de base de datos otorgados al usuario para ejecutar el comando
- La ubicación de Amazon S3 para la descarga

Para obtener más información, consulte la [documentación de Amazon Aurora](#).

Nota: Para hacer la transición de los objetos S3 a niveles de almacenamiento que ahorren costes, recomendamos configurar las reglas [del ciclo de vida de Amazon S3](#) en el depósito de S3 al que se exportan los datos.

Exportación de datos mediante AWS Data Pipeline

AWS Data Pipeline es un servicio web para flujos de trabajo basados en datos. Puede usar Data Pipeline para automatizar el movimiento de datos de Amazon RDS for My SQL a Amazon S3. Data Pipeline ejecuta una serie de acciones o tareas en serie y exporta los datos a Amazon S3 como un archivo.csv.

La canalización se puede crear de cualquiera de las siguientes maneras:

- En la consola, seleccione la plantilla Copia completa de RDS Mi SQL tabla en S3 como fuente y proporcione valores para los parámetros.
- Desde AWSCLI, defina los objetos de la canalización de datos comoJSON, cree la canalización, cárguela y actívela.

La ventaja de usar Data Pipeline es que el trabajo de exportación se puede programar y monitorear. Además, para mesas más grandes, puedes usar Amazon EMR en lugar de AmazonEC2. Para obtener más información y ejemplos, consulte la [documentación de AWS Data Pipeline](#).

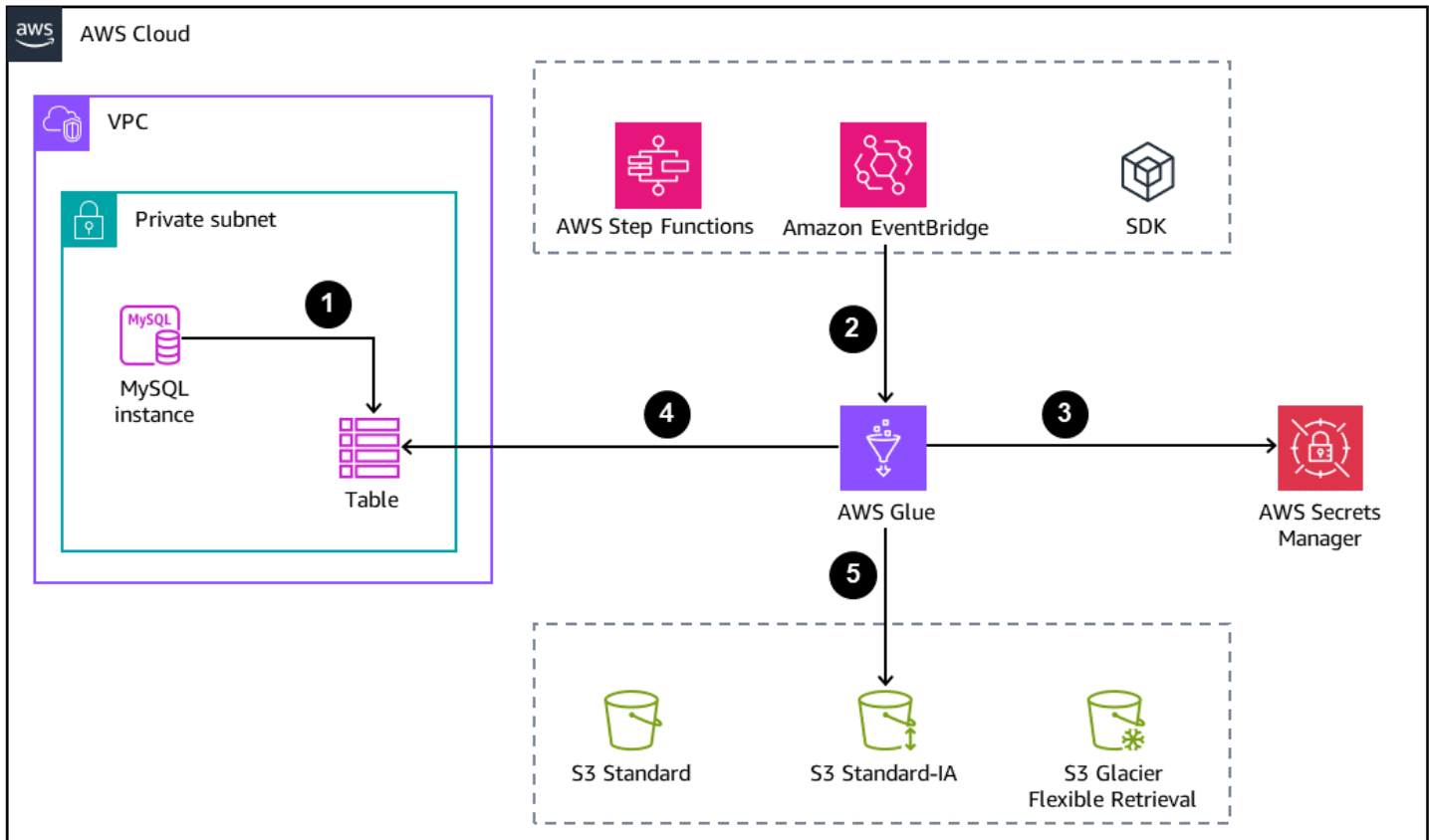
Nota: Para hacer la transición de los objetos S3 a niveles de almacenamiento que ahorren costes, recomendamos configurar las reglas [del ciclo de vida de Amazon S3](#) en el depósito de S3 en el que se realiza la exportación de datos.

Exportación de datos mediante AWS Glue

Puede archivar mis SQL datos en Amazon S3 mediante AWS Glue, que es un servicio analítico sin servidor para escenarios de big data. AWS Glue funciona con Apache Spark, un marco de computación en clúster distribuido ampliamente utilizado que admite muchas fuentes de bases de datos.

La descarga de los datos archivados de la base de datos a Amazon S3 se puede realizar con unas pocas líneas de código en un trabajo de AWS Glue. La mayor ventaja que ofrece AWS Glue es la escalabilidad horizontal y un pay-as-you-go modelo, que proporciona eficiencia operativa y optimización de costos.

El siguiente diagrama muestra una arquitectura básica para el archivado de bases de datos.



1. Mi SQL base de datos crea el archivo o la tabla de respaldo que se descargará en Amazon S3.
2. Un trabajo de AWS Glue se inicia mediante uno de los siguientes enfoques:
 - Sincrónicamente como un paso dentro de una máquina de estados de [AWS Step Functions](#)
 - [De forma asíncrona mediante un evento de Amazon EventBridge](#)
 - Mediante una solicitud manual mediante o un AWS CLI [AWSSDK](#)
3. Las credenciales de base de datos se obtienen de [AWS Secrets Manager](#).
4. El trabajo AWS Glue utiliza una conexión de conectividad de bases de datos Java (JDBC) para acceder a la base de datos y leer la tabla.
5. AWS Glue escribe los datos en Amazon S3 en formato Parquet, que es un formato de datos abierto, en columnas y que ahorra espacio.

Configuración del AWS Glue Job

Para funcionar según lo previsto, el trabajo de AWS Glue requiere los siguientes componentes y configuraciones:

- [AWSConexiones de pegamento](#): se trata de un objeto del catálogo de datos de AWS Glue que se adjunta al trabajo para acceder a la base de datos. Un trabajo puede tener varias conexiones para realizar llamadas a varias bases de datos. Las conexiones contienen las credenciales de la base de datos almacenadas de forma segura.
- [GlueContext](#)— Se trata de un contenedor personalizado sobre [SparkContext](#) la GlueContext clase que proporciona API operaciones de orden superior para interactuar con Amazon S3 y las fuentes de bases de datos. Permite la integración con el catálogo de datos. También elimina la necesidad de depender de los controladores para la conexión a la base de datos, que se gestiona dentro de la conexión Glue. Además, la GlueContext clase proporciona formas de gestionar API las operaciones de Amazon S3, lo que no es posible con la SparkContext clase original.
- IAMpolíticas y funciones: dado que AWS Glue interactúa con otros AWS servicios, debe configurar las funciones adecuadas con el mínimo de privilegios requerido. Entre los servicios que requieren los permisos adecuados para interactuar con AWS Glue se incluyen los siguientes:
 - Amazon S3
 - AWS Secrets Manager
 - AWSServicio de administración de claves (AWSKMS)

prácticas recomendadas

- Para leer tablas enteras que tengan un gran número de filas para descargar, recomendamos utilizar el punto final de réplica de lectura para aumentar el rendimiento de lectura sin degradar el rendimiento de la instancia de escritura principal.
- Para lograr eficiencia en la cantidad de nodos utilizados para procesar el trabajo, active el [escalado automático](#) en AWS Glue 3.0.
- Si el depósito S3 forma parte de la arquitectura de un lago de datos, recomendamos descargar los datos organizándolos en particiones físicas. El esquema de particiones debe basarse en los patrones de acceso. La partición basada en valores de fecha es una de las prácticas más recomendadas.

- Guardar los datos en formatos abiertos, como Parquet o Optimized Row Columnar (ORC), ayuda a que los datos estén disponibles para otros servicios analíticos, como Amazon Athena y Amazon Redshift.
- Para que otros servicios distribuidos optimicen la lectura de los datos descargados, se debe controlar la cantidad de archivos de salida. Casi siempre es beneficioso tener un número menor de archivos grandes en lugar de un gran número de archivos pequeños. Spark tiene archivos de configuración y métodos integrados para controlar la generación de archivos parciales.
- Los datos archivados, por definición, son conjuntos de datos a los que se accede con frecuencia. Para lograr la rentabilidad del almacenamiento, la clase Amazon S3 debe pasar a niveles menos costosos. Esto se puede hacer mediante dos enfoques:
 - Transición sincrónica del nivel durante la descarga: si sabe de antemano que los datos descargados deben transferirse como parte del proceso, puede utilizar el mecanismo GlueContext transition_s3_path [dentro](#) del mismo trabajo de Glue que escribe los datos en Amazon S3. [AWS](#)
 - Transición asíncrona mediante [S3 Lifecycle](#): configure las reglas del ciclo de vida de S3 con los parámetros adecuados para la transición y el vencimiento de las clases de almacenamiento de Amazon S3. Una vez que se configure en el bucket, persistirá para siempre.
- Cree y configure una subred con un [rango de direcciones IP suficiente](#) dentro de la nube privada virtual (VPC) en la que se despliega la base de datos. Esto evitará errores en las tareas de AWS Glue causados por un número insuficiente de direcciones de red cuando se configure una gran cantidad de unidades de procesamiento de datos (DPUs).

Acceso a los datos archivados en Amazon S3

Amazon S3 proporciona una serie de herramientas para leer el contenido de los datos. Sin embargo, según la clase de almacenamiento, es posible que se requieran algunos pasos de preprocesamiento. Esta sección incluye lo siguiente:

- Lectura de objetos S3 archivados con clase de almacenamiento estándar, mediante AWS Glue o Amazon S3 Select
- Lectura de un objeto S3 archivado con las clases de almacenamiento de S3 Glacier mediante Amazon S3 Glacier Select o S3 Batch Operations
- Prácticas recomendadas

Lectura de objetos S3 archivados con clase de almacenamiento estándar

Puede leer los objetos de S3 archivados con la clase de almacenamiento Amazon S3 mediante AWS Glue o Amazon S3 Select.

Uso de AWS Glue

Los datos descargados de MySQL a Amazon S3 conservan la misma rigidez estructural y la misma consistencia típicas de un sistema de administración de bases de datos relacionales (RDBMS).

[Rastreador AWS Glue](#) rastrea los objetos de S3, deduce los tipos de datos y crea los metadatos de la tabla como un DDL de tabla externo. Al configurar el trabajo de rastreador, utilice Amazon S3 como origen y especifique la ubicación del prefijo S3 en la que se crean todos los archivos de datos. En la configuración, incluya lo siguiente:

- Opciones de ejecución del rastreador
- Preferencia de prefijo de tabla opcional
- Base de datos de destino para crear la tabla
- Funciones de IAM con los permisos necesarios

Tras invocar el trabajo, analizará los datos para deducir el esquema y conservarlo en [Catálogo de datos de AWS Glue](#) en [Tablas AWS Glue](#). Las tablas de AWS Glue son básicamente tablas externas

que se pueden consultar con sentencias SQL, como en una tabla de base de datos normal, mediante servicios de análisis como [Amazon Atenea](#), [Amazon Redshift Spectrum](#), y Apache Hive en [Amazon EMR](#). Para obtener más información sobre el rastreador, consulte [Documentación de AWS Glue](#).

En el caso de los archivos.csv con un encabezado de columna especificado, los nombres de las columnas de la tabla resultante reflejarán los mismos nombres de campo. El tipo de datos se deduce en función de los valores del objeto de datos.

En el caso de los archivos Parquet, el esquema se conserva dentro de los propios datos y la tabla resultante reflejará los mismos nombres de campo y tipos de datos.

Como alternativa, puede ejecutar un DDL manualmente en Athena para crear la definición de la tabla con los nombres de columna y el tipo de datos necesarios. Esto crea la definición de la tabla en el catálogo de datos. Para obtener más información sobre la creación de tablas de Athena, consulte [Documentación de Amazon Athena](#).

Nota: Si falta la fila del encabezado en el archivo CSV, el rastreador crea el nombre del campo como c_0, c_1, c_2,...

Uso de Amazon S3 Select

Puede utilizar Amazon S3 Select para leer los objetos de S3 mediante programación mediante expresiones SQL. La operación de la API se puede invocar mediante el comando `AWS CLI select-object-content` mediante un SDK como Boto3 e invocando la operación `select_object_content` de Python.

Las operaciones de la API admiten sentencias SQL como parámetros y solo pueden leer archivos de tipo JSON y Parquet. Las salidas se pueden redirigir como archivos de salida.

Estas operaciones se invocan para cada objeto de S3. Para varios archivos, ejecute las operaciones de forma recursiva.

Para obtener más información sobre la ejecución de las operaciones mediante la CLI de AWS, consulte la [Documentación de AWS CLI](#). Para obtener más información sobre la ejecución de S3 Select mediante el Boto3 del SDK de Python, consulte [Documentación de Boto3](#).

Lectura de objetos S3 archivados con clases de almacenamiento de S3 Glacier

Las clases de Amazon S3 Glacier son clases de almacenamiento especiales con precios económicos pero un tiempo de recuperación elevado. A diferencia de los objetos estándar de S3, los objetos de S3 Glacier no se pueden leer como tablas de AWS Glue. Para que los datos estén disponibles para consultas analíticas o informes, primero debe restaurar los objetos de S3 Glacier. La restauración es un proceso asíncrono que se produce a lo largo del tiempo y tiene un período de retención. Una vez restaurados los objetos, se pueden copiar a una ubicación diferente como objetos estándar de S3. Transcurrido el período de retención, los objetos restaurados vuelven a Amazon S3 Glacier.

Uso de Amazon S3 Glacier Select

Al igual que con Amazon S3 Select con S3 Standard, puede consultar objetos de S3 Glacier para obtener un subconjunto de datos. Esto permite el acceso programático a los datos sin necesidad de ningún procesamiento previo, como la restauración de objetos u otros servicios analíticos de AWS. Para ver un ejemplo, consulte [sintaxis de solicitud](#) para usar el `Boto3initiate_job` operación para leer datos de S3 Glacier Select.

Uso de operaciones por lotes de S3

S3 Batch Operations permite operaciones por lotes a gran escala en Amazon S3 del orden de miles de millones de objetos que contienen exabytes de datos. Amazon S3 realiza un seguimiento del avance, envía notificaciones y guarda un informe de finalización de todas las acciones, por lo que proporciona una experiencia sin servidor, auditable y completamente administrada.

S3 Batch Operations es compatible con [Restaurar](#) operación, que inicia la restauración de objetos de S3 para los siguientes niveles de almacenamiento:

- Objetos archivados en las clases de almacenamiento S3 Glacier Flexible Retrieval o S3 Glacier Deep Archive
- Objetos archivados a través de la clase de almacenamiento S3 Intelligent-Tiering en los niveles de acceso a archivos o acceso profundo a archivos

La operación por lotes se puede invocar tanto mediante programación como en la consola Amazon S3. Para la entrada, necesita un archivo de manifiesto .csv que contenga los objetos de la lista que se van a restaurar.

Puede utilizar un [Inventario de Amazon S3](#) informe como entrada para el trabajo por lotes. El informe de inventario está configurado para un bucket y puede limitarse a objetos con prefijos específicos. Es un informe automatizado y se genera semanal o diariamente en formato CSV, ORC o Parquet.

Para obtener más información sobre la configuración de un informe de inventario, consulte [Documentación de Amazon S3](#). Para obtener información sobre el uso de Boto3 para crear un trabajo de operaciones por lotes de S3, consulte [Documentación de Boto3](#).

Prácticas recomendadas

Recomendamos las siguientes prácticas recomendadas para acceder a los datos archivados:

- Amazon S3 Select y Amazon S3 Glacier Select son adecuados para los casos en los que se obtiene un subconjunto de datos mediante expresiones SQL básicas. Las aplicaciones que deben consultar Amazon S3 mediante programación para obtener conjuntos de datos seleccionados muestran una mejora notable en el rendimiento, en muchos casos, hasta un 400 por ciento. Para obtener más información, consulte la entrada del blog [S3 Select y S3 Glacier Select: recuperación de subconjuntos de objetos](#).
- S3 Select y S3 Glacier Select solo admiten la lectura de datos en formatos CSV, JSON y Parquet. Para otros formatos abiertos, como ORC o Avro, puede utilizar Amazon Athena o Amazon Redshift.
- Para conjuntos de datos archivados de gran tamaño, recomendamos crear tablas de AWS Glue sobre los datos para que puedan leerse mediante motores de consulta como Athena y Amazon Redshift. Tanto Athena como Amazon Redshift ofrecen una escala horizontal del rendimiento de las consultas. También utilizan un pay-per-query modelo, que es rentable en un escenario de consulta única. Además, Amazon Redshift cuenta con motores Advanced Query Accelerator (AQUA) integrados, que aceleran el rendimiento de lectura sin coste adicional.
- Los datos archivados que se descargan regularmente en Amazon S3 no deben almacenarse como un vertedero. En su lugar, debe guardarse como una partición nueva. Una partición de fechas separará los datos en dimensiones de fecha (por ejemplo, `year=<value>/month=<value>/day=<value>`). Esto es extremadamente beneficioso en dos situaciones:
 - Si las tablas de AWS Glue las crean los rastreadores de AWS Glue, estas particiones actúan como pseudocolumnas. Esto mejora el rendimiento de lectura al restringir los datos escaneados a las particiones de la consulta de rango.
 - Esto ayuda en una operación de restauración de S3 Glacier cuando se restaura solo un subconjunto del objeto como estándar de S3.

- Los rastreadores de AWS Glue muestran un gran valor cuando los datos archivados guardados en Amazon S3 se particionan físicamente. Cada vez que los datos se descargan como una nueva partición de prefijo, el rastreador escanea solo la nueva partición y actualiza los metadatos de esa partición. Si el esquema de la tabla cambia, esos cambios se capturarán en los metadatos a nivel de partición.

Limpieza de tablas de archivado

La etapa final del proceso de archivado consiste en limpiar las tablas del esquema de archivado. Puede hacerlo después de confirmar que los datos archivados están archivados de forma segura en Amazon S3. Para evitar cualquier impacto en la aplicación, se recomienda eliminar las tablas del esquema de archivado durante un período de inactividad programado o durante un período de mantenimiento o durante un período de muy poco tráfico en la aplicación. La aplicación no consulta estas tablas de forma activa y no debería ser motivo de alarma por su impacto en las transacciones en curso. Aun así, se recomienda ejecutar los DDL durante un tiempo de inactividad.

Una vez liberado el almacenamiento para tablas de esquemas de archivo de gran tamaño, Amazon Aurora utiliza [redimensionamiento dinámico](#) para ayudarle a ahorrar en costes de almacenamiento.

Recursos

Documentación

- [Amazon Aurora](#)
- [AWS Data Pipeline](#)
- [AWS Glue](#)
- [Amazon S3](#)
- [Consultar el inventario de Amazon S3 con Amazon Athena](#)
- [Configuración del inventario de Amazon S3](#)
- [Realizar operaciones por lotes a gran escala en objetos de Amazon S3](#)
- [Documentación de Boto3](#)
- [Documentación sobre particiones de MySQL](#)

Precios

- [Precios de Amazon RDS para MySQL Multi-AZ GP2](#)
- [Precios de Amazon S3 Glacier](#)

Herramientas

- [Kit de herramientas Percona](#)
- [pt-archiver](#)
- [sysbench](#)

Publicaciones de blog

- [S3 Select y S3 Glacier Select: recuperación de subconjuntos de objetos](#)
- [Archive y purgue datos para Amazon RDS para PostgreSQL y Amazon Aurora con compatibilidad con PostgreSQL mediante pg_partman y Amazon S3](#)

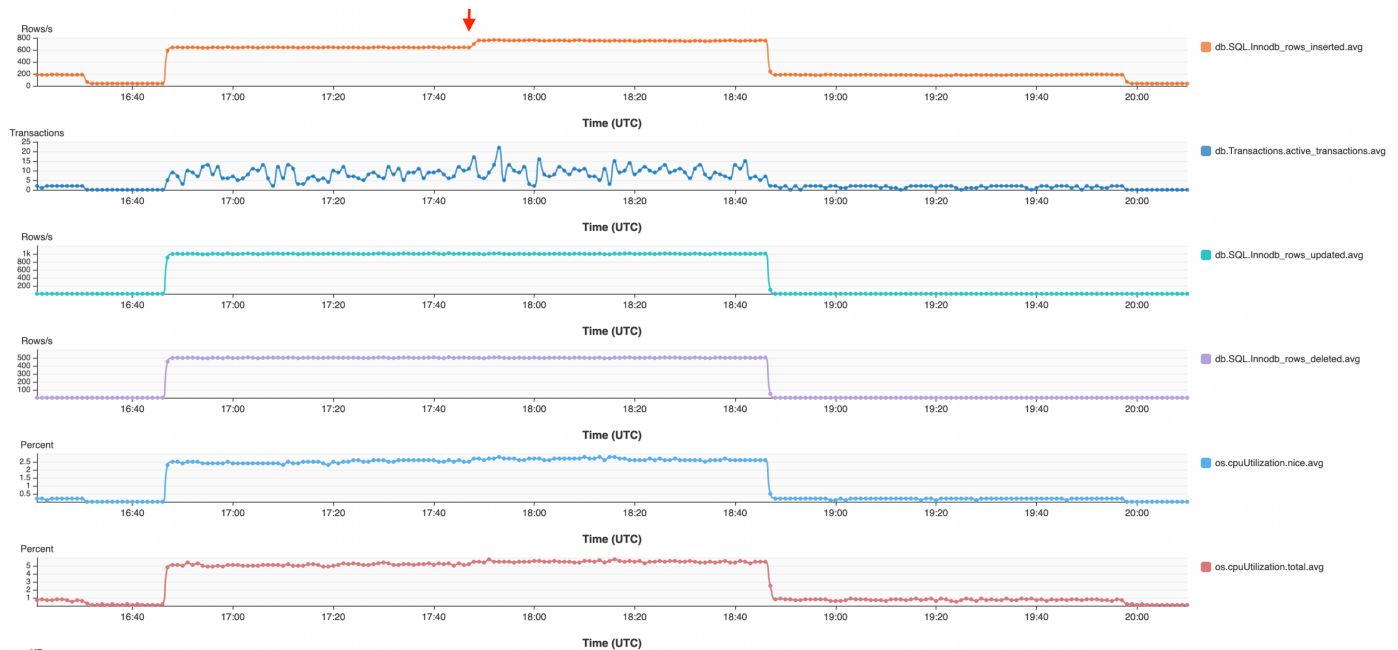
Apéndice I

Todas las pruebas se realizan en una instancia de Amazon RDS para MySQL que se ejecuta en `endb.r6g.8xlarge` clase de instancia.

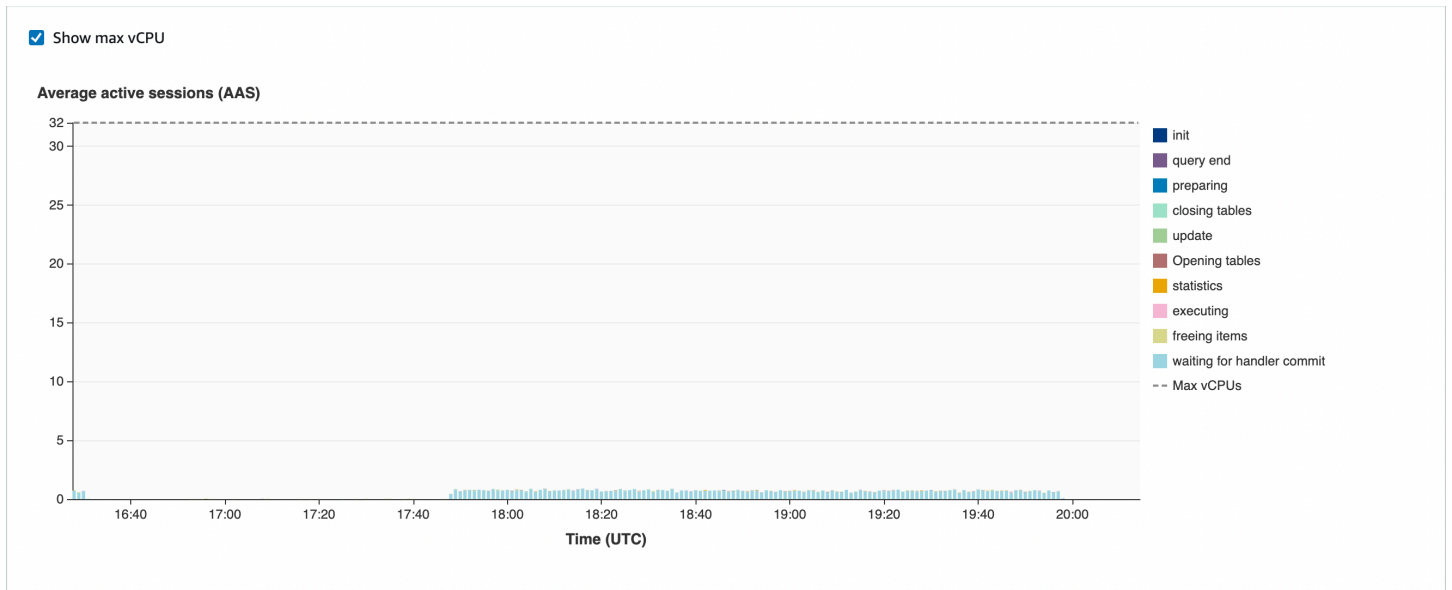
Los siguientes [sysbench](#) se usaron comandos para preparar y ejecutar la carga en la base de datos.

```
sysbench oltp_read_write --db-driver=mysql --mysql-db=<DATABASE> --mysql-user=<USER> --mysql-password=<PASSWORD> --mysql-host=<ENDPOINT> --tables=500 --table-size=2000000 --threads=500 prepare
sysbench oltp_read_write --db-driver=mysql --mysql-db=employees --mysql-user=admin --mysql-password=qwertyuiop --mysql-host=mysql8.cbbhujzeoxed.us-east-1.rds.amazonaws.com --tables=500 --rate=500 --time=7200 run
```

En el siguiente gráfico, se estaba ejecutando una carga de trabajo de OLTP y el proceso `pt-archiver` se inició donde está marcada la flecha.



No hay ningún cambio significativo en la utilización de la CPU, ya que `pt-archiver` se ejecuta en paralelo, lo que deduce que `pt-archiver` no afecta a las consultas OLTP mientras se ejecuta.



Apéndice II

Esta sección proporciona los resultados de la evaluación comparativa de las herramientas pt-archiver en diferentes escenarios. El [sysbench](#)La herramienta se utiliza en esta prueba para cargar la base de datos. Todas las pruebas se realizan en la instancia de Amazon RDS para MySQL que se ejecuta en `db.r6g.8xlarge` clase de instancia.

Se usaron los siguientes comandos de sysbench para preparar y ejecutar la carga en la base de datos:

```
sysbench oltp_read_write --db-driver=mysql --mysql-db=<DATABASE> --mysql-user=<USER> --mysql-password=<PASSWORD> --mysql-host=<ENDPOINT> --tables=1000 --table-size=2000000 --threads=500 prepare
sysbench oltp_read_write --db-driver=mysql --mysql-db=<DATABASE> --mysql-user=<USER> --mysql-password=<PASSWORD> --mysql-host=<ENDPOINT> --tables=1000 --rate=500 --threads=500 run
```

Archivar una tabla que no tiene una clave principal y solo un índice (sin carga en la base de datos)

```
Started at 2022-11-07T05:29:12, ended at 2022-11-07T06:03:31
Action      Count          Time           Pct
commit     600050         1715.3582     83.31
select     300025         166.5470      8.09
inserting  300024         165.4025      8.03
other       0              11.6644       0.57
```

Se tardaron unos 34 minutos en archivar 300.024 filas. Esta tabla tenía 2 millones de filas, pero la herramienta archivaba solo las filas con datos exclusivos de la columna indexada.

Archivar una tabla que tiene una clave principal (sin carga en la base de datos)

```
Started at 2022-11-16T08:53:49, ended at 2022-11-16T12:38:18
Action      Count          Time           Pct
commit     4000000        11065.9534    82.16
select     2000000        1278.1854     9.49
inserting  1999999        1050.4961     7.80
other       0              74.1519       0.55
```

Se necesitaron alrededor de 3 horas, 44 minutos y 29 segundos para archivar 1.999.999 filas.

El siguiente gráfico muestra que pt-archiver consume muy poca CPU y recursos cuando se ejecuta solo sin que exista ninguna carga en el sistema.

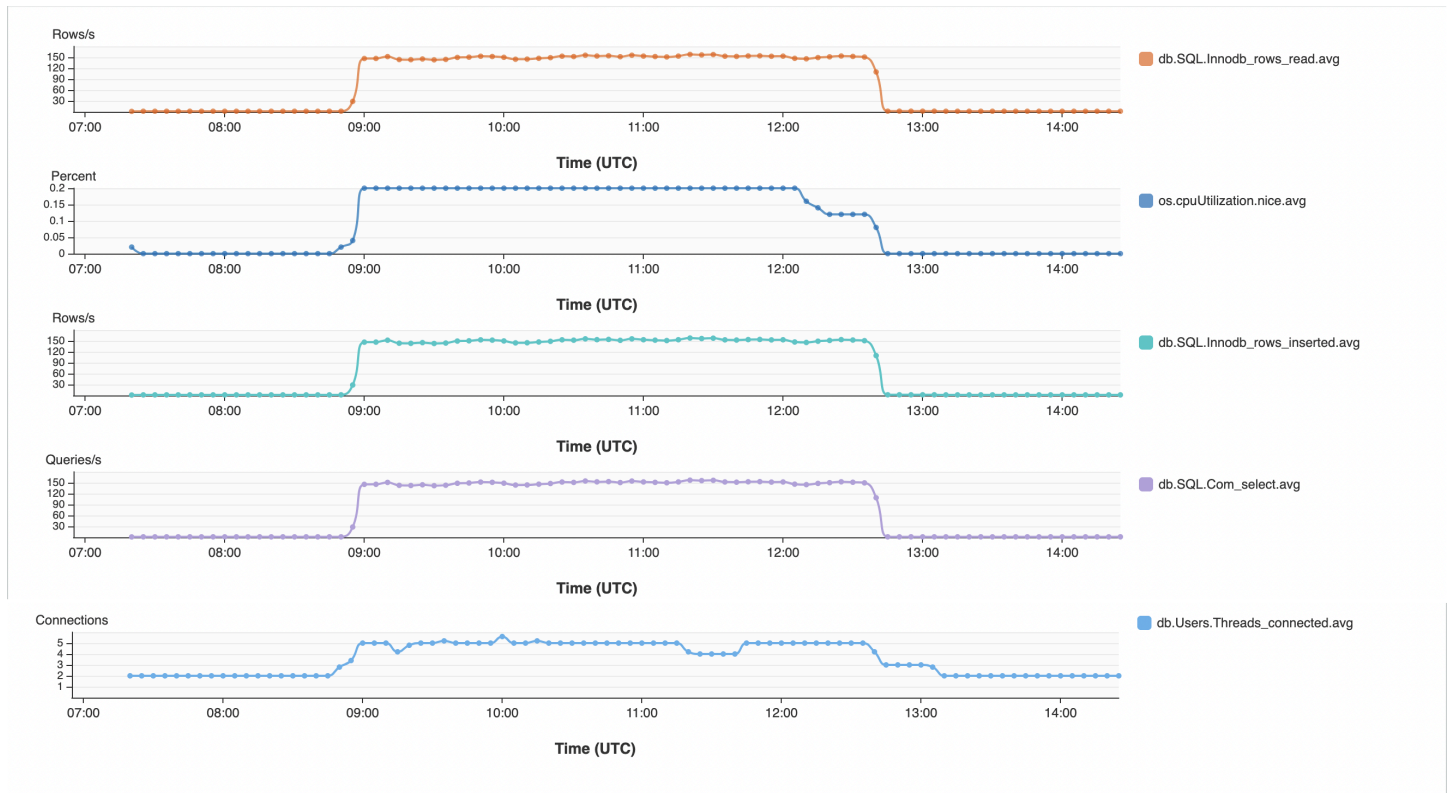


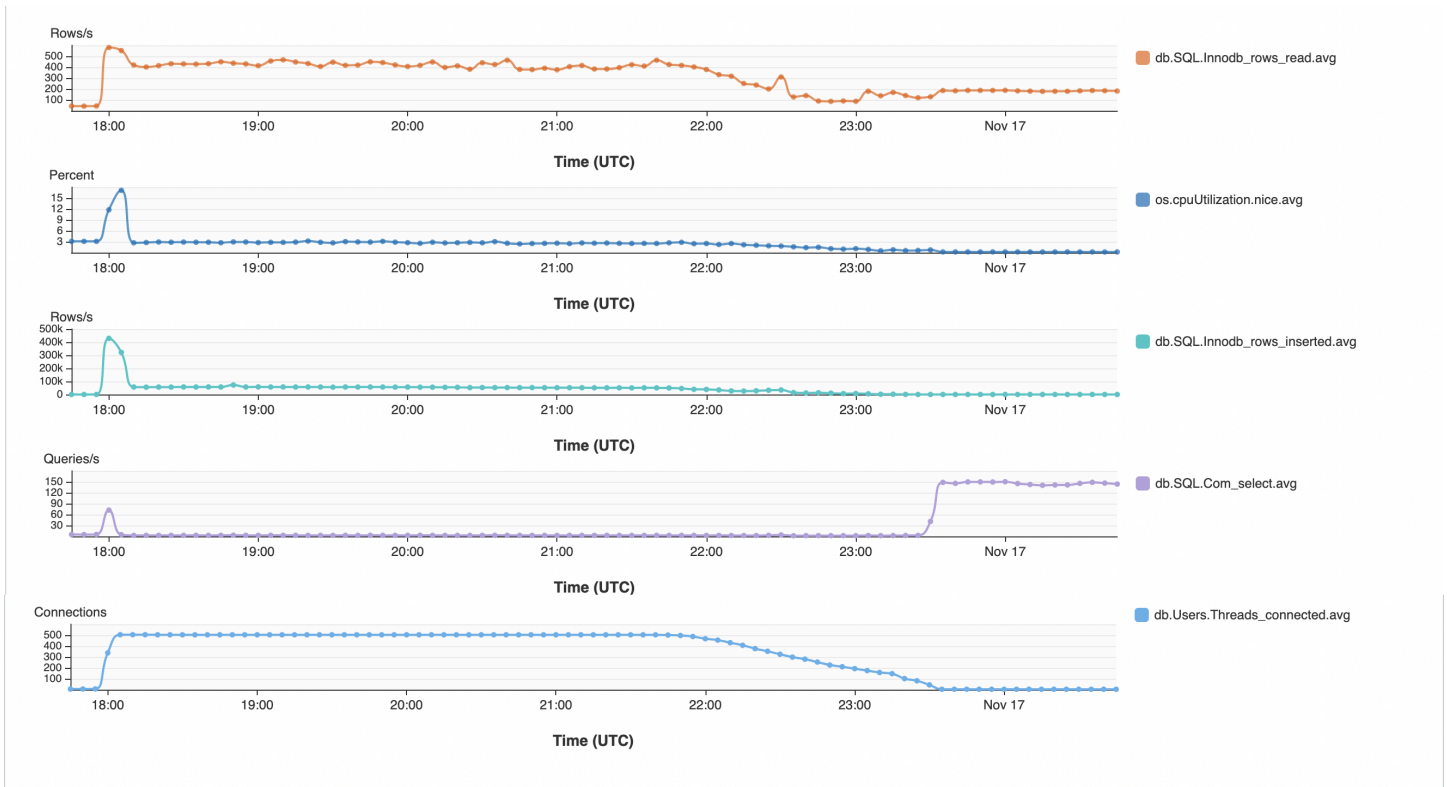
Tabla de archivado que tiene una clave principal (con carga en la base de datos)

Started at 2022-11-16T17:37:07, ended at 2022-11-17T03:20:43

Action	Count	Time	Pct
commit	4000000	19688.8362	56.23
inserting	1999999	13933.4418	39.79
select	2000000	1305.1770	3.73
other	0	89.1787	0.25

Se necesitaron alrededor de 9 horas, 43 minutos y 36 segundos para archivar 1999999 filas.

El siguiente gráfico muestra que, durante la prueba, la utilización de la CPU fue de hasta un 15 por ciento debido a la carga aplicada por sysbench. Una vez completada la carga, pt-archiver siguió funcionando consumiendo una cantidad mínima de CPU, tal como se esperaba para completar el archivado.



Como se desprende de los gráficos, pt-archiver no archiva de forma agresiva cuando hay una carga en la base de datos.

Historial de documentos

En la siguiente tabla se describen los cambios importantes de esta guía. Si quieres recibir notificaciones sobre futuras actualizaciones, puedes suscribirte a un [Fuente RSS](#).

Cambio	Descripción	Fecha
Publicación inicial	—	8 de junio de 2023

AWS Glosario de orientación prescriptiva

Los siguientes son términos de uso común en las estrategias, guías y patrones proporcionados por AWS Prescriptive Guidance. Para sugerir entradas, utilice el enlace [Enviar comentarios](#) al final del glosario.

Números

Las 7 R

Siete estrategias de migración comunes para trasladar aplicaciones a la nube. Estas estrategias se basan en las 5 R que Gartner identificó en 2011 y consisten en lo siguiente:

- **Refactorizar/rediseñar:** traslade una aplicación y modifique su arquitectura mediante el máximo aprovechamiento de las características nativas en la nube para mejorar la agilidad, el rendimiento y la escalabilidad. Por lo general, esto implica trasladar el sistema operativo y la base de datos. Ejemplo: migre su base de datos Oracle local a la edición compatible con PostgreSQL de Amazon Aurora.
- **Redefinir la plataforma (transportar y redefinir):** traslade una aplicación a la nube e introduzca algún nivel de optimización para aprovechar las capacidades de la nube. Ejemplo: migre su base de datos Oracle local a Amazon Relational Database Service (Amazon RDS) para Oracle en el. Nube de AWS
- **Recomprar (readquirir):** cambie a un producto diferente, lo cual se suele llevar a cabo al pasar de una licencia tradicional a un modelo SaaS. Ejemplo: migre su sistema de gestión de relaciones con los clientes (CRM) a Salesforce.com.
- **Volver a alojar (migrar mediante lift-and-shift):** traslade una aplicación a la nube sin realizar cambios para aprovechar las capacidades de la nube. Ejemplo: migre su base de datos Oracle local a Oracle en una instancia EC2 del. Nube de AWS
- **Reubicar:** (migrar el hipervisor mediante lift and shift): traslade la infraestructura a la nube sin comprar equipo nuevo, reescribir aplicaciones o modificar las operaciones actuales. Los servidores se migran de una plataforma local a un servicio en la nube para la misma plataforma. Ejemplo: migrar una Microsoft Hyper-V aplicación a AWS.
- **Retener (revisitar):** conserve las aplicaciones en el entorno de origen. Estas pueden incluir las aplicaciones que requieren una refactorización importante, que desee posponer para más adelante, y las aplicaciones heredadas que desee retener, ya que no hay ninguna justificación empresarial para migrarlas.

- Retirar: retire o elimine las aplicaciones que ya no sean necesarias en un entorno de origen.

A

ABAC

Consulte control de [acceso basado en atributos](#).

servicios abstractos

Consulte [servicios gestionados](#).

ACID

Consulte [atomicidad, consistencia, aislamiento y durabilidad](#).

migración activa-activa

Método de migración de bases de datos en el que las bases de datos de origen y destino se mantienen sincronizadas (mediante una herramienta de replicación bidireccional o mediante operaciones de escritura doble) y ambas bases de datos gestionan las transacciones de las aplicaciones conectadas durante la migración. Este método permite la migración en lotes pequeños y controlados, en lugar de requerir una transición única. Es más flexible, pero requiere más trabajo que la migración [activa-pasiva](#).

migración activa-pasiva

Método de migración de bases de datos en el que las bases de datos de origen y destino se mantienen sincronizadas, pero solo la base de datos de origen gestiona las transacciones de las aplicaciones conectadas, mientras los datos se replican en la base de datos de destino. La base de datos de destino no acepta ninguna transacción durante la migración.

función agregada

Función SQL que opera en un grupo de filas y calcula un único valor de retorno para el grupo. Entre los ejemplos de funciones agregadas se incluyen SUM y MAX.

IA

Véase [inteligencia artificial](#).

AIOps

Consulte las [operaciones de inteligencia artificial](#).

anonimización

El proceso de eliminar permanentemente la información personal de un conjunto de datos. La anonimización puede ayudar a proteger la privacidad personal. Los datos anonimizados ya no se consideran datos personales.

antipatronos

Una solución que se utiliza con frecuencia para un problema recurrente en el que la solución es contraproducente, ineficaz o menos eficaz que una alternativa.

control de aplicaciones

Un enfoque de seguridad que permite el uso únicamente de aplicaciones aprobadas para ayudar a proteger un sistema contra el malware.

cartera de aplicaciones

Recopilación de información detallada sobre cada aplicación que utiliza una organización, incluido el costo de creación y mantenimiento de la aplicación y su valor empresarial. Esta información es clave para [el proceso de detección y análisis de la cartera](#) y ayuda a identificar y priorizar las aplicaciones que se van a migrar, modernizar y optimizar.

inteligencia artificial (IA)

El campo de la informática que se dedica al uso de tecnologías informáticas para realizar funciones cognitivas que suelen estar asociadas a los seres humanos, como el aprendizaje, la resolución de problemas y el reconocimiento de patrones. Para más información, consulte [¿Qué es la inteligencia artificial?](#)

operaciones de inteligencia artificial (AIOps)

El proceso de utilizar técnicas de machine learning para resolver problemas operativos, reducir los incidentes operativos y la intervención humana, y mejorar la calidad del servicio. Para obtener más información sobre cómo se utiliza AIOps en la estrategia de migración de AWS, consulte la [Guía de integración de operaciones](#).

cifrado asimétrico

Algoritmo de cifrado que utiliza un par de claves, una clave pública para el cifrado y una clave privada para el descifrado. Puede compartir la clave pública porque no se utiliza para el descifrado, pero el acceso a la clave privada debe estar sumamente restringido.

atomicidad, consistencia, aislamiento, durabilidad (ACID)

Conjunto de propiedades de software que garantizan la validez de los datos y la fiabilidad operativa de una base de datos, incluso en caso de errores, cortes de energía u otros problemas.

control de acceso basado en atributos (ABAC)

La práctica de crear permisos detallados basados en los atributos del usuario, como el departamento, el puesto de trabajo y el nombre del equipo. Para obtener más información, consulte [ABAC AWS en la](#) documentación AWS Identity and Access Management (IAM).

origen de datos fidedigno

Ubicación en la que se almacena la versión principal de los datos, que se considera la fuente de información más fiable. Puede copiar los datos del origen de datos autorizado a otras ubicaciones con el fin de procesarlos o modificarlos, por ejemplo, anonimizarlos, redactarlos o seudonimizarlos.

Zona de disponibilidad

Una ubicación distinta dentro de una Región de AWS que está aislada de los fallos en otras zonas de disponibilidad y que proporciona una conectividad de red económica y de baja latencia a otras zonas de disponibilidad de la misma región.

AWS Marco de adopción de la nube (AWS CAF)

Un marco de directrices y mejores prácticas AWS para ayudar a las organizaciones a desarrollar un plan eficiente y eficaz para migrar con éxito a la nube. AWS CAF organiza la orientación en seis áreas de enfoque denominadas perspectivas: negocios, personas, gobierno, plataforma, seguridad y operaciones. Las perspectivas empresariales, humanas y de gobernanza se centran en las habilidades y los procesos empresariales; las perspectivas de plataforma, seguridad y operaciones se centran en las habilidades y los procesos técnicos. Por ejemplo, la perspectiva humana se dirige a las partes interesadas que se ocupan de los Recursos Humanos (RR. HH.), las funciones del personal y la administración de las personas. Desde esta perspectiva, AWS CAF proporciona orientación para el desarrollo, la formación y la comunicación de las personas a fin de preparar a la organización para una adopción exitosa de la nube. Para obtener más información, consulte la [Página web de AWS CAF](#) y el [Documento técnico de AWS CAF](#).

AWS Marco de calificación de la carga de trabajo (AWS WQF)

Herramienta que evalúa las cargas de trabajo de migración de bases de datos, recomienda estrategias de migración y proporciona estimaciones de trabajo. AWS WQF se incluye con AWS

Schema Conversion Tool (). AWS SCT Analiza los esquemas de bases de datos y los objetos de código, el código de las aplicaciones, las dependencias y las características de rendimiento y proporciona informes de evaluación.

B

Un bot malo

Un [bot](#) destinado a interrumpir o causar daño a personas u organizaciones.

BCP

Consulte la [planificación de la continuidad del negocio](#).

gráfico de comportamiento

Una vista unificada e interactiva del comportamiento de los recursos y de las interacciones a lo largo del tiempo. Puede utilizar un gráfico de comportamiento con Amazon Detective para examinar los intentos de inicio de sesión fallidos, las llamadas sospechosas a la API y acciones similares. Para obtener más información, consulte [Datos en un gráfico de comportamiento](#) en la documentación de Detective.

sistema big-endian

Un sistema que almacena primero el byte más significativo. Véase también [endianness](#).

clasificación binaria

Un proceso que predice un resultado binario (una de las dos clases posibles). Por ejemplo, es posible que su modelo de ML necesite predecir problemas como “¿Este correo electrónico es spam o no es spam?” o “¿Este producto es un libro o un automóvil?”.

filtro de floración

Estructura de datos probabilística y eficiente en términos de memoria que se utiliza para comprobar si un elemento es miembro de un conjunto.

implementación azul/verde

Una estrategia de despliegue en la que se crean dos entornos separados pero idénticos. La versión actual de la aplicación se ejecuta en un entorno (azul) y la nueva versión de la aplicación en el otro entorno (verde). Esta estrategia le ayuda a revertirla rápidamente con un impacto mínimo.

bot

Aplicación de software que ejecuta tareas automatizadas a través de Internet y simula la actividad o interacción humana. Algunos bots son útiles o beneficiosos, como los rastreadores web que indexan información en Internet. Algunos otros bots, conocidos como bots malos, tienen como objetivo interrumpir o causar daños a personas u organizaciones.

botnet

Redes de [bots](#) que están infectadas por [malware](#) y que están bajo el control de una sola parte, conocida como pastor u operador de bots. Las botnets son el mecanismo más conocido para escalar los bots y su impacto.

rama

Área contenida de un repositorio de código. La primera rama que se crea en un repositorio es la rama principal. Puede crear una rama nueva a partir de una rama existente y, a continuación, desarrollar características o corregir errores en la rama nueva. Una rama que se genera para crear una característica se denomina comúnmente rama de característica. Cuando la característica se encuentra lista para su lanzamiento, se vuelve a combinar la rama de característica con la rama principal. Para obtener más información, consulte [Acerca de las sucursales](#) (GitHub documentación).

acceso con cristales rotos

En circunstancias excepcionales y mediante un proceso aprobado, un usuario puede acceder rápidamente a un sitio para el Cuenta de AWS que normalmente no tiene permisos de acceso. Para obtener más información, consulte el indicador [Implemente procedimientos de rotura de cristales en la guía Well-Architected AWS](#) .

estrategia de implementación sobre infraestructura existente

La infraestructura existente en su entorno. Al adoptar una estrategia de implementación sobre infraestructura existente para una arquitectura de sistemas, se diseña la arquitectura en función de las limitaciones de los sistemas y la infraestructura actuales. Si está ampliando la infraestructura existente, puede combinar las estrategias de implementación sobre infraestructuras existentes y de [implementación desde cero](#).

caché de búfer

El área de memoria donde se almacenan los datos a los que se accede con más frecuencia.

capacidad empresarial

Lo que hace una empresa para generar valor (por ejemplo, ventas, servicio al cliente o marketing). Las arquitecturas de microservicios y las decisiones de desarrollo pueden estar impulsadas por las capacidades empresariales. Para obtener más información, consulte la sección [Organizado en torno a las capacidades empresariales](#) del documento técnico [Ejecutar microservicios en contenedores en AWS](#).

planificación de la continuidad del negocio (BCP)

Plan que aborda el posible impacto de un evento disruptivo, como una migración a gran escala en las operaciones y permite a la empresa reanudar las operaciones rápidamente.

C

CAF

[Consulte el marco AWS de adopción de la nube.](#)

despliegue canario

El lanzamiento lento e incremental de una versión para los usuarios finales. Cuando se tiene confianza, se despliega la nueva versión y se reemplaza la versión actual en su totalidad.

CCoE

Consulte el [Centro de excelencia en la nube](#).

CDC

Consulte la [captura de datos de cambios](#).

captura de datos de cambio (CDC)

Proceso de seguimiento de los cambios en un origen de datos, como una tabla de base de datos, y registro de los metadatos relacionados con el cambio. Puede utilizar los CDC para diversos fines, como auditar o replicar los cambios en un sistema de destino para mantener la sincronización.

ingeniería del caos

Introducir intencionalmente fallos o eventos disruptivos para poner a prueba la resiliencia de un sistema. Puedes usar [AWS Fault Injection Service \(AWS FIS\)](#) para realizar experimentos que estresen tus AWS cargas de trabajo y evalúen su respuesta.

CI/CD

Consulte la [integración continua y la entrega continua](#).

clasificación

Un proceso de categorización que permite generar predicciones. Los modelos de ML para problemas de clasificación predicen un valor discreto. Los valores discretos siempre son distintos entre sí. Por ejemplo, es posible que un modelo necesite evaluar si hay o no un automóvil en una imagen.

cifrado del cliente

Cifrado de datos localmente, antes de que el objetivo los Servicio de AWS reciba.

Centro de excelencia en la nube (CCoE)

Equipo multidisciplinario que impulsa los esfuerzos de adopción de la nube en toda la organización, incluido el desarrollo de las prácticas recomendadas en la nube, la movilización de recursos, el establecimiento de plazos de migración y la dirección de la organización durante las transformaciones a gran escala. Para obtener más información, consulte las [publicaciones de CCoE](#) en el blog de estrategia Nube de AWS empresarial.

computación en la nube

La tecnología en la nube que se utiliza normalmente para la administración de dispositivos de IoT y el almacenamiento de datos de forma remota. La computación en la nube suele estar conectada a la tecnología de [computación perimetral](#).

modelo operativo en la nube

En una organización de TI, el modelo operativo que se utiliza para crear, madurar y optimizar uno o más entornos de nube. Para obtener más información, consulte [Creación de su modelo operativo de nube](#).

etapas de adopción de la nube

Las cuatro fases por las que suelen pasar las organizaciones cuando migran a Nube de AWS:

- Proyecto: ejecución de algunos proyectos relacionados con la nube con fines de prueba de concepto y aprendizaje
- Fundamento: realización de inversiones fundamentales para escalar la adopción de la nube (p. ej., crear una zona de aterrizaje, definir un CCoE, establecer un modelo de operaciones)
- Migración: migración de aplicaciones individuales

- Reinención: optimización de productos y servicios e innovación en la nube

Stephen Orban definió estas etapas en la entrada del blog [The Journey Toward Cloud-First & the Stages of Adoption del](#) blog Nube de AWS Enterprise Strategy. Para obtener información sobre su relación con la estrategia de AWS migración, consulte la guía de [preparación para la migración](#).

CMDB

Consulte la [base de datos de administración de la configuración](#).

repositorio de código

Una ubicación donde el código fuente y otros activos, como documentación, muestras y scripts, se almacenan y actualizan mediante procesos de control de versiones. Los repositorios en la nube más comunes incluyen GitHub o AWS CodeCommit. Cada versión del código se denomina rama. En una estructura de microservicios, cada repositorio se encuentra dedicado a una única funcionalidad. Una sola canalización de CI/CD puede utilizar varios repositorios.

caché en frío

Una caché de búfer que está vacía no está bien poblada o contiene datos obsoletos o irrelevantes. Esto afecta al rendimiento, ya que la instancia de la base de datos debe leer desde la memoria principal o el disco, lo que es más lento que leer desde la memoria caché del búfer.

datos fríos

Datos a los que se accede con poca frecuencia y que suelen ser históricos. Al consultar este tipo de datos, normalmente se aceptan consultas lentas. Trasladar estos datos a niveles o clases de almacenamiento de menor rendimiento y menos costosos puede reducir los costos.

visión artificial (CV)

Campo de la [IA](#) que utiliza el aprendizaje automático para analizar y extraer información de formatos visuales, como imágenes y vídeos digitales. Por ejemplo, AWS Panorama ofrece dispositivos que añaden CV a las redes de cámaras locales, y Amazon SageMaker proporciona algoritmos de procesamiento de imágenes para CV.

desviación de configuración

En el caso de una carga de trabajo, un cambio de configuración con respecto al estado esperado. Puede provocar que la carga de trabajo deje de cumplir las normas y, por lo general, es gradual e involuntario.

base de datos de administración de configuración (CMDB)

Repositorio que almacena y administra información sobre una base de datos y su entorno de TI, incluidos los componentes de hardware y software y sus configuraciones. Por lo general, los datos de una CMDB se utilizan en la etapa de detección y análisis de la cartera de productos durante la migración.

paquete de conformidad

Conjunto de AWS Config reglas y medidas correctivas que puede reunir para personalizar sus comprobaciones de conformidad y seguridad. Puede implementar un paquete de conformidad como una entidad única en una región Cuenta de AWS y, o en una organización, mediante una plantilla YAML. Para obtener más información, consulta los [paquetes de conformidad](#) en la documentación. AWS Config

integración y entrega continuas (CI/CD)

El proceso de automatización de las etapas de origen, compilación, prueba, presentación y producción del proceso de lanzamiento del software. La CI/CD se describe comúnmente como una canalización. La CI/CD puede ayudarlo a automatizar los procesos, mejorar la productividad, mejorar la calidad del código y entregar con mayor rapidez. Para obtener más información, consulte [Beneficios de la entrega continua](#). CD también puede significar implementación continua. Para obtener más información, consulte [Entrega continua frente a implementación continua](#).

CV

Consulte [visión artificial](#).

D

datos en reposo

Datos que están estacionarios en la red, como los datos que se encuentran almacenados.

clasificación de datos

Un proceso para identificar y clasificar los datos de su red en función de su importancia y sensibilidad. Es un componente fundamental de cualquier estrategia de administración de riesgos de ciberseguridad porque lo ayuda a determinar los controles de protección y retención adecuados para los datos. La clasificación de datos es un componente del pilar de seguridad

del AWS Well-Architected Framework. Para obtener más información, consulte [Clasificación de datos](#).

desviación de datos

Una variación significativa entre los datos de producción y los datos que se utilizaron para entrenar un modelo de machine learning, o un cambio significativo en los datos de entrada a lo largo del tiempo. La desviación de los datos puede reducir la calidad, la precisión y la imparcialidad generales de las predicciones de los modelos de machine learning.

datos en tránsito

Datos que se mueven de forma activa por la red, por ejemplo, entre los recursos de la red.

malla de datos

Un marco arquitectónico que proporciona una propiedad de datos distribuida y descentralizada con administración y gobierno centralizados.

minimización de datos

El principio de recopilar y procesar solo los datos estrictamente necesarios. Practicar la minimización de los datos Nube de AWS puede reducir los riesgos de privacidad, los costos y la huella de carbono de la analítica.

perímetro de datos

Un conjunto de barreras preventivas en su AWS entorno que ayudan a garantizar que solo las identidades confiables accedan a los recursos confiables desde las redes esperadas. Para obtener más información, consulte [Crear un perímetro de datos sobre](#) AWS

preprocesamiento de datos

Transformar los datos sin procesar en un formato que su modelo de ML pueda analizar fácilmente. El preprocesamiento de datos puede implicar eliminar determinadas columnas o filas y corregir los valores faltantes, incoherentes o duplicados.

procedencia de los datos

El proceso de rastrear el origen y el historial de los datos a lo largo de su ciclo de vida, por ejemplo, la forma en que se generaron, transmitieron y almacenaron los datos.

titular de los datos

Persona cuyos datos se recopilan y procesan.

almacenamiento de datos

Un sistema de administración de datos que respalde la inteligencia empresarial, como el análisis. Los almacenes de datos suelen contener grandes cantidades de datos históricos y, por lo general, se utilizan para consultas y análisis.

lenguaje de definición de datos (DDL)

Instrucciones o comandos para crear o modificar la estructura de tablas y objetos de una base de datos.

lenguaje de manipulación de datos (DML)

Instrucciones o comandos para modificar (insertar, actualizar y eliminar) la información de una base de datos.

DDL

Consulte el [lenguaje de definición de bases](#) de datos.

conjunto profundo

Combinar varios modelos de aprendizaje profundo para la predicción. Puede utilizar conjuntos profundos para obtener una predicción más precisa o para estimar la incertidumbre de las predicciones.

aprendizaje profundo

Un subcampo del ML que utiliza múltiples capas de redes neuronales artificiales para identificar el mapeo entre los datos de entrada y las variables objetivo de interés.

defense-in-depth

Un enfoque de seguridad de la información en el que se distribuyen cuidadosamente una serie de mecanismos y controles de seguridad en una red informática para proteger la confidencialidad, la integridad y la disponibilidad de la red y de los datos que contiene. Al adoptar esta estrategia AWS, se añaden varios controles en diferentes capas de la AWS Organizations estructura para ayudar a proteger los recursos. Por ejemplo, un defense-in-depth enfoque podría combinar la autenticación multifactorial, la segmentación de la red y el cifrado.

administrador delegado

En AWS Organizations, un servicio compatible puede registrar una cuenta de AWS miembro para administrar las cuentas de la organización y gestionar los permisos de ese servicio. Esta

cuenta se denomina administrador delegado para ese servicio. Para obtener más información y una lista de servicios compatibles, consulte [Servicios que funcionan con AWS Organizations](#) en la documentación de AWS Organizations .

Implementación

El proceso de hacer que una aplicación, características nuevas o correcciones de código se encuentren disponibles en el entorno de destino. La implementación abarca implementar cambios en una base de código y, a continuación, crear y ejecutar esa base en los entornos de la aplicación.

entorno de desarrollo

Consulte [entorno](#).

control de detección

Un control de seguridad que se ha diseñado para detectar, registrar y alertar después de que se produzca un evento. Estos controles son una segunda línea de defensa, ya que lo advierten sobre los eventos de seguridad que han eludido los controles preventivos establecidos. Para obtener más información, consulte [Controles de detección](#) en Implementación de controles de seguridad en AWS.

asignación de flujos de valor para el desarrollo (DVSM)

Proceso que se utiliza para identificar y priorizar las restricciones que afectan negativamente a la velocidad y la calidad en el ciclo de vida del desarrollo de software. DVSM amplía el proceso de asignación del flujo de valor diseñado originalmente para las prácticas de fabricación ajustada. Se centra en los pasos y los equipos necesarios para crear y transferir valor a través del proceso de desarrollo de software.

gemelo digital

Representación virtual de un sistema del mundo real, como un edificio, una fábrica, un equipo industrial o una línea de producción. Los gemelos digitales son compatibles con el mantenimiento predictivo, la supervisión remota y la optimización de la producción.

tabla de dimensiones

En un [esquema en estrella](#), tabla más pequeña que contiene los atributos de datos sobre los datos cuantitativos de una tabla de hechos. Los atributos de la tabla de dimensiones suelen ser campos de texto o números discretos que se comportan como texto. Estos atributos se utilizan habitualmente para restringir consultas, filtrar y etiquetar conjuntos de resultados.

desastre

Un evento que impide que una carga de trabajo o un sistema cumplan sus objetivos empresariales en su ubicación principal de implementación. Estos eventos pueden ser desastres naturales, fallos técnicos o el resultado de acciones humanas, como una configuración incorrecta involuntaria o un ataque de malware.

recuperación de desastres (DR)

La estrategia y el proceso que se utilizan para minimizar el tiempo de inactividad y la pérdida de datos ocasionados por un [desastre](#). Para obtener más información, consulte [Recuperación ante desastres de cargas de trabajo en AWS: Recovery in the Cloud in the AWS Well-Architected Framework](#).

DML

Consulte el lenguaje de manipulación de [bases de datos](#).

diseño basado en el dominio

Un enfoque para desarrollar un sistema de software complejo mediante la conexión de sus componentes a dominios en evolución, o a los objetivos empresariales principales, a los que sirve cada componente. Este concepto lo introdujo Eric Evans en su libro, *Diseño impulsado por el dominio: abordando la complejidad en el corazón del software* (Boston: Addison-Wesley Professional, 2003). Para obtener información sobre cómo utilizar el diseño basado en dominios con el patrón de higos estranguladores, consulte [Modernización gradual de los servicios web antiguos de Microsoft ASP.NET \(ASMX\) mediante contenedores y Amazon API Gateway](#).

DR

Consulte [recuperación ante desastres](#).

detección de desviaciones

Seguimiento de las desviaciones con respecto a una configuración de referencia. Por ejemplo, puedes usarlo AWS CloudFormation para [detectar desviaciones en los recursos del sistema](#) o puedes usarlo AWS Control Tower para [detectar cambios en tu landing zone](#) que puedan afectar al cumplimiento de los requisitos de gobierno.

DVSM

Consulte [el mapeo del flujo de valor del desarrollo](#).

E

EDA

Consulte el [análisis exploratorio de datos](#).

computación en la periferia

La tecnología que aumenta la potencia de cálculo de los dispositivos inteligentes en la periferia de una red de IoT. En comparación con [la computación en nube, la computación](#) perimetral puede reducir la latencia de la comunicación y mejorar el tiempo de respuesta.

cifrado

Proceso informático que transforma datos de texto plano, legibles por humanos, en texto cifrado.

clave de cifrado

Cadena criptográfica de bits aleatorios que se genera mediante un algoritmo de cifrado. Las claves pueden variar en longitud y cada una se ha diseñado para ser impredecible y única.

endianidad

El orden en el que se almacenan los bytes en la memoria del ordenador. Los sistemas big-endianos almacenan primero el byte más significativo. Los sistemas Little-Endian almacenan primero el byte menos significativo.

punto de conexión

[Consulte el punto final del servicio](#).

servicio de punto de conexión

Servicio que puede alojar en una nube privada virtual (VPC) para compartir con otros usuarios. Puede crear un servicio de punto final AWS PrivateLink y conceder permisos a otros directores Cuentas de AWS o a AWS Identity and Access Management (IAM). Estas cuentas o entidades principales pueden conectarse a su servicio de punto de conexión de forma privada mediante la creación de puntos de conexión de VPC de interfaz. Para obtener más información, consulte [Creación de un servicio de punto de conexión](#) en la documentación de Amazon Virtual Private Cloud (Amazon VPC).

planificación de recursos empresariales (ERP)

Un sistema que automatiza y gestiona los procesos empresariales clave (como la contabilidad, el [MES](#) y la gestión de proyectos) de una empresa.

cifrado de sobre

El proceso de cifrar una clave de cifrado con otra clave de cifrado. Para obtener más información, consulte el [cifrado de sobres](#) en la documentación de AWS Key Management Service (AWS KMS).

environment

Una instancia de una aplicación en ejecución. Los siguientes son los tipos de entornos más comunes en la computación en la nube:

- entorno de desarrollo: instancia de una aplicación en ejecución que solo se encuentra disponible para el equipo principal responsable del mantenimiento de la aplicación. Los entornos de desarrollo se utilizan para probar los cambios antes de promocionarlos a los entornos superiores. Este tipo de entorno a veces se denomina entorno de prueba.
- entornos inferiores: todos los entornos de desarrollo de una aplicación, como los que se utilizan para las compilaciones y pruebas iniciales.
- entorno de producción: instancia de una aplicación en ejecución a la que pueden acceder los usuarios finales. En una canalización de CI/CD, el entorno de producción es el último entorno de implementación.
- entornos superiores: todos los entornos a los que pueden acceder usuarios que no sean del equipo de desarrollo principal. Esto puede incluir un entorno de producción, entornos de preproducción y entornos para las pruebas de aceptación por parte de los usuarios.

epopeya

En las metodologías ágiles, son categorías funcionales que ayudan a organizar y priorizar el trabajo. Las epopeyas brindan una descripción detallada de los requisitos y las tareas de implementación. Por ejemplo, las epopeyas AWS de seguridad de CAF incluyen la gestión de identidades y accesos, los controles de detección, la seguridad de la infraestructura, la protección de datos y la respuesta a incidentes. Para obtener más información sobre las epopeyas en la estrategia de migración de AWS , consulte la [Guía de implementación del programa](#).

PERP

Consulte [planificación de recursos empresariales](#).

análisis de datos de tipo exploratorio (EDA)

El proceso de analizar un conjunto de datos para comprender sus características principales. Se recopilan o agregan datos y, a continuación, se realizan las investigaciones iniciales para

encontrar patrones, detectar anomalías y comprobar las suposiciones. El EDA se realiza mediante el cálculo de estadísticas resumidas y la creación de visualizaciones de datos.

F

tabla de datos

La tabla central de un [esquema en forma de estrella](#). Almacena datos cuantitativos sobre las operaciones comerciales. Normalmente, una tabla de hechos contiene dos tipos de columnas: las que contienen medidas y las que contienen una clave externa para una tabla de dimensiones.

fallan rápidamente

Una filosofía que utiliza pruebas frecuentes e incrementales para reducir el ciclo de vida del desarrollo. Es una parte fundamental de un enfoque ágil.

límite de aislamiento de fallas

En el Nube de AWS, un límite, como una zona de disponibilidad Región de AWS, un plano de control o un plano de datos, que limita el efecto de una falla y ayuda a mejorar la resiliencia de las cargas de trabajo. Para obtener más información, consulte [Límites de AWS aislamiento de errores](#).

rama de característica

Consulte la [sucursal](#).

características

Los datos de entrada que se utilizan para hacer una predicción. Por ejemplo, en un contexto de fabricación, las características pueden ser imágenes que se capturan periódicamente desde la línea de fabricación.

importancia de las características

La importancia que tiene una característica para las predicciones de un modelo. Por lo general, esto se expresa como una puntuación numérica que se puede calcular mediante diversas técnicas, como las explicaciones aditivas de Shapley (SHAP) y los gradientes integrados. Para obtener más información, consulte [Interpretabilidad del modelo de aprendizaje automático con:AWS](#).

transformación de funciones

Optimizar los datos para el proceso de ML, lo que incluye enriquecer los datos con fuentes adicionales, escalar los valores o extraer varios conjuntos de información de un solo campo de datos. Esto permite que el modelo de ML se beneficie de los datos. Por ejemplo, si divide la fecha del “27 de mayo de 2021 00:15:37” en “jueves”, “mayo”, “2021” y “15”, puede ayudar al algoritmo de aprendizaje a aprender patrones matizados asociados a los diferentes componentes de los datos.

FGAC

Consulte el control [de acceso detallado](#).

control de acceso preciso (FGAC)

El uso de varias condiciones que tienen por objetivo permitir o denegar una solicitud de acceso.

migración relámpago

Método de migración de bases de datos que utiliza la replicación continua de datos mediante la [captura de datos modificados](#) para migrar los datos en el menor tiempo posible, en lugar de utilizar un enfoque gradual. El objetivo es reducir al mínimo el tiempo de inactividad.

G

bloqueo geográfico

Consulta [las restricciones geográficas](#).

restricciones geográficas (bloqueo geográfico)

En Amazon CloudFront, una opción para impedir que los usuarios de países específicos accedan a las distribuciones de contenido. Puede utilizar una lista de permitidos o bloqueados para especificar los países aprobados y prohibidos. Para obtener más información, consulta [Restringir la distribución geográfica del contenido](#) en la CloudFront documentación.

Flujo de trabajo de Gitflow

Un enfoque en el que los entornos inferiores y superiores utilizan diferentes ramas en un repositorio de código fuente. El flujo de trabajo de Gitflow se considera heredado, y el [flujo de trabajo basado en enlaces troncales](#) es el enfoque moderno preferido.

estrategia de implementación desde cero

La ausencia de infraestructura existente en un entorno nuevo. Al adoptar una estrategia de implementación desde cero para una arquitectura de sistemas, puede seleccionar todas las tecnologías nuevas sin que estas deban ser compatibles con una infraestructura existente, lo que también se conoce como [implementación sobre infraestructura existente](#). Si está ampliando la infraestructura existente, puede combinar las estrategias de implementación sobre infraestructuras existentes y de implementación desde cero.

barrera de protección

Una regla de alto nivel que ayuda a regular los recursos, las políticas y la conformidad en todas las unidades organizativas (OU). Las barreras de protección preventivas aplican políticas para garantizar la alineación con los estándares de conformidad. Se implementan mediante políticas de control de servicios y límites de permisos de IAM. Las barreras de protección de detección detectan las vulneraciones de las políticas y los problemas de conformidad, y generan alertas para su corrección. Se implementan mediante Amazon AWS Config AWS Security Hub GuardDuty AWS Trusted Advisor, Amazon Inspector y AWS Lambda cheques personalizados.

H

JA

Consulte [alta disponibilidad](#).

migración heterogénea de bases de datos

Migración de la base de datos de origen a una base de datos de destino que utilice un motor de base de datos diferente (por ejemplo, de Oracle a Amazon Aurora). La migración heterogénea suele ser parte de un esfuerzo de rediseño de la arquitectura y convertir el esquema puede ser una tarea compleja. [AWS ofrece AWS SCT](#), lo cual ayuda con las conversiones de esquemas.

alta disponibilidad (HA)

La capacidad de una carga de trabajo para funcionar de forma continua, sin intervención, en caso de desafíos o desastres. Los sistemas de alta disponibilidad están diseñados para realizar una conmutación por error automática, ofrecer un rendimiento de alta calidad de forma constante y gestionar diferentes cargas y fallos con un impacto mínimo en el rendimiento.

modernización histórica

Un enfoque utilizado para modernizar y actualizar los sistemas de tecnología operativa (TO) a fin de satisfacer mejor las necesidades de la industria manufacturera. Un histórico es un tipo de base de datos que se utiliza para recopilar y almacenar datos de diversas fuentes en una fábrica.

migración homogénea de bases de datos

Migración de la base de datos de origen a una base de datos de destino que comparte el mismo motor de base de datos (por ejemplo, Microsoft SQL Server a Amazon RDS para SQL Server).

La migración homogénea suele formar parte de un esfuerzo para volver a alojar o redefinir la plataforma. Puede utilizar las utilidades de bases de datos nativas para migrar el esquema.

datos recientes

Datos a los que se accede con frecuencia, como datos en tiempo real o datos traslacionales recientes. Por lo general, estos datos requieren un nivel o una clase de almacenamiento de alto rendimiento para proporcionar respuestas rápidas a las consultas.

hotfix

Una solución urgente para un problema crítico en un entorno de producción. Debido a su urgencia, las revisiones suelen realizarse fuera del flujo de trabajo habitual de las DevOps versiones.

periodo de hiperatención

Periodo, inmediatamente después de la transición, durante el cual un equipo de migración administra y monitorea las aplicaciones migradas en la nube para solucionar cualquier problema. Por lo general, este periodo dura de 1 a 4 días. Al final del periodo de hiperatención, el equipo de migración suele transferir la responsabilidad de las aplicaciones al equipo de operaciones en la nube.

I

IaC

Vea [la infraestructura como código](#).

políticas basadas en identidad

Política asociada a uno o más directores de IAM que define sus permisos en el Nube de AWS entorno.

I

aplicación inactiva

Aplicación que utiliza un promedio de CPU y memoria de entre 5 y 20 por ciento durante un periodo de 90 días. En un proyecto de migración, es habitual retirar estas aplicaciones o mantenerlas en las instalaciones.

IloT

Consulte [Internet de las cosas industrial](#).

infraestructura inmutable

Un modelo que implementa una nueva infraestructura para las cargas de trabajo de producción en lugar de actualizar, aplicar parches o modificar la infraestructura existente. [Las infraestructuras inmutables son intrínsecamente más consistentes, fiables y predecibles que las infraestructuras mutables](#). Para obtener más información, consulte las prácticas recomendadas para [implementar con una infraestructura inmutable](#) en Well-Architected Framework AWS .

VPC entrante (de entrada)

En una arquitectura de AWS cuentas múltiples, una VPC que acepta, inspecciona y enruta las conexiones de red desde fuera de una aplicación. La [Arquitectura de referencia de seguridad de AWS](#) recomienda configurar su cuenta de red con VPC entrantes, salientes y de inspección para proteger la interfaz bidireccional entre su aplicación e Internet en general.

migración gradual

Estrategia de transición en la que se migra la aplicación en partes pequeñas en lugar de realizar una transición única y completa. Por ejemplo, puede trasladar inicialmente solo unos pocos microservicios o usuarios al nuevo sistema. Tras comprobar que todo funciona correctamente, puede trasladar microservicios o usuarios adicionales de forma gradual hasta que pueda retirar su sistema heredado. Esta estrategia reduce los riesgos asociados a las grandes migraciones.

Industria 4.0

Un término que [Klaus Schwab](#) introdujo en 2016 para referirse a la modernización de los procesos de fabricación mediante avances en la conectividad, los datos en tiempo real, la automatización, el análisis y la inteligencia artificial/aprendizaje automático.

infraestructura

Todos los recursos y activos que se encuentran en el entorno de una aplicación.

infraestructura como código (IaC)

Proceso de aprovisionamiento y administración de la infraestructura de una aplicación mediante un conjunto de archivos de configuración. La IaC se ha diseñado para ayudarlo a centralizar la administración de la infraestructura, estandarizar los recursos y escalar con rapidez a fin de que los entornos nuevos sean repetibles, fiables y consistentes.

Internet de las cosas industrial (IIoT)

El uso de sensores y dispositivos conectados a Internet en los sectores industriales, como el productivo, el eléctrico, el automotriz, el sanitario, el de las ciencias de la vida y el de la agricultura. Para obtener más información, consulte [Creación de una estrategia de transformación digital del Internet de las cosas industrial \(IIoT\)](#).

VPC de inspección

En una arquitectura de AWS cuentas múltiples, una VPC centralizada que gestiona las inspecciones del tráfico de red entre las VPC (iguales o Regiones de AWS diferentes), Internet y las redes locales. La [Arquitectura de referencia de seguridad de AWS](#) recomienda configurar su cuenta de red con VPC entrantes, salientes y de inspección para proteger la interfaz bidireccional entre su aplicación e Internet en general.

Internet de las cosas (IoT)

Red de objetos físicos conectados con sensores o procesadores integrados que se comunican con otros dispositivos y sistemas a través de Internet o de una red de comunicación local. Para obtener más información, consulte [¿Qué es IoT?](#).

interpretabilidad

Característica de un modelo de machine learning que describe el grado en que un ser humano puede entender cómo las predicciones del modelo dependen de sus entradas. Para más información, consulte [Interpretabilidad del modelo de machine learning con AWS](#).

IoT

[Consulte Internet de las cosas.](#)

biblioteca de información de TI (ITIL)

Conjunto de prácticas recomendadas para ofrecer servicios de TI y alinearlos con los requisitos empresariales. La ITIL proporciona la base para la ITSM.

administración de servicios de TI (ITSM)

Actividades asociadas con el diseño, la implementación, la administración y el soporte de los servicios de TI para una organización. Para obtener información sobre la integración de las operaciones en la nube con las herramientas de ITSM, consulte la [Guía de integración de operaciones](#).

ITIL

Consulte la [biblioteca de información de TI](#).

ITSM

Consulte [Administración de servicios de TI](#).

L

control de acceso basado en etiquetas (LBAC)

Una implementación del control de acceso obligatorio (MAC) en la que a los usuarios y a los propios datos se les asigna explícitamente un valor de etiqueta de seguridad. La intersección entre la etiqueta de seguridad del usuario y la etiqueta de seguridad de los datos determina qué filas y columnas puede ver el usuario.

zona de aterrizaje

Una landing zone es un AWS entorno multicuenta bien diseñado, escalable y seguro. Este es un punto de partida desde el cual las empresas pueden lanzar e implementar rápidamente cargas de trabajo y aplicaciones con confianza en su entorno de seguridad e infraestructura. Para obtener más información sobre las zonas de aterrizaje, consulte [Configuración de un entorno de AWS seguro y escalable con varias cuentas](#).

migración grande

Migración de 300 servidores o más.

LBAC

Consulte el control de acceso basado en [etiquetas](#).

privilegio mínimo

La práctica recomendada de seguridad que consiste en conceder los permisos mínimos necesarios para realizar una tarea. Para obtener más información, consulte [Aplicar permisos de privilegio mínimo](#) en la documentación de IAM.

migrar mediante lift-and-shift

Ver [7 Rs](#).

sistema little-endian

Un sistema que almacena primero el byte menos significativo. Véase también [endianness](#).

entornos inferiores

[Véase entorno](#).

M

machine learning (ML)

Un tipo de inteligencia artificial que utiliza algoritmos y técnicas para el reconocimiento y el aprendizaje de patrones. El ML analiza y aprende de los datos registrados, como los datos del Internet de las cosas (IoT), para generar un modelo estadístico basado en patrones. Para más información, consulte [Machine learning](#).

rama principal

Ver [sucursal](#).

malware

Software diseñado para comprometer la seguridad o la privacidad de la computadora. El malware puede interrumpir los sistemas informáticos, filtrar información confidencial u obtener acceso no autorizado. Algunos ejemplos de malware son los virus, los gusanos, el ransomware, los troyanos, el spyware y los registradores de pulsaciones de teclas.

servicios gestionados

Servicios de AWS para los que AWS opera la capa de infraestructura, el sistema operativo y las plataformas, y usted accede a los puntos finales para almacenar y recuperar datos. Amazon Simple Storage Service (Amazon S3) y Amazon DynamoDB son ejemplos de servicios gestionados. También se conocen como servicios abstractos.

sistema de ejecución de fabricación (MES)

Un sistema de software para rastrear, monitorear, documentar y controlar los procesos de producción que convierten las materias primas en productos terminados en el taller.

MAP

Consulte [Migration Acceleration Program](#).

mecanismo

Un proceso completo en el que se crea una herramienta, se impulsa su adopción y, a continuación, se inspeccionan los resultados para realizar ajustes. Un mecanismo es un ciclo que se refuerza y mejora a sí mismo a medida que funciona. Para obtener más información, consulte [Creación de mecanismos](#) en el AWS Well-Architected Framework.

cuenta de miembro

Todas las Cuentas de AWS demás cuentas, excepto la de administración, que forman parte de una organización. AWS Organizations Una cuenta no puede pertenecer a más de una organización a la vez.

MES

Consulte el [sistema de ejecución de la fabricación](#).

Transporte telemétrico de Message Queue Queue (MQTT)

[Un protocolo de comunicación ligero machine-to-machine \(M2M\), basado en el patrón de publicación/suscripción, para dispositivos de IoT con recursos limitados.](#)

microservicio

Un servicio pequeño e independiente que se comunica a través de API bien definidas y que, por lo general, es propiedad de equipos pequeños e independientes. Por ejemplo, un sistema de seguros puede incluir microservicios que se adapten a las capacidades empresariales, como las de ventas o marketing, o a subdominios, como las de compras, reclamaciones o análisis. Los beneficios de los microservicios incluyen la agilidad, la escalabilidad flexible, la facilidad de implementación, el código reutilizable y la resiliencia. Para obtener más información, consulte [Integrar](#) microservicios mediante servicios sin servidor. AWS

arquitectura de microservicios

Un enfoque para crear una aplicación con componentes independientes que ejecutan cada proceso de la aplicación como un microservicio. Estos microservicios se comunican a través de

una interfaz bien definida mediante API ligeras. Cada microservicio de esta arquitectura se puede actualizar, implementar y escalar para satisfacer la demanda de funciones específicas de una aplicación. Para obtener más información, consulte [Implementación de microservicios](#) en. AWS

Programa de aceleración de la migración (MAP)

Un AWS programa que proporciona soporte de consultoría, formación y servicios para ayudar a las organizaciones a crear una base operativa sólida para migrar a la nube y para ayudar a compensar el costo inicial de las migraciones. El MAP incluye una metodología de migración para ejecutar las migraciones antiguas de forma metódica y un conjunto de herramientas para automatizar y acelerar los escenarios de migración más comunes.

migración a escala

Proceso de transferencia de la mayoría de la cartera de aplicaciones a la nube en oleadas, con más aplicaciones desplazadas a un ritmo más rápido en cada oleada. En esta fase, se utilizan las prácticas recomendadas y las lecciones aprendidas en las fases anteriores para implementar una fábrica de migración de equipos, herramientas y procesos con el fin de agilizar la migración de las cargas de trabajo mediante la automatización y la entrega ágil. Esta es la tercera fase de la [estrategia de migración de AWS](#).

fábrica de migración

Equipos multifuncionales que agilizan la migración de las cargas de trabajo mediante enfoques automatizados y ágiles. Los equipos de las fábricas de migración suelen incluir a analistas y propietarios de operaciones, empresas, ingenieros de migración, desarrolladores y DevOps profesionales que trabajan a pasos agigantados. Entre el 20 y el 50 por ciento de la cartera de aplicaciones empresariales se compone de patrones repetidos que pueden optimizarse mediante un enfoque de fábrica. Para obtener más información, consulte la [discusión sobre las fábricas de migración](#) y la [Guía de fábricas de migración a la nube](#) en este contenido.

metadatos de migración

Información sobre la aplicación y el servidor que se necesita para completar la migración. Cada patrón de migración requiere un conjunto diferente de metadatos de migración. Algunos ejemplos de metadatos de migración son la subred de destino, el grupo de seguridad y AWS la cuenta.

patrón de migración

Tarea de migración repetible que detalla la estrategia de migración, el destino de la migración y la aplicación o el servicio de migración utilizados. Ejemplo: rehospede la migración a Amazon EC2 AWS con Application Migration Service.

Migration Portfolio Assessment (MPA)

Una herramienta en línea que proporciona información para validar el modelo de negocio para migrar a. Nube de AWS La MPA ofrece una evaluación detallada de la cartera (adecuación del tamaño de los servidores, precios, comparaciones del costo total de propiedad, análisis de los costos de migración), así como una planificación de la migración (análisis y recopilación de datos de aplicaciones, agrupación de aplicaciones, priorización de la migración y planificación de oleadas). La [herramienta MPA](#) (requiere iniciar sesión) está disponible de forma gratuita para todos los AWS consultores y consultores asociados de APN.

Evaluación de la preparación para la migración (MRA)

Proceso que consiste en obtener información sobre el estado de preparación de una organización para la nube, identificar sus puntos fuertes y débiles y elaborar un plan de acción para cerrar las brechas identificadas mediante el AWS CAF. Para obtener más información, consulte la [Guía de preparación para la migración](#). La MRA es la primera fase de la [estrategia de migración de AWS](#).

estrategia de migración

El enfoque utilizado para migrar una carga de trabajo a. Nube de AWS Para obtener más información, consulte la entrada de las [7 R](#) de este glosario y consulte [Movilice a su organización para acelerar las migraciones a gran escala](#).

ML

[Consulte el aprendizaje automático.](#)

modernización

Transformar una aplicación obsoleta (antigua o monolítica) y su infraestructura en un sistema ágil, elástico y de alta disponibilidad en la nube para reducir los gastos, aumentar la eficiencia y aprovechar las innovaciones. Para obtener más información, consulte [Estrategia para modernizar las aplicaciones en el Nube de AWS](#).

evaluación de la preparación para la modernización

Evaluación que ayuda a determinar la preparación para la modernización de las aplicaciones de una organización; identifica los beneficios, los riesgos y las dependencias; y determina qué tan bien la organización puede soportar el estado futuro de esas aplicaciones. El resultado de la evaluación es un esquema de la arquitectura objetivo, una hoja de ruta que detalla las fases de desarrollo y los hitos del proceso de modernización y un plan de acción para abordar las brechas identificadas. Para obtener más información, consulte [Evaluación de la preparación para la modernización de las aplicaciones en el Nube de AWS](#).

aplicaciones monolíticas (monolitos)

Aplicaciones que se ejecutan como un único servicio con procesos estrechamente acoplados. Las aplicaciones monolíticas presentan varios inconvenientes. Si una característica de la aplicación experimenta un aumento en la demanda, se debe escalar toda la arquitectura. Agregar o mejorar las características de una aplicación monolítica también se vuelve más complejo a medida que crece la base de código. Para solucionar problemas con la aplicación, puede utilizar una arquitectura de microservicios. Para obtener más información, consulte [Descomposición de monolitos en microservicios](#).

MAPA

Consulte [la evaluación de la cartera de migración](#).

MQTT

Consulte [Message Queue Queue Telemetría](#) y Transporte.

clasificación multiclase

Un proceso que ayuda a generar predicciones para varias clases (predice uno de más de dos resultados). Por ejemplo, un modelo de ML podría preguntar “¿Este producto es un libro, un automóvil o un teléfono?” o “¿Qué categoría de productos es más interesante para este cliente?”.

infraestructura mutable

Un modelo que actualiza y modifica la infraestructura existente para las cargas de trabajo de producción. Para mejorar la coherencia, la fiabilidad y la previsibilidad, el AWS Well-Architected Framework recomienda el uso [de una infraestructura inmutable](#) como práctica recomendada.

O

OAC

[Consulte el control de acceso de origen](#).

OAI

Consulte la [identidad de acceso de origen](#).

OCM

Consulte [gestión del cambio organizacional](#).

migración fuera de línea

Método de migración en el que la carga de trabajo de origen se elimina durante el proceso de migración. Este método implica un tiempo de inactividad prolongado y, por lo general, se utiliza para cargas de trabajo pequeñas y no críticas.

OI

Consulte [integración de operaciones](#).

OLA

Véase el [acuerdo a nivel operativo](#).

migración en línea

Método de migración en el que la carga de trabajo de origen se copia al sistema de destino sin que se desconecte. Las aplicaciones que están conectadas a la carga de trabajo pueden seguir funcionando durante la migración. Este método implica un tiempo de inactividad nulo o mínimo y, por lo general, se utiliza para cargas de trabajo de producción críticas.

OPC-UA

Consulte [Open Process Communications: arquitectura unificada](#).

Comunicaciones de proceso abierto: arquitectura unificada (OPC-UA)

Un protocolo de comunicación machine-to-machine (M2M) para la automatización industrial. El OPC-UA proporciona un estándar de interoperabilidad con esquemas de cifrado, autenticación y autorización de datos.

acuerdo de nivel operativo (OLA)

Acuerdo que aclara lo que los grupos de TI operativos se comprometen a ofrecerse entre sí, para respaldar un acuerdo de nivel de servicio (SLA).

revisión de la preparación operativa (ORR)

Una lista de preguntas y las mejores prácticas asociadas que le ayudan a comprender, evaluar, prevenir o reducir el alcance de los incidentes y posibles fallos. Para obtener más información, consulte [Operational Readiness Reviews \(ORR\)](#) en AWS Well-Architected Framework.

tecnología operativa (OT)

Sistemas de hardware y software que funcionan con el entorno físico para controlar las operaciones, los equipos y la infraestructura industriales. En la industria manufacturera, la

integración de los sistemas de TO y tecnología de la información (TI) es un enfoque clave para las transformaciones de [la industria 4.0](#).

integración de operaciones (OI)

Proceso de modernización de las operaciones en la nube, que implica la planificación de la preparación, la automatización y la integración. Para obtener más información, consulte la [Guía de integración de las operaciones](#).

registro de seguimiento organizativo

Un registro creado por el AWS CloudTrail que se registran todos los eventos para todos Cuentas de AWS los miembros de una organización AWS Organizations. Este registro de seguimiento se crea en cada Cuenta de AWS que forma parte de la organización y realiza un seguimiento de la actividad en cada cuenta. Para obtener más información, consulte [Crear un registro para una organización](#) en la CloudTrail documentación.

administración del cambio organizacional (OCM)

Marco para administrar las transformaciones empresariales importantes y disruptivas desde la perspectiva de las personas, la cultura y el liderazgo. La OCM ayuda a las empresas a prepararse para nuevos sistemas y estrategias y a realizar la transición a ellos, al acelerar la adopción de cambios, abordar los problemas de transición e impulsar cambios culturales y organizacionales. En la estrategia de AWS migración, este marco se denomina aceleración de personal, debido a la velocidad de cambio que requieren los proyectos de adopción de la nube. Para obtener más información, consulte la [Guía de OCM](#).

control de acceso de origen (OAC)

En CloudFront, una opción mejorada para restringir el acceso y proteger el contenido del Amazon Simple Storage Service (Amazon S3). El OAC admite todos los buckets de S3 Regiones de AWS, el cifrado del lado del servidor AWS KMS (SSE-KMS) y las solicitudes dinámicas PUT y DELETE dirigidas al bucket de S3.

identidad de acceso de origen (OAI)

En CloudFront, una opción para restringir el acceso y proteger el contenido de Amazon S3. Cuando utiliza OAI, CloudFront crea un principal con el que Amazon S3 puede autenticarse. Los directores autenticados solo pueden acceder al contenido de un bucket de S3 a través de una distribución específica. CloudFront Consulte también el [OAC](#), que proporciona un control de acceso más detallado y mejorado.

O

Consulte la [revisión de la preparación operativa](#).

NO

Consulte [tecnología operativa](#).

VPC saliente (de salida)

En una arquitectura de AWS cuentas múltiples, una VPC que gestiona las conexiones de red que se inician desde una aplicación. La [Arquitectura de referencia de seguridad de AWS](#) recomienda configurar su cuenta de red con VPC entrantes, salientes y de inspección para proteger la interfaz bidireccional entre su aplicación e Internet en general.

P

límite de permisos

Una política de administración de IAM que se adjunta a las entidades principales de IAM para establecer los permisos máximos que puede tener el usuario o el rol. Para obtener más información, consulte [Límites de permisos](#) en la documentación de IAM.

información de identificación personal (PII)

Información que, vista directamente o combinada con otros datos relacionados, puede utilizarse para deducir de manera razonable la identidad de una persona. Algunos ejemplos de información de identificación personal son los nombres, las direcciones y la información de contacto.

PII

Consulte la información de [identificación personal](#).

manual de estrategias

Conjunto de pasos predefinidos que capturan el trabajo asociado a las migraciones, como la entrega de las funciones de operaciones principales en la nube. Un manual puede adoptar la forma de scripts, manuales de procedimientos automatizados o resúmenes de los procesos o pasos necesarios para operar un entorno modernizado.

PLC

Consulte [controlador lógico programable](#).

PLM

Consulte la [gestión del ciclo de vida del producto](#).

política

Un objeto que puede definir los permisos (consulte la [política basada en la identidad](#)), especifique las condiciones de acceso (consulte la [política basada en los recursos](#)) o defina los permisos máximos para todas las cuentas de una organización AWS Organizations (consulte la política de control de [servicios](#)).

persistencia políglota

Elegir de forma independiente la tecnología de almacenamiento de datos de un microservicio en función de los patrones de acceso a los datos y otros requisitos. Si sus microservicios tienen la misma tecnología de almacenamiento de datos, pueden enfrentarse a desafíos de implementación o experimentar un rendimiento deficiente. Los microservicios se implementan más fácilmente y logran un mejor rendimiento y escalabilidad si utilizan el almacén de datos que mejor se adapte a sus necesidades. Para obtener más información, consulte [Habilitación de la persistencia de datos en los microservicios](#).

evaluación de cartera

Proceso de detección, análisis y priorización de la cartera de aplicaciones para planificar la migración. Para obtener más información, consulte la [Evaluación de la preparación para la migración](#).

predicate

Una condición de consulta que devuelve true o false, por lo general, se encuentra en una cláusula. WHERE

pulsar un predicado

Técnica de optimización de consultas de bases de datos que filtra los datos de la consulta antes de transferirlos. Esto reduce la cantidad de datos que se deben recuperar y procesar de la base de datos relacional y mejora el rendimiento de las consultas.

control preventivo

Un control de seguridad diseñado para evitar que ocurra un evento. Estos controles son la primera línea de defensa para evitar el acceso no autorizado o los cambios no deseados en la red. Para obtener más información, consulte [Controles preventivos](#) en Implementación de controles de seguridad en AWS.

entidad principal

Una entidad AWS que puede realizar acciones y acceder a los recursos. Esta entidad suele ser un usuario raíz para un Cuenta de AWS rol de IAM o un usuario. Para obtener más información, consulte Entidad principal en [Términos y conceptos de roles](#) en la documentación de IAM.

Privacidad desde el diseño

Un enfoque de ingeniería de sistemas que tiene en cuenta la privacidad durante todo el proceso de ingeniería.

zonas alojadas privadas

Contenedor que aloja información acerca de cómo desea que responda Amazon Route 53 a las consultas de DNS de un dominio y sus subdominios en una o varias VPC. Para obtener más información, consulte [Uso de zonas alojadas privadas](#) en la documentación de Route 53.

control proactivo

Un [control de seguridad](#) diseñado para evitar el despliegue de recursos no conformes. Estos controles escanean los recursos antes de aprovisionarlos. Si el recurso no cumple con el control, significa que no está aprovisionado. Para obtener más información, consulte la [guía de referencia de controles](#) en la AWS Control Tower documentación y consulte [Controles proactivos](#) en Implementación de controles de seguridad en AWS.

gestión del ciclo de vida del producto (PLM)

La gestión de los datos y los procesos de un producto a lo largo de todo su ciclo de vida, desde el diseño, el desarrollo y el lanzamiento, pasando por el crecimiento y la madurez, hasta el rechazo y la retirada.

entorno de producción

Consulte [el entorno](#).

controlador lógico programable (PLC)

En la fabricación, una computadora adaptable y altamente confiable que monitorea las máquinas y automatiza los procesos de fabricación.

seudonimización

El proceso de reemplazar los identificadores personales de un conjunto de datos por valores de marcadores de posición. La seudonimización puede ayudar a proteger la privacidad personal. Los datos seudonimizados siguen considerándose datos personales.

publicar/suscribirse (pub/sub)

Un patrón que permite las comunicaciones asíncronas entre microservicios para mejorar la escalabilidad y la capacidad de respuesta. Por ejemplo, en un [MES](#) basado en microservicios, un microservicio puede publicar mensajes de eventos en un canal al que se puedan suscribir otros microservicios. El sistema puede añadir nuevos microservicios sin cambiar el servicio de publicación.

Q

plan de consulta

Serie de pasos, como instrucciones, que se utilizan para acceder a los datos de un sistema de base de datos relacional SQL.

regresión del plan de consulta

El optimizador de servicios de la base de datos elige un plan menos óptimo que antes de un cambio determinado en el entorno de la base de datos. Los cambios en estadísticas, restricciones, configuración del entorno, enlaces de parámetros de consultas y actualizaciones del motor de base de datos PostgreSQL pueden provocar una regresión del plan.

R

Matriz RACI

Véase [responsable, responsable, consultado, informado \(RACI\)](#).

ransomware

Software malicioso que se ha diseñado para bloquear el acceso a un sistema informático o a los datos hasta que se efectúe un pago.

Matriz RASCI

Véase [responsable, responsable, consultado, informado \(RACI\)](#).

RCAC

Consulte control de [acceso por filas y columnas](#).

read replica

Una copia de una base de datos que se utiliza con fines de solo lectura. Puede enrutar las consultas a la réplica de lectura para reducir la carga en la base de datos principal.

rediseñar

Ver [7 Rs.](#)

objetivo de punto de recuperación (RPO)

La cantidad de tiempo máximo aceptable desde el último punto de recuperación de datos. Esto determina qué se considera una pérdida de datos aceptable entre el último punto de recuperación y la interrupción del servicio.

objetivo de tiempo de recuperación (RTO)

La demora máxima aceptable entre la interrupción del servicio y el restablecimiento del servicio.

refactorizar

Ver [7 Rs.](#)

Región

Una colección de AWS recursos en un área geográfica. Cada uno Región de AWS está aislado e independiente de los demás para proporcionar tolerancia a las fallas, estabilidad y resiliencia. Para obtener más información, consulte [Regiones de AWS Especificar qué cuenta puede usar.](#)

regresión

Una técnica de ML que predice un valor numérico. Por ejemplo, para resolver el problema de “¿A qué precio se venderá esta casa?”, un modelo de ML podría utilizar un modelo de regresión lineal para predecir el precio de venta de una vivienda en función de datos conocidos sobre ella (por ejemplo, los metros cuadrados).

volver a alojar

Consulte [7 Rs.](#)

versión

En un proceso de implementación, el acto de promover cambios en un entorno de producción.

trasladarse

Ver [7 Rs.](#)

redefinir la plataforma

Ver [7 Rs.](#)

recompra

Ver [7 Rs.](#)

resiliencia

La capacidad de una aplicación para resistir las interrupciones o recuperarse de ellas. [La alta disponibilidad](#) y la [recuperación ante desastres](#) son consideraciones comunes a la hora de planificar la resiliencia en el. Nube de AWS Para obtener más información, consulte [Nube de AWS Resiliencia](#).

política basada en recursos

Una política asociada a un recurso, como un bucket de Amazon S3, un punto de conexión o una clave de cifrado. Este tipo de política especifica a qué entidades principales se les permite el acceso, las acciones compatibles y cualquier otra condición que deba cumplirse.

matriz responsable, confiable, consultada e informada (RACI)

Una matriz que define las funciones y responsabilidades de todas las partes involucradas en las actividades de migración y las operaciones de la nube. El nombre de la matriz se deriva de los tipos de responsabilidad definidos en la matriz: responsable (R), contable (A), consultado (C) e informado (I). El tipo de soporte (S) es opcional. Si incluye el soporte, la matriz se denomina matriz RASCI y, si la excluye, se denomina matriz RACI.

control receptivo

Un control de seguridad que se ha diseñado para corregir los eventos adversos o las desviaciones con respecto a su base de seguridad. Para obtener más información, consulte [Controles receptivos](#) en Implementación de controles de seguridad en AWS.

retain

Consulte [7 Rs.](#)

jubilarse

Ver [7 Rs.](#)

rotación

Proceso de actualizar periódicamente un [secreto](#) para dificultar el acceso de un atacante a las credenciales.

control de acceso por filas y columnas (RCAC)

El uso de expresiones SQL básicas y flexibles que tienen reglas de acceso definidas. El RCAC consta de permisos de fila y máscaras de columnas.

RPO

Consulte el [objetivo del punto de recuperación](#).

RTO

Consulte el [objetivo de tiempo de recuperación](#).

manual de procedimientos

Conjunto de procedimientos manuales o automatizados necesarios para realizar una tarea específica. Por lo general, se diseñan para agilizar las operaciones o los procedimientos repetitivos con altas tasas de error.

S

SAML 2.0

Un estándar abierto que utilizan muchos proveedores de identidad (IdPs). Esta función permite el inicio de sesión único (SSO) federado, de modo que los usuarios pueden iniciar sesión AWS Management Console o llamar a las operaciones de la AWS API sin tener que crear un usuario en IAM para todos los miembros de la organización. Para obtener más información sobre la federación basada en SAML 2.0, consulte [Acerca de la federación basada en SAML 2.0](#) en la documentación de IAM.

SCADA

Consulte el [control de supervisión y la adquisición de datos](#).

SCP

Consulte la [política de control de servicios](#).

secreta

Información confidencial o restringida, como una contraseña o credenciales de usuario, que almacene de forma cifrada. AWS Secrets Manager Se compone del valor secreto y sus

metadatos. El valor secreto puede ser binario, una sola cadena o varias cadenas. Para obtener más información, consulta [¿Qué hay en un secreto de Secrets Manager?](#) en la documentación de Secrets Manager.

control de seguridad

Barrera de protección técnica o administrativa que impide, detecta o reduce la capacidad de un agente de amenazas para aprovechar una vulnerabilidad de seguridad. Hay cuatro tipos principales de controles de seguridad: [preventivos](#), de detección, de [respuesta](#) y [proactivos](#).

refuerzo de la seguridad

Proceso de reducir la superficie expuesta a ataques para hacerla más resistente a los ataques. Esto puede incluir acciones, como la eliminación de los recursos que ya no se necesitan, la implementación de prácticas recomendadas de seguridad consistente en conceder privilegios mínimos o la desactivación de características innecesarias en los archivos de configuración.

sistema de información sobre seguridad y administración de eventos (SIEM)

Herramientas y servicios que combinan sistemas de administración de información sobre seguridad (SIM) y de administración de eventos de seguridad (SEM). Un sistema de SIEM recopila, monitorea y analiza los datos de servidores, redes, dispositivos y otras fuentes para detectar amenazas y brechas de seguridad y generar alertas.

automatización de la respuesta de seguridad

Una acción predefinida y programada que está diseñada para responder automáticamente a un evento de seguridad o remediarlo. Estas automatizaciones sirven como controles de seguridad [detectables](#) o [adaptables](#) que le ayudan a implementar las mejores prácticas AWS de seguridad. Algunos ejemplos de acciones de respuesta automatizadas incluyen la modificación de un grupo de seguridad de VPC, la aplicación de parches a una instancia de Amazon EC2 o la rotación de credenciales.

cifrado del servidor

Cifrado de los datos en su destino, por parte de quien Servicio de AWS los recibe.

política de control de servicio (SCP)

Una política que proporciona un control centralizado de los permisos de todas las cuentas de una organización en AWS Organizations. Las SCP definen barreras de protección o establecen límites a las acciones que un administrador puede delegar en los usuarios o roles. Puede utilizar las SCP como listas de permitidos o rechazados, para especificar qué servicios o acciones se

encuentra permitidos o prohibidos. Para obtener más información, consulte [las políticas de control de servicios](#) en la AWS Organizations documentación.

punto de enlace de servicio

La URL del punto de entrada de un Servicio de AWS. Para conectarse mediante programación a un servicio de destino, puede utilizar un punto de conexión. Para obtener más información, consulte [Puntos de conexión de Servicio de AWS](#) en Referencia general de AWS.

acuerdo de nivel de servicio (SLA)

Acuerdo que aclara lo que un equipo de TI se compromete a ofrecer a los clientes, como el tiempo de actividad y el rendimiento del servicio.

indicador de nivel de servicio (SLI)

Medición de un aspecto del rendimiento de un servicio, como la tasa de errores, la disponibilidad o el rendimiento.

objetivo de nivel de servicio (SLO)

[Una métrica objetivo que representa el estado de un servicio, medido mediante un indicador de nivel de servicio.](#)

modelo de responsabilidad compartida

Un modelo que describe la responsabilidad que compartes con respecto a la seguridad y AWS el cumplimiento de la nube. AWS es responsable de la seguridad de la nube, mientras que usted es responsable de la seguridad en la nube. Para obtener más información, consulte el [Modelo de responsabilidad compartida](#).

SIEM

Consulte [la información de seguridad y el sistema de gestión de eventos](#).

punto único de fallo (SPOF)

Una falla en un único componente crítico de una aplicación que puede interrumpir el sistema.

SLA

Consulte el acuerdo [de nivel de servicio](#).

SLI

Consulte el indicador de [nivel de servicio](#).

ASÍ QUE

Consulte el objetivo de [nivel de servicio](#).

split-and-seed modelo

Un patrón para escalar y acelerar los proyectos de modernización. A medida que se definen las nuevas funciones y los lanzamientos de los productos, el equipo principal se divide para crear nuevos equipos de productos. Esto ayuda a ampliar las capacidades y los servicios de su organización, mejora la productividad de los desarrolladores y apoya la innovación rápida. Para obtener más información, consulte [Enfoque gradual para modernizar las aplicaciones en el. Nube de AWS](#)

SPOT

Consulte el [punto único de falla](#).

esquema en forma de estrella

Estructura organizativa de una base de datos que utiliza una tabla de datos grande para almacenar datos transaccionales o medidos y una o más tablas dimensionales más pequeñas para almacenar los atributos de los datos. Esta estructura está diseñada para usarse en un [almacén de datos](#) o con fines de inteligencia empresarial.

patrón de higo estrangulador

Un enfoque para modernizar los sistemas monolíticos mediante la reescritura y el reemplazo gradual de las funciones del sistema hasta que se pueda desmantelar el sistema heredado. Este patrón utiliza la analogía de una higuera que crece hasta convertirse en un árbol estable y, finalmente, se apodera y reemplaza a su host. El patrón fue [presentado por Martin Fowler](#) como una forma de gestionar el riesgo al reescribir sistemas monolíticos. Para ver un ejemplo con la aplicación de este patrón, consulte [Modernización gradual de los servicios web antiguos de Microsoft ASP.NET \(ASMX\) mediante contenedores y Amazon API Gateway](#).

subred

Un intervalo de direcciones IP en la VPC. Una subred debe residir en una sola zona de disponibilidad.

supervisión, control y adquisición de datos (SCADA)

En la industria manufacturera, un sistema que utiliza hardware y software para monitorear los activos físicos y las operaciones de producción.

cifrado simétrico

Un algoritmo de cifrado que utiliza la misma clave para cifrar y descifrar los datos.

pruebas sintéticas

Probar un sistema de manera que simule las interacciones de los usuarios para detectar posibles problemas o monitorear el rendimiento. Puede usar [Amazon CloudWatch Synthetics](#) para crear estas pruebas.

T

etiquetas

Pares clave-valor que actúan como metadatos para organizar los recursos. AWS Las etiquetas pueden ayudarle a administrar, identificar, organizar, buscar y filtrar recursos. Para obtener más información, consulte [Etiquetado de los recursos de AWS](#).

variable de destino

El valor que intenta predecir en el ML supervisado. Esto también se conoce como variable de resultado. Por ejemplo, en un entorno de fabricación, la variable objetivo podría ser un defecto del producto.

lista de tareas

Herramienta que se utiliza para hacer un seguimiento del progreso mediante un manual de procedimientos. La lista de tareas contiene una descripción general del manual de procedimientos y una lista de las tareas generales que deben completarse. Para cada tarea general, se incluye la cantidad estimada de tiempo necesario, el propietario y el progreso.

entorno de prueba

[Consulte entorno.](#)

entrenamiento

Proporcionar datos de los que pueda aprender su modelo de ML. Los datos de entrenamiento deben contener la respuesta correcta. El algoritmo de aprendizaje encuentra patrones en los datos de entrenamiento que asignan los atributos de los datos de entrada al destino (la respuesta que desea predecir). Genera un modelo de ML que captura estos patrones. Luego, el modelo de ML se puede utilizar para obtener predicciones sobre datos nuevos para los que no se conoce el destino.

puerta de enlace de tránsito

Centro de tránsito de red que puede utilizar para interconectar las VPC y las redes en las instalaciones. Para obtener más información, consulte [Qué es una pasarela de tránsito](#) en la AWS Transit Gateway documentación.

flujo de trabajo basado en enlaces troncales

Un enfoque en el que los desarrolladores crean y prueban características de forma local en una rama de característica y, a continuación, combinan esos cambios en la rama principal. Luego, la rama principal se adapta a los entornos de desarrollo, preproducción y producción, de forma secuencial.

acceso de confianza

Otorgar permisos a un servicio que especifique para realizar tareas en su organización AWS Organizations y en sus cuentas en su nombre. El servicio de confianza crea un rol vinculado al servicio en cada cuenta, cuando ese rol es necesario, para realizar las tareas de administración por usted. Para obtener más información, consulte [AWS Organizations Utilización con otros AWS servicios](#) en la AWS Organizations documentación.

ajuste

Cambiar aspectos de su proceso de formación a fin de mejorar la precisión del modelo de ML. Por ejemplo, puede entrenar el modelo de ML al generar un conjunto de etiquetas, incorporar etiquetas y, luego, repetir estos pasos varias veces con diferentes ajustes para optimizar el modelo.

equipo de dos pizzas

Un DevOps equipo pequeño al que puedes alimentar con dos pizzas. Un equipo formado por dos integrantes garantiza la mejor oportunidad posible de colaboración en el desarrollo de software.

U

incertidumbre

Un concepto que hace referencia a información imprecisa, incompleta o desconocida que puede socavar la fiabilidad de los modelos predictivos de ML. Hay dos tipos de incertidumbre: la incertidumbre epistémica se debe a datos limitados e incompletos, mientras que la incertidumbre aleatoria se debe al ruido y la aleatoriedad inherentes a los datos. Para más información, consulte la guía [Cuantificación de la incertidumbre en los sistemas de aprendizaje profundo](#).

tareas indiferenciadas

También conocido como tareas arduas, es el trabajo que es necesario para crear y operar una aplicación, pero que no proporciona un valor directo al usuario final ni proporciona una ventaja competitiva. Algunos ejemplos de tareas indiferenciadas son la adquisición, el mantenimiento y la planificación de la capacidad.

entornos superiores

Ver [entorno](#).

V

succión

Una operación de mantenimiento de bases de datos que implica limpiar después de las actualizaciones incrementales para recuperar espacio de almacenamiento y mejorar el rendimiento.

control de versión

Procesos y herramientas que realizan un seguimiento de los cambios, como los cambios en el código fuente de un repositorio.

Emparejamiento de VPC

Conexión entre dos VPC que permite enrutar el tráfico mediante direcciones IP privadas. Para obtener más información, consulte [¿Qué es una interconexión de VPC?](#) en la documentación de Amazon VPC.

vulnerabilidad

Defecto de software o hardware que pone en peligro la seguridad del sistema.

W

caché caliente

Un búfer caché que contiene datos actuales y relevantes a los que se accede con frecuencia. La instancia de base de datos puede leer desde la caché del búfer, lo que es más rápido que leer desde la memoria principal o el disco.

datos templados

Datos a los que el acceso es infrecuente. Al consultar este tipo de datos, normalmente se aceptan consultas moderadamente lentas.

función de ventana

Función SQL que realiza un cálculo en un grupo de filas que se relacionan de alguna manera con el registro actual. Las funciones de ventana son útiles para procesar tareas, como calcular una media móvil o acceder al valor de las filas en función de la posición relativa de la fila actual.

carga de trabajo

Conjunto de recursos y código que ofrece valor comercial, como una aplicación orientada al cliente o un proceso de backend.

flujo de trabajo

Grupos funcionales de un proyecto de migración que son responsables de un conjunto específico de tareas. Cada flujo de trabajo es independiente, pero respalda a los demás flujos de trabajo del proyecto. Por ejemplo, el flujo de trabajo de la cartera es responsable de priorizar las aplicaciones, planificar las oleadas y recopilar los metadatos de migración. El flujo de trabajo de la cartera entrega estos recursos al flujo de trabajo de migración, que luego migra los servidores y las aplicaciones.

GUSANO

Mira, [escribe una vez, lee muchas](#).

WQF

Consulte el [marco de calificación de cargas de trabajo de AWS](#).

escribe una vez, lee muchas (WORM)

Un modelo de almacenamiento que escribe los datos una sola vez y evita que los datos se eliminen o modifiquen. Los usuarios autorizados pueden leer los datos tantas veces como sea necesario, pero no pueden cambiarlos. Esta infraestructura de almacenamiento de datos se considera [inmutable](#).

Z

ataque de día cero

Un ataque, normalmente de malware, que aprovecha una vulnerabilidad de [día cero](#).

vulnerabilidad de día cero

Un defecto o una vulnerabilidad sin mitigación en un sistema de producción. Los agentes de amenazas pueden usar este tipo de vulnerabilidad para atacar el sistema. Los desarrolladores suelen darse cuenta de la vulnerabilidad a raíz del ataque.

aplicación zombi

Aplicación que utiliza un promedio de CPU y memoria menor al 5 por ciento. En un proyecto de migración, es habitual retirar estas aplicaciones.

Las traducciones son generadas a través de traducción automática. En caso de conflicto entre la traducción y la versión original de inglés, prevalecerá la versión en inglés.