



Diseñar un lago de datos para crecer y escalar en el Nube de AWS

AWS Guía prescriptiva



AWS Guía prescriptiva: Diseñar un lago de datos para crecer y escalar en el Nube de AWS

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
Resultados comerciales específicos	2
Desafíos comunes de escalado	3
Despliegue inicial del lago de datos	3
Aumentan los consumidores de datos	4
Aumentan los productores de datos	5
Arquitectura de referencia	7
Componentes de la arquitectura de referencia	9
Productores de datos	10
Catálogo centralizado	12
AWS Lake Formation	13
AWS Glue	14
Consumidores de datos	15
Incorporación y concesión de acceso	19
Incorporación de productores de datos	19
Incorporación de consumidores de datos	20
Otorgar acceso a Select en una cuenta de consumidor de datos	21
Preguntas frecuentes	23
¿Esta arquitectura de referencia del lago de datos es más aplicable a las organizaciones empresariales?	23
¿Puedo seguir usando esta arquitectura de referencia si mi organización solo tiene un productor de datos?	23
Mi lago de datos conecta directamente a un productor de datos con varios consumidores de datos. ¿Sigue siendo relevante la arquitectura de referencia del lago de datos de esta guía?	23
¿Debería mi organización seguir el flujo de trabajo de incorporación y concesión de acceso sin realizar cambios en él?	24
Recursos	25
Historial de documentos	26
Glosario	27
#	27
A	28
B	31
C	33
D	36

E	41
F	43
G	44
H	45
I	46
L	49
M	50
O	54
P	57
Q	60
R	60
S	63
T	67
U	68
V	69
W	69
Z	70
.....	lxxii

Diseñar un lago de datos para crecer y escalar en el Nube de AWS

Wei Shao, Amazon Web Services (AWS)

Octubre de 2021 ([historial del documento](#))

Las Organizaciones diseñan y crean cada vez más lagos de datos en la nube de Amazon Web Services (AWS) como parte de su estrategia de modernización de datos. Los lagos de datos pueden ser repositorios que almacenan los datos estructurados y no estructurados a cualquier escala y los ponen a disposición de una amplia gama de usuarios internos y externos.

Sin embargo, incorporar datos en lagos de datos que sirvan a los casos de uso de su empresa en crecimiento puede llevar mucho tiempo y esfuerzo. Para ayudar a reducir los costos y maximizar el valor que se genera a partir de los datos, muchas organizaciones planifican una ingestión única de datos en su lago de datos y, a continuación, consumen estos datos varias veces. Diseñar una arquitectura de lago de datos que se amplíe con la producción, el intercambio y el consumo de datos es fundamental para ofrecer valor a las partes interesadas de la empresa cuando su lago de datos crece.

Una arquitectura de lago de datos escalable proporciona a su organización una base sólida para obtener valor de su lago de datos y, al mismo tiempo, incorporar más datos. Al obtener información sobre los datos de forma continua sin ralentizarse ni interrumpirse debido a las restricciones de escalabilidad, un lago de datos escalable también ayuda a su organización a seguir siendo competitiva.

Por lo general, un lago de datos tiene productores y consumidores de datos. Los productores de datos crean activos de datos mediante la recopilación, el procesamiento y el almacenamiento de datos de su dominio de datos. Estos activos de datos colectivos forman el contenido de su lago de datos. Los productores de datos pueden optar por compartir selectivamente sus activos de datos con los consumidores de datos del lago de datos.

Los consumidores de datos necesitan los datos de los productores de datos para cumplir con sus casos de uso empresarial y, en ocasiones, también pueden combinar estos datos con sus propios datos. Los productores y consumidores de datos suelen formar parte de su organización, aunque no siempre. Es importante destacar que ambos pueden ser productores de datos o consumidores de datos al mismo tiempo.

Una arquitectura de lago de datos escalable le ayuda a lograr los siguientes resultados:

- Incorpore a los productores de datos a gran escala sin necesidad de que mantengan todo el proceso de intercambio de datos. Esto ayuda a los productores de datos a incorporar sus datos en el lago de datos y a centrarse en recopilar, procesar y almacenar datos de su dominio de datos.
- Permita que los consumidores de datos accedan a los datos de varios productores de datos sin aumentar los costos generales ni los gastos generales de administración.

Esta guía describe los desafíos de escalado comunes que pueden surgir cuando las organizaciones hacen crecer su lago de datos, proporciona una arquitectura de referencia de lagos de datos y describe los enfoques para incorporar y conceder acceso a los productores y consumidores de datos. La arquitectura de referencia del lago de datos de esta guía aprovecha las diferentes características y capacidades que ofrece [AWS Lake Formation](#). La guía está dirigida a los equipos responsables de diseñar lagos de datos en el Nube de AWS, incluidos los arquitectos de datos empresariales, los arquitectos de plataformas de datos, los diseñadores o los líderes de dominios de datos.

Resultados comerciales específicos

Debe esperar los siguientes tres resultados después de diseñar un lago de datos para crecer y escalar en Nube de AWS:

- Reduzca los gastos generales de uso compartido y consumo de datos en varias líneas de negocio de su organización.
- Un enfoque seguro y coherente que ayuda a su organización a incluir a productores de datos externos y a compartir datos con ellos en su lago de datos.
- Obtenga información sobre los datos de forma continua sin ralentizarse ni interrumpirse debido a las restricciones de escalabilidad.

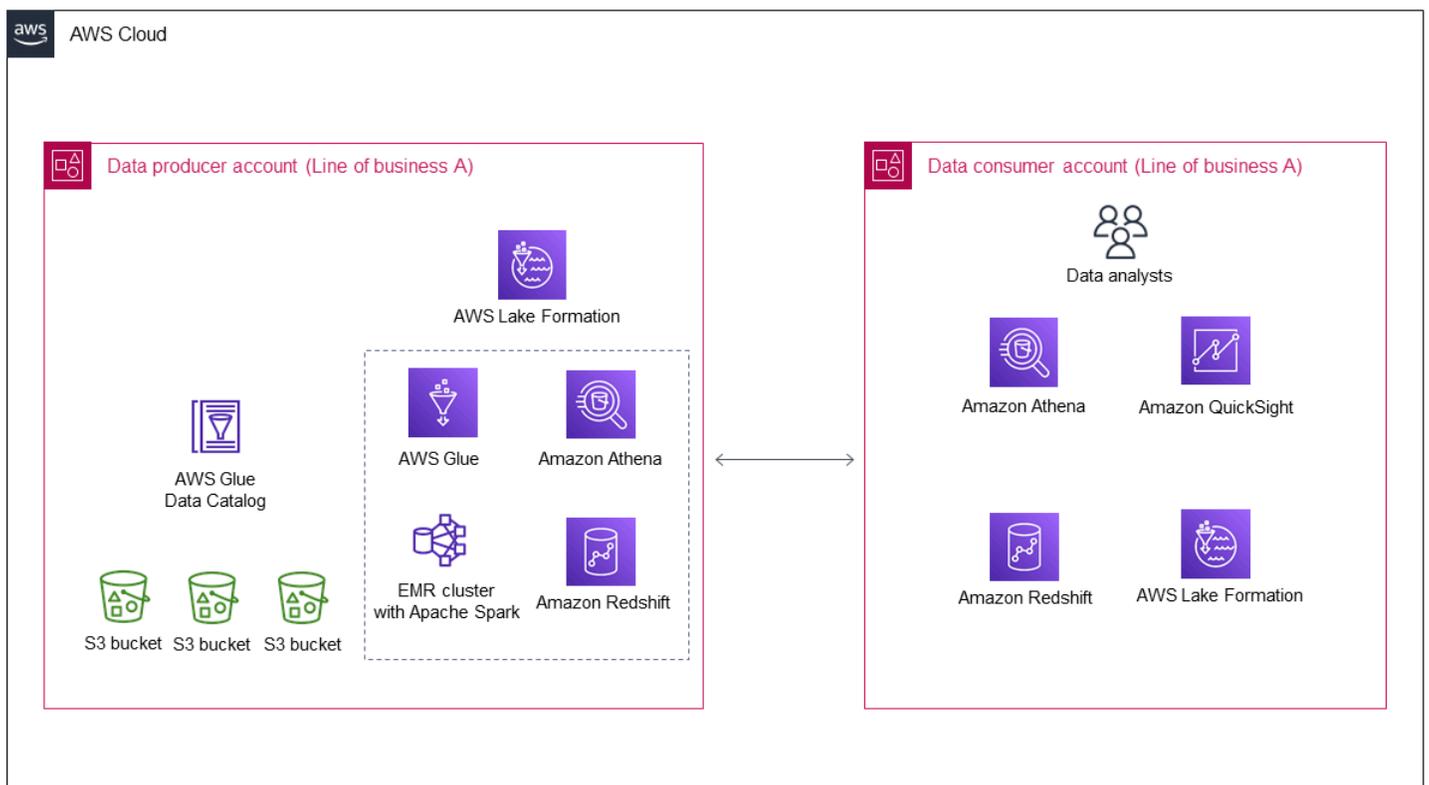
Desafíos comunes de escalado

Un lago de datos pasa por varias etapas cuando sus datos crecen después de la implementación inicial. Si no utilizó una arquitectura escalable para diseñar su lago de datos, su organización podría enfrentarse a desafíos y quedar en desventaja por el crecimiento del lago de datos.

En las siguientes secciones se explica cómo el crecimiento de un lago de datos típico puede provocar desafíos de escalado.

Despliegue inicial del lago de datos

El siguiente diagrama muestra la arquitectura de un lago de datos tras su implementación inicial por parte de la línea de negocio A.



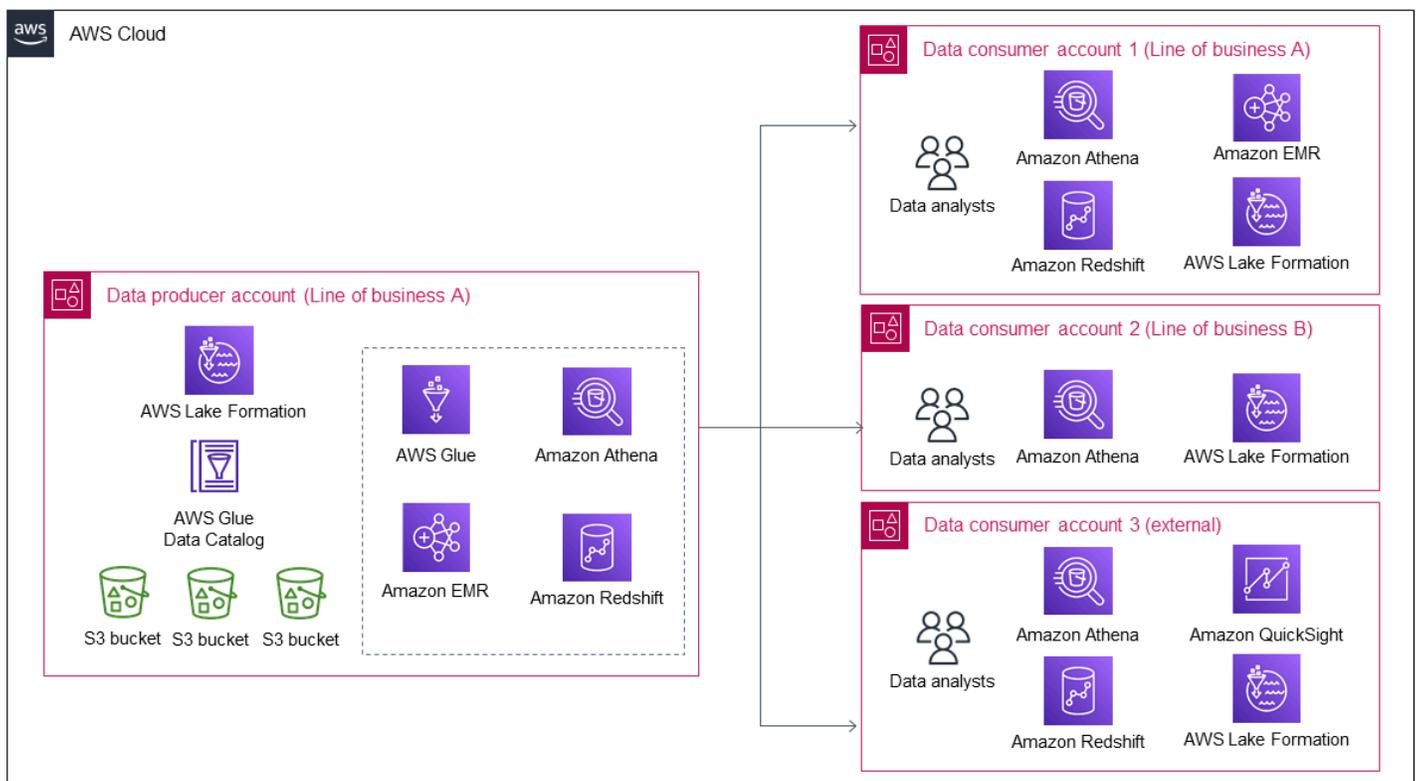
En el diagrama se muestra los siguientes componentes:

- La cuenta del productor de datos recopila y procesa datos, almacena los datos procesados y los prepara para el consumo.
- Los datos de la cuenta del productor de datos se almacenan en [bucket de Amazon Simple Storage Service \(Amazon S3\)](#), que pueden tener varias capas de datos.

- Puede utilizar AWS servicios para el procesamiento de datos (por ejemplo, [AWS Glue](#) [Amazon EMR](#)).
- El productor de datos no solo produce y almacena datos en el lago de datos, sino que también debe decidir qué datos compartir con un consumidor de datos y cómo compartirlos. [AWS Lake Formation](#) gestiona el lago de datos en la cuenta del productor de datos, además de gestionar el intercambio de datos entre cuentas del productor de datos al consumidor de datos.
- La cuenta de consumidor de datos consume los datos compartidos de la cuenta del productor de datos para casos de uso empresarial específicos.

Aumentan los consumidores de datos

El siguiente diagrama muestra que se introducen más datos al lago de datos cuando aumentan los datos de la línea de negocios A. El lago de datos atrae entonces a más consumidores de datos para aprovechar los datos y obtener valor de ellos.

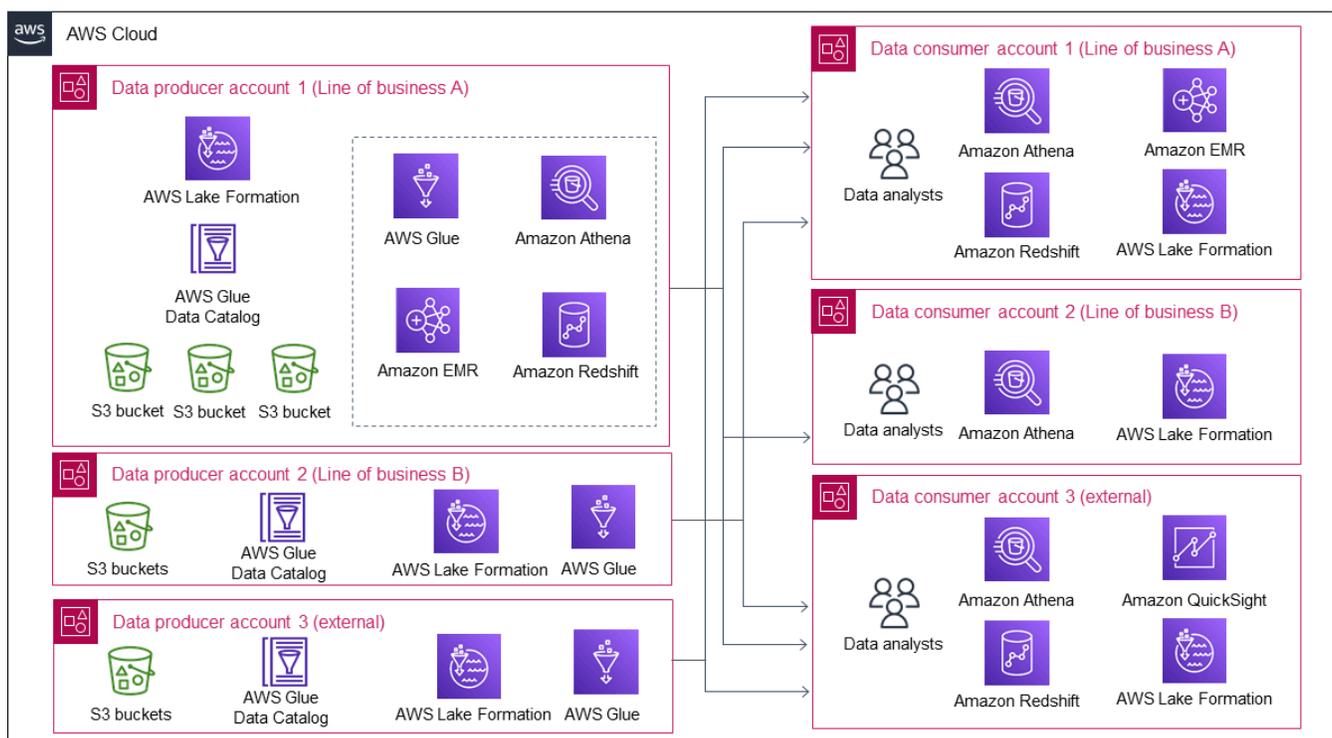


El diagrama muestra cómo una organización genera valor casi continuo a partir de un activo de datos existente y cómo esto atrae a más consumidores de datos. Sin embargo, cuando los consumidores de datos aumentan, el productor de datos solo tiene las dos opciones siguientes para adaptarse a este crecimiento:

- Administre manualmente el intercambio y el acceso a los datos por parte de los consumidores de datos individuales, lo cual no es un enfoque escalable.
- Desarrolle un proceso automatizado o semiautomatizado para compartir datos y gestionar el acceso a los datos. Si bien esta opción podría ser escalable, su diseño y creación requieren mucho tiempo y esfuerzo, ya que los consumidores de datos internos y externos tienen diferentes requisitos de control de seguridad. En el futuro, también se necesitaría tiempo y esfuerzo adicionales para cualquier mejora de la solución.

Aumentan los productores de datos

El siguiente diagrama muestra la arquitectura del lago de datos cuando varias líneas de negocio se unen como productores de datos.



La arquitectura del lago de datos se vuelve cada vez más complicada, incluso con solo tres productores de datos y tres consumidores de datos.

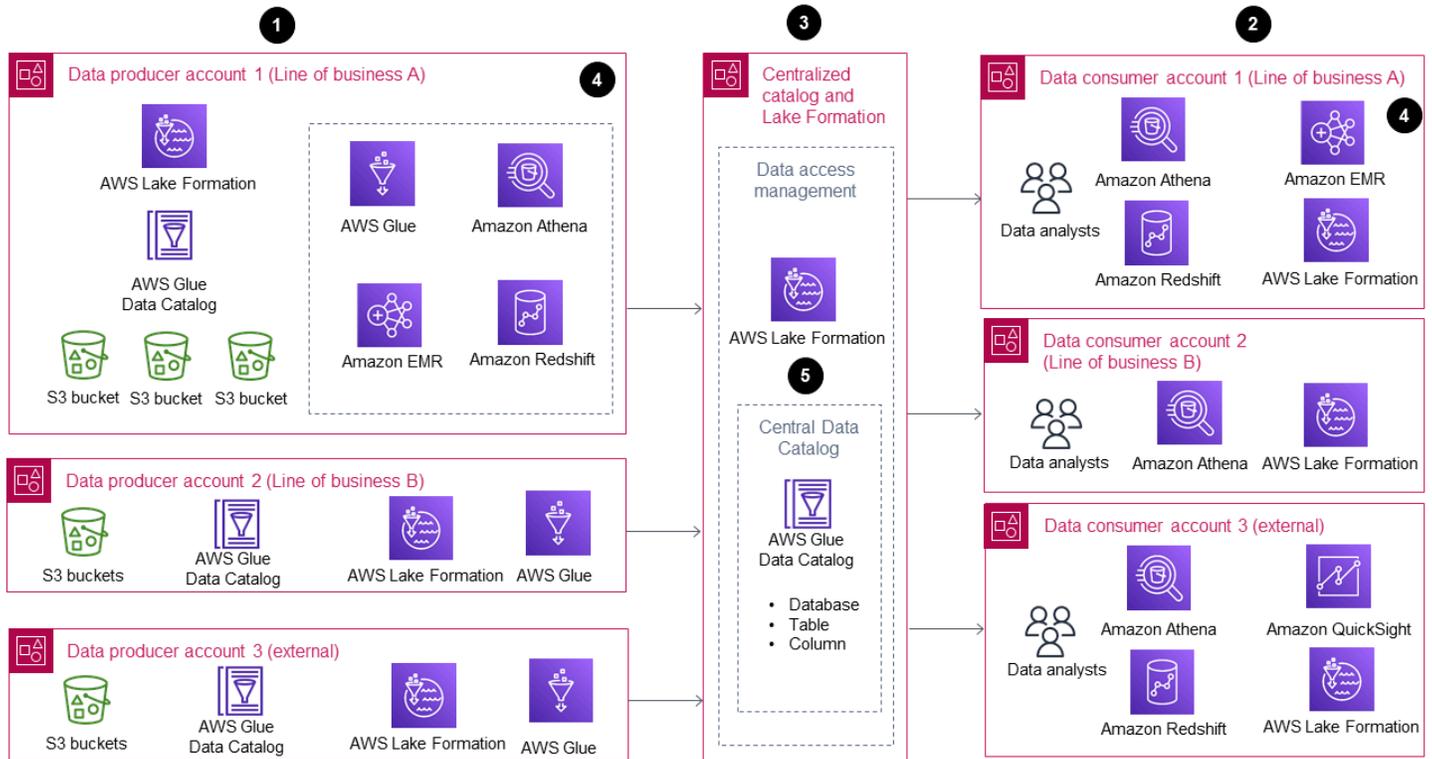
Cada productor de datos debe gestionar el intercambio de datos y la gestión del acceso a los datos para varios consumidores de datos. No es realista esperar que todos los productores de datos desarrollen un proceso automatizado o semiautomatizado para el intercambio de datos y la gestión

del acceso a los datos. Algunos productores de datos pueden optar por no compartir sus datos y, por lo tanto, evitar una sobrecarga de administración inasequible. Del mismo modo, cada consumidor de datos debe interactuar con varios productores de datos para entender sus diferentes procesos de consumo de datos. Esto significa que los consumidores de datos individuales se enfrentan a una sobrecarga de administración cada vez mayor para gestionar diferentes patrones de intercambio de datos.

En muchas organizaciones, este lago de datos provoca cuellos de botella y no puede crecer ni escalar. Esto puede significar que su organización debe rediseñar y reconstruir su lago de datos para eliminar el embotellamiento, lo que puede costar mucho tiempo, recursos y dinero.

Arquitectura de referencia

El siguiente diagrama muestra la arquitectura de referencia de esta guía para hacer crecer y escalar un lago de datos en Nube de AWS.



En el diagrama se muestra los siguientes componentes:

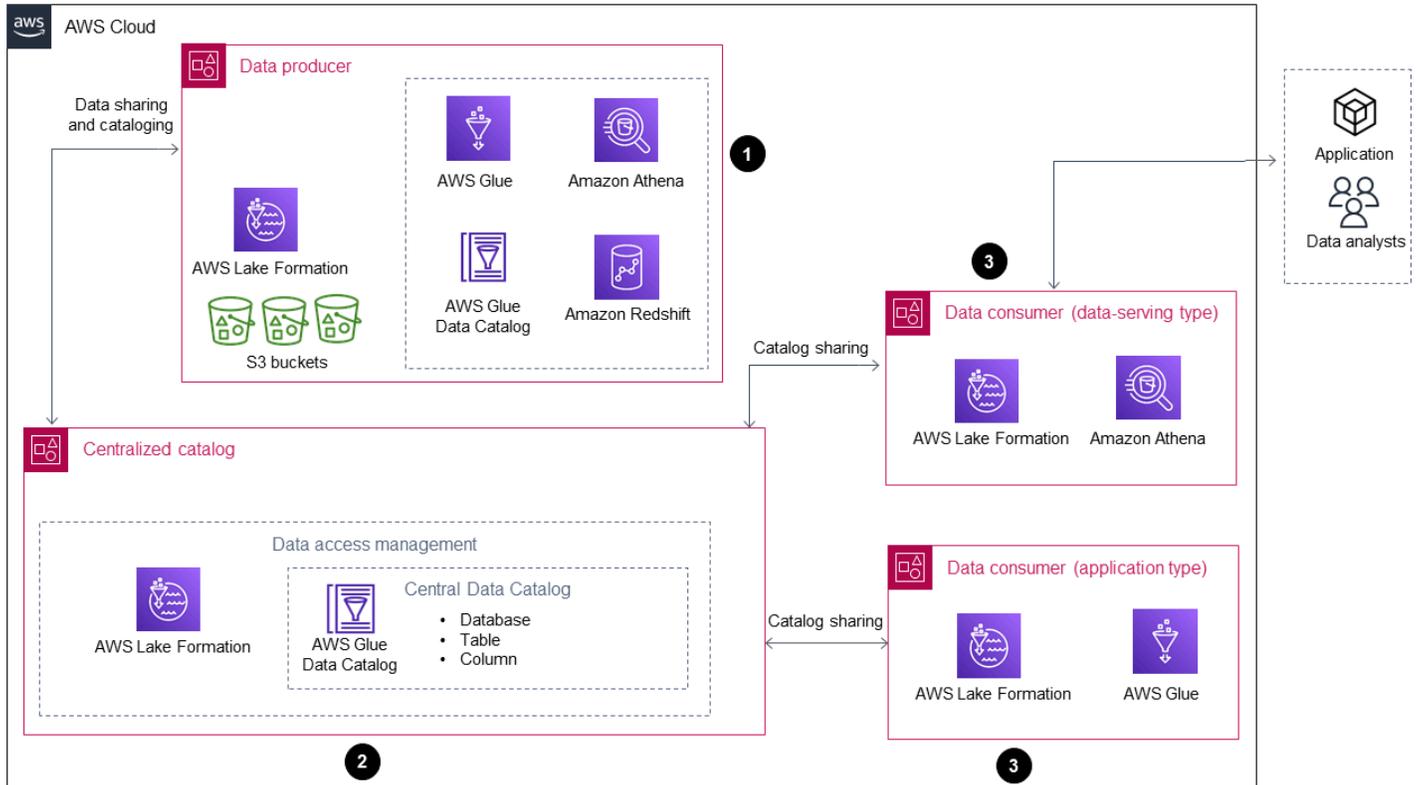
1. Una capa de producción de datos es diferente Cuentas de AWS.
2. Una capa de consumidores de datos es diferente Cuentas de AWS.
3. Un catálogo centralizado en un Cuenta de AWS.
4. Si bien cada línea de negocio solo tiene un productor de datos y un consumidor de datos, la arquitectura de referencia de la guía es compatible con varios productores y consumidores de datos para cada línea de negocio. Es habitual integrar a un productor de datos con uno o varios consumidores de datos que incluyan tanto el servicio de datos como los tipos de aplicaciones. Para obtener más información al respecto, consulte la [Componentes de la arquitectura de referencia](#) sección de esta guía.
5. El catálogo centralizado es la interfaz que utilizan los productores y consumidores de datos para compartir y consumir datos.

El enfoque de la arquitectura de referencia permite estandarizar el uso compartido y el consumo de datos y escalar de forma independiente a los productores y consumidores de datos sin aumentar la sobrecarga de administración. La arquitectura de referencia también permite la producción y distribución de datos entre diferentes productores de datos. Cualquier productor de datos puede formar parte del lago de datos, compartir sus datos y contribuir al valor general que proporciona el lago de datos.

Este enfoque permite a su organización recopilar el valor de los datos en todas sus líneas de negocio y en los propietarios de datos externos, sin provocar un embotellamiento al limitar la recopilación y el procesamiento de datos en una sola canalización.

Componentes de la arquitectura de referencia

El siguiente diagrama muestra los tres componentes de la arquitectura de referencia de esta guía: los productores de datos, los consumidores de datos y el catálogo centralizado.



En el diagrama se muestran los siguientes componentes:

1. Una cuenta de productor de datos.
2. Un catálogo centralizado.
3. Dos cuentas de consumidores de datos, que incluyen tipos de aplicaciones y de servicio de datos.

Note

El diagrama muestra solo un productor de datos y dos consumidores de datos, pero la arquitectura de referencia de esta guía es compatible con varios productores y consumidores de datos cuando el lago de datos crece.

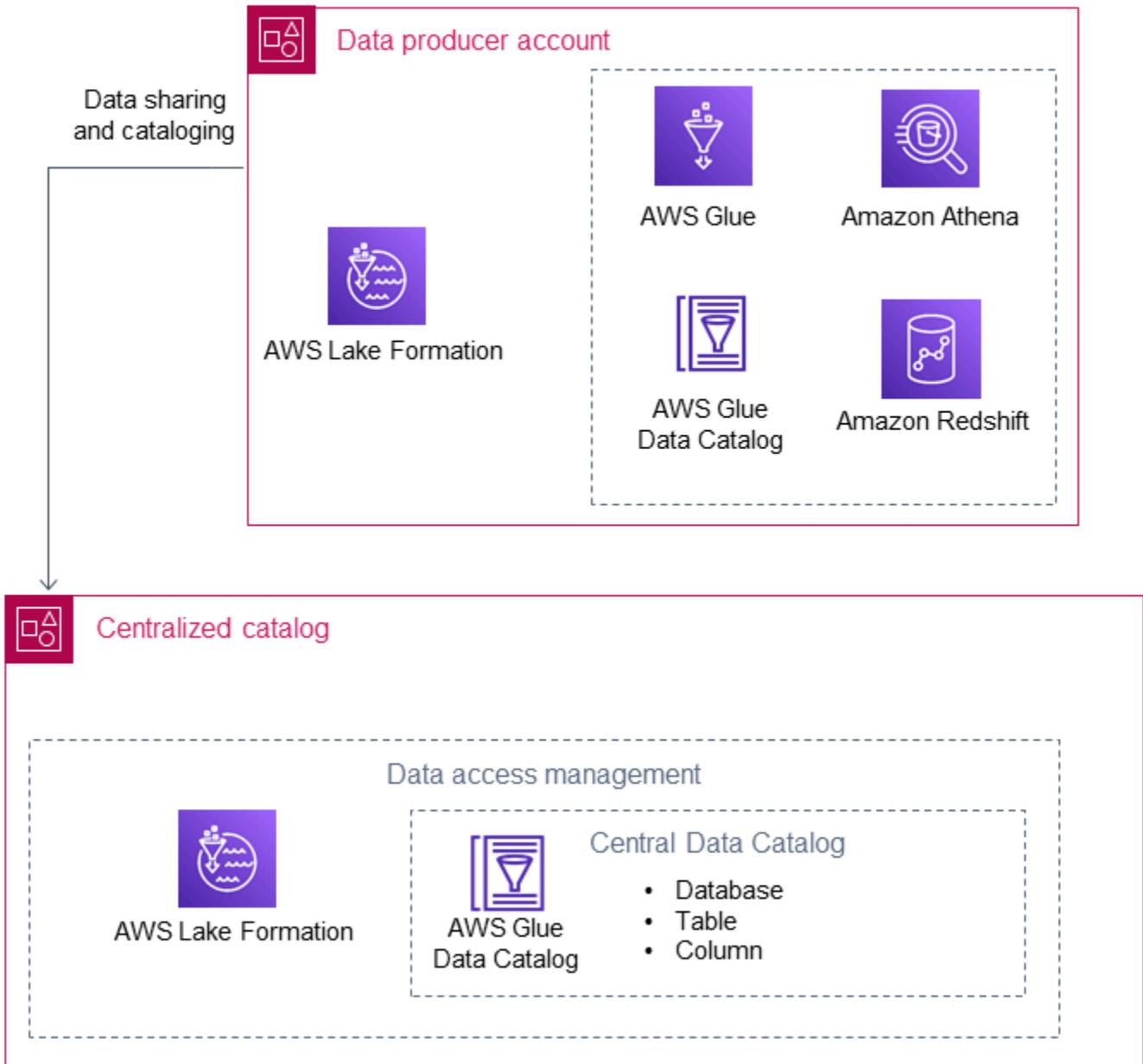
Comprender cada componente ayuda a su organización a implementar patrones reutilizables y escalar su lago de datos al incorporar de manera independiente y coherente a nuevos productores y consumidores de datos.

En las siguientes secciones se analizan los componentes de la arquitectura de referencia:

- [Productores de datos](#)
- [Catálogo centralizado](#)
- [Consumidores de datos](#)

Productores de datos

Un productor de datos recopila, procesa y almacena datos de su dominio de datos, además de monitorear y garantizar la calidad de sus activos de datos. El siguiente diagrama muestra la cuenta del productor de datos como componente de la arquitectura de referencia de esta guía.



Cada productor de datos tiene un catálogo de datos privado gestionado por AWS Lake Formation en su Cuenta de AWS propio proceso de datos interno. Los productores de datos proporcionan al catálogo centralizado permisos selectivos para sus datos, lo que significa que Lake Formation, en la cuenta del catálogo centralizado, puede acceder a los datos que el productor de datos desea compartir.

Esto significa que los productores de datos no interactúan directamente con los consumidores de datos. En cambio, la cuenta del productor de datos y su ubicación de almacenamiento de datos están completamente abstractas y ocultas al consumidor de datos. Este enfoque reduce los costos al

eliminar los gastos innecesarios para los productores de datos que experimentan un aumento en sus consumidores de datos.

Un cambio en la ubicación de los datos del productor de datos no afecta al consumidor de datos si la nueva ubicación de datos está registrada en el catálogo centralizado. Si el productor de datos quiere dejar de compartir un dato en particular, puede eliminar los permisos del catálogo centralizado. Esto evita que los consumidores de datos accedan a los datos y elimina la necesidad de revocar manualmente el acceso para cada consumidor de datos.

Al utilizar catálogos de datos públicos y privados, los productores de datos pueden elegir qué compartir con los consumidores de datos y, al mismo tiempo, gestionar de forma independiente el acceso interno a los datos a través de un catálogo de datos privado.

En la siguiente tabla se describen los dos AWS servicios que los productores de datos utilizan para compartir datos con el catálogo centralizado.

[Amazon Simple Storage Service \(Amazon S3\)](#)

Ajuste la política de depósitos para los depósitos de S3 para proporcionar acceso a los datos a las funciones AWS Identity and Access Management (IAM) del catálogo centralizado.

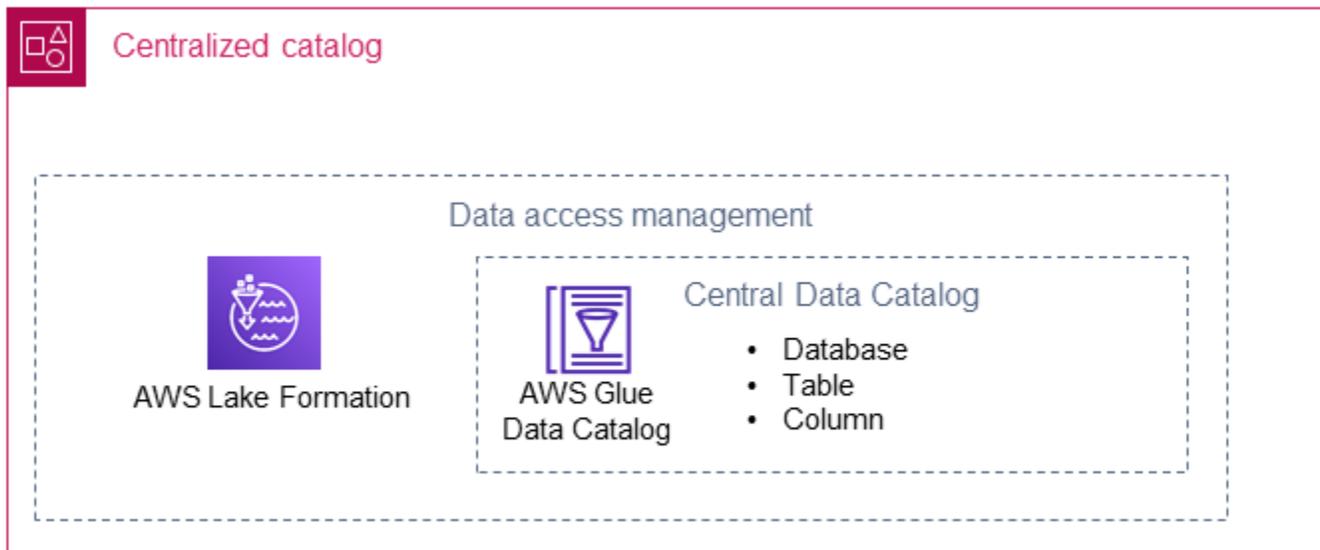
Los productores de datos también pueden compartir los datos almacenados en otras aplicaciones o servicios de datos mediante Amazon S3 como capa de datos intermedia.

[AWS Key Management Service \(AWS KMS\)](#)

Proporcione permisos para las claves AWS gestionadas de las funciones de IAM en el catálogo centralizado y las que AWS KMS keys se utilizan para cifrar los datos compartidos de Amazon S3 en las cuentas del productor de datos.

Catálogo centralizado

El siguiente diagrama muestra cómo el catálogo centralizado conecta a los productores de datos con los consumidores de datos en el lago de datos.



El catálogo centralizado almacena y administra el catálogo de datos compartidos para las cuentas de los productores de datos. El catálogo centralizado también aloja los metadatos técnicos de los datos compartidos (por ejemplo, el nombre de la tabla y el esquema) y es el lugar al que los consumidores de datos acuden para acceder a los datos.

Los consumidores de datos pueden acceder a los datos de varios productores de datos del catálogo centralizado y, a continuación, mezclar estos datos con sus propios datos para su posterior procesamiento. El uso de un catálogo centralizado elimina la necesidad de que los consumidores de datos se conecten directamente con diferentes productores de datos y reduce la sobrecarga operativa.

Dado que el catálogo centralizado ofrece visibilidad sobre el intercambio y el consumo de datos por parte de los productores y consumidores de datos, puede ser un lugar ideal para aplicar sus funciones de gobernanza de datos centralizada (por ejemplo, la auditoría de acceso).

En las siguientes secciones se describe cómo utiliza AWS Lake Formation y el catálogo centralizado AWS Glue.

AWS Lake Formation

[AWS Lake Formation](#) ayuda a crear bases de datos de AWS Glue en un catálogo de datos que apuntan a las ubicaciones de varios productores de datos en su lago de datos. Se crea un rol AWS Identity and Access Management (IAM) para Lake Formation en el catálogo centralizado. Al usar Lake Formation, el catálogo centralizado puede compartir recursos de datos de forma selectiva (por ejemplo, bases de datos, tablas o columnas) con los consumidores de datos. Los recursos administrados de

Lake Formation se comparten con los consumidores de datos utilizando uno de los dos métodos siguientes:

- [Método de recursos con nombre](#) asignado: este método comparte los recursos administrados entre las cuentas. Se deben especificar las bases de datos, tablas o nombres de columnas y se puede compartir un recurso con una organización, unidad organizativa (OU) o Cuenta de AWS. Para reducir la sobrecarga de uso compartido y administración, le recomendamos que comparta los recursos en niveles superiores siempre que sea posible (por ejemplo, en una organización o unidad organizativa en lugar de en una Cuenta de AWS). Sin embargo, debe asegurarse de que este enfoque cumpla con los requisitos de control de seguridad de datos de su organización.
- Nota: Este método funciona bien para los consumidores de datos con un tipo de aplicación, en el que los AWS servicios consumen datos del productor de datos. El requisito de acceso a los datos de este tipo de consumidor de datos está basado en la aplicación, es prescriptivo y es relativamente estático.
- [Método de control de acceso basado en etiquetas \(LF-TBAC\) de Lake Formation](#): el LF-TBAC es particularmente útil para los consumidores de datos con un tipo de servidor de datos. Sin embargo, los recursos etiquetados con Lake Formation actualmente solo se pueden compartir a Cuenta de AWS nivel y no a nivel de organización ni de unidad organizativa.

AWS Glue

Debe crear bases de datos AWS Glue para cada productor de datos de su catálogo centralizado. Dado que el catálogo centralizado AWS Glue aloja bases de datos de todos los productores de datos, debe asegurarse de que el nombre de la base de datos sea único en todos los productores de datos y de que refleje al productor de datos y su tipo de datos. Por ejemplo, puede utilizar la siguiente estructura de nomenclatura de bases de datos: <Data_Producer>-<Environment>-<Data_Group>

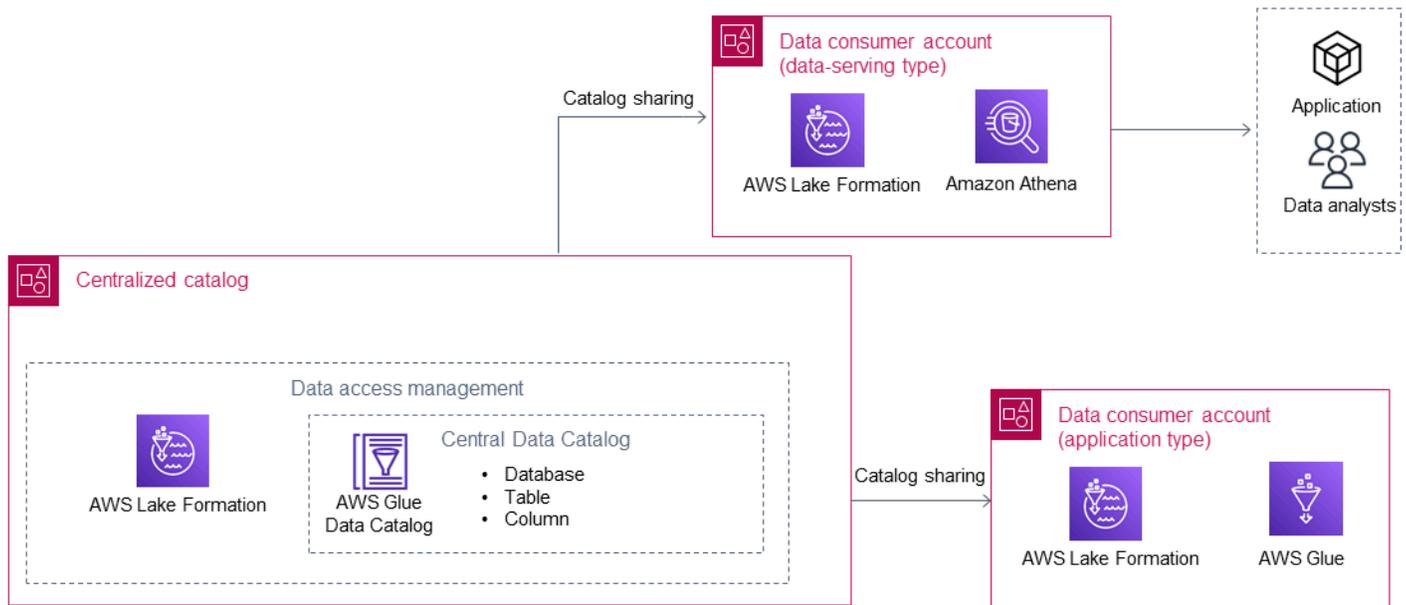
- <Data_Producer>— El nombre del productor de datos.
- <Environment>— El entorno del lago de datos, dev por ejemplo, un entorno de desarrollo, sit un entorno de pruebas de integración de sistemas o prod un entorno de producción.
- <Data_Group>— El nombre del grupo de datos que se usa para separar los datos de un productor de datos en grupos lógicos. Puede utilizar el nombre, el identificador o la abreviatura del sistema de origen como nombre. Una descripción de la base de datos también ayuda a describir el contenido y el propósito de la base de datos.

Puede utilizar unAWS Glue rastreador de los datos del productor de datos para mantener su esquema en la base de datos del catálogo centralizado. Si un productor de datos crea datos regularmente con la misma frecuencia, puede utilizar un únicoAWS Glue rastreador. En todos los demás casos, debe utilizar variosAWS Glue rastreadores para adaptarse a diferentes frecuencias de rastreo. Según el caso de uso empresarial, el rastreador puede programarse para una frecuencia predefinida o iniciarse mediante eventos.

También puede mantener el esquema de la tablaAWS Glue llamando a laAWS Glue API para crear o actualizar el esquema. Si bien esto puede proporcionar flexibilidad, se requiere un esfuerzo adicional para el desarrollo y el mantenimiento del código. Asegúrese de evaluar el caso de uso y el valor empresarial y, a continuación, elija la opción que cumpla con sus requisitos y tenga menos gastos generales.

Consumidores de datos

Los consumidores de datos consumen los datos del productor de datos después de que el catálogo centralizado los comparta utilizandoAWS Lake Formation. En el siguiente diagrama se muestran dos consumidores de datos en el lago de datos.



Hay dos tipos de consumidores de datos: aplicaciones y servidores de datos. En la tabla siguiente, se describen estos dos tipos.

Tipo de aplicación	Los consumidores de datos de aplicaciones ejecutan las aplicaciones por su cuenta
--------------------	---

ntas de AWS. Las aplicaciones consumen las funciones AWS Identity and Access Management (IAM) para acceder a los datos compartidos de un productor de datos y, a continuación, los procesan de acuerdo con su lógica.

Por lo general, este tipo de consumidor de datos tiene requisitos de datos prescriptivos para cumplir con las necesidades de una aplicación.

Tipo de servicio de datos

Los consumidores de datos que sirven datos suelen estar destinados a personas (por ejemplo, analistas de datos o científicos de datos) y aplicaciones (por ejemplo, una aplicación de inteligencia empresarial) que no tienen las suyas propias Cuentas de AWS.

Pueden existir varios consumidores de datos que sirven datos en el lago de datos de una organización. Por ejemplo, diferentes líneas de negocio podrían optar por configurar sus propios consumidores de datos de servicio de datos para ayudar a los usuarios a consumir los datos del lago de datos. Estos consumidores de datos tienen sus propios roles de IAM configurados en sus roles Cuenta de AWS (por ejemplo, los roles de IAM asociados a ellos [AWS IAM Identity Center](#)) que los utilizan los usuarios finales de la cuenta del consumidor de datos para acceder a los datos compartidos a través de AWS servicios (por ejemplo, [Amazon Athena](#)).

Por lo general, este tipo de consumidor de datos tiene requisitos de datos amplios y en continuo aumento.

AWS Lake Formations es el AWS servicio más importante que utiliza un consumidor de datos para compartir datos entre cuentas y acceder al catálogo centralizado. Una vez que el catálogo centralizado comparte las bases de datos, los recursos compartidos están disponibles en Lake Formation en la cuenta del consumidor de datos. A continuación, se puede conceder el acceso a los datos a los directores de IAM locales en la cuenta del consumidor de datos, con el permiso del productor de datos, si es necesario. Los datos compartidos pueden luego ser utilizados por AWS servicios integrados con Lake Formation (por ejemplo, Amazon Athena y AWS Glue). Puede utilizar los siguientes AWS servicios para acceder a los datos compartidos en la cuenta de consumidor de datos:

- [Amazon Athena](#) es un servicio interactivo de consultas que ayuda a analizar datos directamente en Amazon Simple Storage Service (Amazon S3) usando SQL estándar. Para obtener más información sobre Athena y Lake Formation, consulte [Cómo accede Athena a los datos registrados en Lake Formation](#) en la documentación de Amazon Athena.
- [Amazon Redshift Spectrum](#) le ayuda a consultar y recuperar de manera eficiente los datos estructurados y semiestructurados de los archivos en Amazon S3 sin tener que cargar los datos en tablas de Amazon Redshift. Para obtener más información sobre el espectro de desplazamiento al rojo y Lake Formation, consulte [Uso del espectro de Redshift con Lake Formation](#) en la documentación de Amazon Redshift.
- [AWS Glue](#) es un servicio completamente administrado de extracción, transformación y carga (ETL) con el que resulta más rentable y sencillo categorizar los datos, limpiarlos, enriquecerlos y moverlos de manera fiable entre distintos almacenes y flujos de datos. La función de IAM asociada a un trabajo de AWS Glue ETL puede acceder a los datos del lago de datos gestionados por Lake Formation si tiene los permisos de acceso necesarios.
- [Amazon EMR](#) ayuda a ejecutar marcos de macrodatos (por ejemplo, [Apache Hadoop](#) y [Apache Spark](#)) para procesar y analizar grandes cantidades de datos. Para obtener más información sobre Amazon EMR y Lake Formation, consulte [Integrar Amazon EMR con Lake Formation](#) en la documentación de Amazon EMR.
- [Amazon QuickSight](#) es un servicio de inteligencia empresarial escalable, sin servidor, integrable y basado en aprendizaje automático (ML) que puede utilizar para analizar y visualizar los datos de su lago de datos. Para obtener más información sobre Amazon QuickSight y Lake Formation, consulte [Autorizar conexiones a través de Lake Formation](#) en la QuickSight documentación de Amazon.
- [Amazon SageMaker Data Wrangler \(Data Wrangler\)](#) reduce el tiempo necesario para agregar y preparar los datos para el aprendizaje automático. Para obtener más información sobre Data Wrangler y Lake Formation, consulte [Preparar datos de aprendizaje automático con Amazon SageMaker Data Wrangler](#) en la SageMaker documentación de Amazon.

Incorporación y concesión de acceso

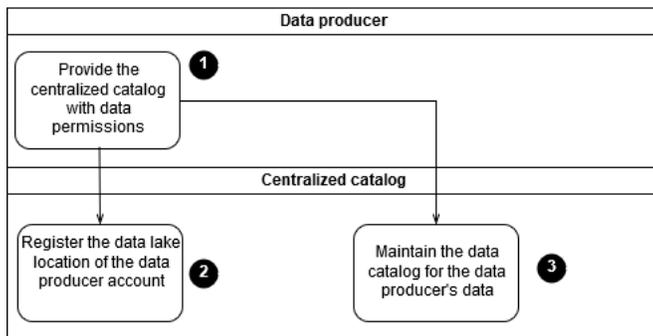
La arquitectura de referencia de lagos de datos de esta guía le ayuda a escalar de forma independiente a los productores y consumidores de datos, además de definir y establecer un proceso coherente para incorporar y conceder acceso a esos consumidores de datos.

En las siguientes secciones se describe el proceso de incorporación para los productores y consumidores de datos y cómo conceder el acceso a una cuenta de consumidor de datos. Esta guía utiliza el [método de recursos con nombre asignado](#) entre el catálogo centralizado y los consumidores de datos. El proceso para el [método LF-TBAC](#) es similar pero ligeramente diferente. Le recomendamos que evalúe y configure estos enfoques para cumplir con las prácticas y políticas de gobierno de datos de su organización.

Para obtener más información sobre estos dos métodos, consulte la [Catálogo centralizado](#) sección de esta guía.

Incorporación de productores de datos

El siguiente diagrama muestra cómo incorporar un nuevo productor de datos a su lago de datos.



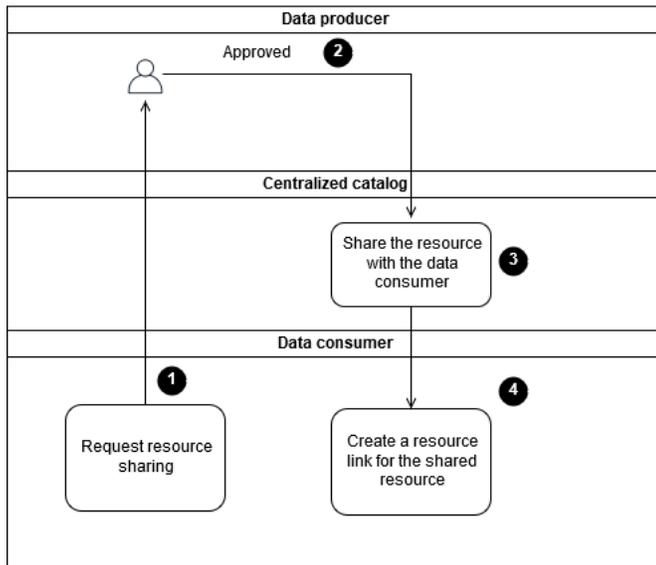
El diagrama muestra el siguiente proceso de incorporación:

1. El productor de datos proporciona acceso selectivo al catálogo centralizado a sus datos (por ejemplo, un bucket de Amazon Simple Storage Service (Amazon S3) y AWS KMS key). Se proporciona acceso a los directores del catálogo centralizado AWS Identity and Access Management (IAM) para registrar la ubicación del lago de datos del productor de datos AWS Lake Formation y a los principios de IAM que se utilizan para mantener el catálogo del productor de datos.
2. Registre la ubicación del lago de datos del productor de datos (por ejemplo, un depósito S3) que utilice la formación de lagos del catálogo centralizado.

3. Cree la base de datos, las tablas y los esquemas de tablas para los nuevos datos del productor de datos en el catálogoAWS Glue de datos.

Incorporación de consumidores de datos

El siguiente diagrama muestra cómo incorporar un nuevo consumidor de datos a su lago de datos.



El diagrama muestra el siguiente proceso de incorporación:

1. El consumidor de datos solicita autorización para ver los datos del productor de datos y especifica los datos a los que debe acceder.
2. El administrador de datos del productor de datos revisa la solicitud del consumidor de datos y evalúa si debe:
 - Comparta algunas o todas las tablas de las bases de datos solicitadas. Recomendamos compartir a nivel de base de datos cuando compartir todas las tablas con el consumidor de datos no implique implicaciones para la seguridad de los datos, lo que ayuda a evitar la sobrecarga de administración que supone compartir a nivel de tabla.
 - Comparta los datos a nivel de organización, unidad organizativa o cuenta del consumidor.
3. Cuando el productor de datos los aprueba, los recursos del catálogo de datos requeridos se comparten con el consumidor de datos en el catálogo centralizado.
4. Los enlaces a recursos se pueden crear en la cuenta del consumidor de datos mediante Lake Formation y, a continuación, dirigir a los recursos del catálogo de datos compartido del catálogo centralizado.

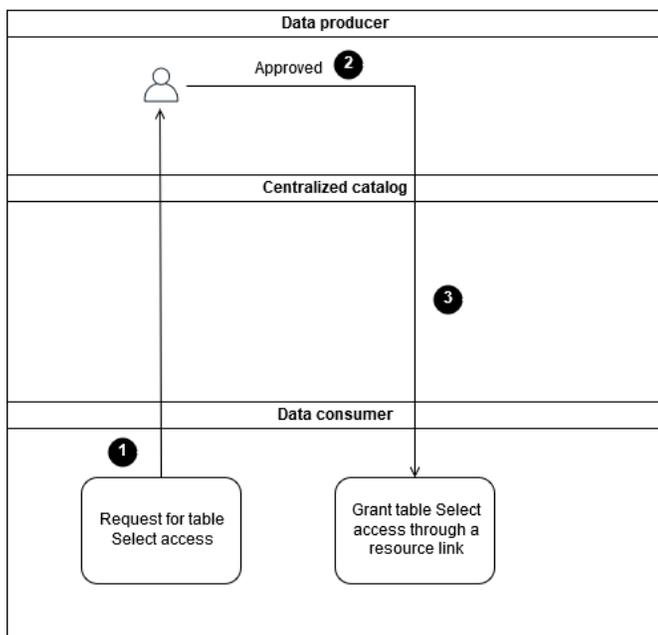
Una vez completado el proceso de incorporación, el administrador de Lake Formation del consumidor de datos puede ver el recurso del catálogo de la base de datos desde el catálogo centralizado y el enlace al recurso. En este momento, nadie más en la cuenta del consumidor de datos puede acceder a los datos del productor de datos.

Otorgar acceso a Select en una cuenta de consumidor de datos

El siguiente diagrama muestra el proceso para conceder [Select acceso a](#) los recursos de datos compartidos con un director de IAM local en la cuenta del consumidor de datos. El principal de IAM local puede ser el rol de IAM para usuarios individuales o un rol de IAM que consumen AWS servicios específicos.

Note

Cuando los datos que se comparten son de baja sensibilidad, puede delegar la concesión de acceso al propio consumidor de datos sin necesidad de la aprobación del productor de datos. Esto se debe a que la confianza y el intercambio ya están establecidos entre ellos.



En el diagrama se muestra el siguiente proceso:

1. El principal de IAM individual de la cuenta del consumidor de datos solicita el `Select` acceso al enlace de recursos desde el principal de IAM de la cuenta del consumidor de datos.

2. El administrador de datos del productor de datos revisa la solicitud del consumidor de datos y da su aprobación si se cumplen todos los requisitos.
3. Se le concede el acceso y esto permite que el director de IAM consuma los datos solicitados.

Preguntas frecuentes

Esta sección proporciona respuestas a las preguntas más frecuentes sobre el diseño de un lago de datos para crecer y escalar en el Nube de AWS.

¿Esta arquitectura de referencia del lago de datos es más aplicable a las organizaciones empresariales?

La arquitectura de referencia de lagos de datos de esta guía se puede aplicar a lagos de datos que pertenecen a organizaciones de cualquier tamaño. La arquitectura de referencia estandariza la interfaz de intercambio de datos, reduce la sobrecarga y el costo de mantener y hacer crecer el lago de datos, y se puede aplicar a cualquier escala a la que crezca el lago de datos de su organización.

¿Puedo seguir usando esta arquitectura de referencia si mi organización solo tiene un productor de datos?

La arquitectura de referencia del lago de datos de esta guía sigue siendo relevante y beneficiosa, incluso si su organización solo tiene un productor de datos. Sin el catálogo centralizado, el productor de datos tiene que gestionar el crecimiento de los consumidores de datos, lo que aumenta la complejidad y los gastos generales. Su lago de datos también es un activo a largo plazo para su organización y, por lo general, las organizaciones agregan más productores de datos. Por ejemplo, es posible que necesite un productor de datos adicional para almacenar datos confidenciales por motivos de cumplimiento o porque su organización adquiera otra unidad de negocio que tenga su propio productor de datos.

Mi lago de datos conecta directamente a un productor de datos con varios consumidores de datos. ¿Sigue siendo relevante la arquitectura de referencia del lago de datos de esta guía?

La arquitectura de referencia del lago de datos beneficiaría a su organización a largo plazo. Puede utilizar un enfoque de dos pasos y empezar por crear un catálogo centralizado para los nuevos consumidores de datos. A continuación, puede conectar sus consumidores de datos actuales al catálogo centralizado.

¿Debería mi organización seguir el flujo de trabajo de incorporación y concesión de acceso sin realizar cambios en él?

No, el objetivo principal de esa sección es ilustrar los bloques de actividad lógica requeridos durante el proceso de incorporación. Todas las organizaciones deben personalizar el proceso e incluso pueden tener varios procesos, según la confidencialidad de sus datos.

Otra consideración es que el flujo del proceso utiliza el enfoque de intercambio basado en recursos enAWS Lake Formation. Existen otros métodos de intercambio de datos compatibles con Lake Formation, como el intercambio basado en etiquetas, en los que las diferencias en el proceso se pueden adaptar al método de intercambio específico.

Recursos

- [AWS Glueguía para desarrolladores](#)
- [AWS Lake Formationguía para desarrolladores](#)
- [Acceso entre cuentas en Lake Formation](#)
- [Otorgar y revocar los permisos de catálogo de datos en Lake Formation](#)
- [Rellenar el catálogoAWS Glue de datos](#)

Historial de documentos

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

Cambio	Descripción	Fecha
—	Publicación inicial	1 de octubre de 2021

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.

antipatrones

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.

mallado 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](#) 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

laC

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.

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.