Aplicación del marco AWS Well-Architected para Amazon Neptune Analytics

AWS Guía prescriptiva

AWS Guía prescriptiva: Aplicación del marco AWS Well-Architected para Amazon Neptune Analytics

Table of Contents

Introducción 1
Destinatarios previstos 2
Objetivos 2
Pilar de excelencia operativa 3
Automatice la implementación mediante un enfoque de IaC 3
Diseño de operaciones 4
Realice cambios frecuentes, pequeños y reversibles 5
Implemente la observabilidad para obtener información procesable 5
Aprenda de todos los fallos operativos 6
Utilice las funciones de registro para supervisar la actividad no autorizada o anómala 6
Pilar de seguridad 8
Implementar la seguridad de los datos 9
Proteja sus redes 9
Implemente la autenticación y la autorización 10
Pilar de fiabilidad 11
Comprenda las cuotas de servicio de Neptune 11
Comprenda los patrones de despliegue de Neptune 12
Gestione y escale los clústeres de Neptune 13
Gestione las copias de seguridad y los eventos de conmutación por error 14
Pilar de eficiencia de rendimiento 15
Comprenda el modelado de gráficos para el análisis 15
Optimización de consultas 18
Optimice las escrituras 19
Gráficos del tamaño correcto 20
Pilar de optimización de costos 22
Comprenda los patrones de uso y los servicios necesarios 22
Seleccione los recursos prestando atención al costo 24
Pilar de sostenibilidad 26
Tenga en cuenta su Región de AWS selección 26
Optimice el consumo 26
Optimice los patrones de desarrollo y arquitectura del software 27
Recursos 29
Referencias 29
Publicaciones de blog y vídeos 29

Formación 29
Historial de documentos 30
Glosario 31
..... 31
A 32
B 35
C 37
D 40
E 45
F 47
G 49
H 50
I 52
L 54
M 55
O 60
P 63
Q 66
R 66
S 69
T 73
U 75
V 75
W 76
Z 77
..... lxxviii

Aplicación del marco AWS Well-Architected para Amazon Neptune Analytics

Michael Havey, Amazon Web Services ()AWS

Diciembre de 2024 (historial del documento)

Puede crear soluciones basadas en gráficos en Amazon Web Services (AWS) mediante Amazon Neptune. Neptune incluye Neptune Analytics, un motor de análisis de gráficos optimizado para la memoria que puede analizar rápidamente grandes cantidades de datos de gráficos para obtener información y encontrar tendencias. Puede realizar análisis de los datos de su clúster de base de datos de Neptune existente, o puede cargar y analizar datos de conjuntos de datos externos. Esta guía proporciona una guía prescriptiva para aplicar los principios del Marco AWS de Buena Arquitectura al planificar la implementación de Neptune Analytics. La aplicación del AWS marco Well-Architected para Amazon Neptune trata el mismo tema para una base de datos de Neptune.

El AWS Well-Architected Framework le ayuda a crear infraestructuras seguras, de alto rendimiento, resilientes y eficientes para una variedad de aplicaciones y cargas de trabajo. También proporciona un enfoque coherente para evaluar las arquitecturas e implementar diseños escalables.

El AWS Well-Architected Framework se basa en los seis pilares siguientes:

- Excelencia operativa
- Seguridad
- Fiabilidad
- Eficiencia del rendimiento
- Optimización de costos
- Sostenibilidad

Esta guía proporciona información sobre los pilares de diseño y las prácticas recomendadas de Well-Architected Framework, así como consideraciones que se deben tener en cuenta al implementar Neptune Analytics en. AWS

Destinatarios previstos

Esta guía está destinada a ingenieros de datos, arquitectos de soluciones y analistas de datos que diseñan e implementan soluciones basadas en gráficos. AWS

Objetivos

Esta guía puede ayudarle a usted y a su organización a hacer lo siguiente:

- Elija entre las opciones de implementación compatibles.
- Siga los patrones de diseño de AWS Well-Architected que le ayudan a mejorar la resiliencia y la seguridad.
- Diseñe sus consultas para obtener un rendimiento óptimo y ahorrar costes.
- Aprenda a ser eficiente desde el punto de vista operativo al gestionar su gráfico de Neptune Analytics en producción.

Pilar de excelencia operativa

El pilar de excelencia operativa del AWS Well-Architected Framework se centra en ejecutar y monitorear los sistemas, y en mejorar continuamente los procesos y procedimientos. Incluye la capacidad de respaldar el desarrollo y ejecutar las cargas de trabajo de manera eficaz, obtener información sobre su funcionamiento y mejorar continuamente los procesos y procedimientos de apoyo para ofrecer valor empresarial. Puede reducir la complejidad operativa mediante la autorreparación de las cargas de trabajo, que detectan y solucionan la mayoría de los problemas sin intervención humana. Puede trabajar para lograr este objetivo siguiendo las prácticas recomendadas que se describen en esta sección y utilizar las métricas y los mecanismos de Amazon Neptune Analytics para responder adecuadamente cuando su carga de trabajo se desvíe del comportamiento esperado. APIs

Este análisis del pilar de la excelencia operativa se centra en las siguientes áreas clave:

- Infraestructura como código (IaC)
- Administración de cambios
- Estrategias de resiliencia
- Administración de incidentes
- Informes de auditoría para garantizar el cumplimiento
- Registro y supervisión

Automatice la implementación mediante un enfoque de IaC

Entre las prácticas recomendadas para automatizar el despliegue en Neptune mediante IaC se incluyen las siguientes:

- Aplique IaC para implementar gráficos de Neptune Analytics y recursos relacionados. Para una configuración coherente del entorno, utilice el soporte para Neptune Analytics que proporciona para AWS CloudFormationaprovisionar gráficos y puntos finales privados.
- Se utiliza CloudFormation para aprovisionar instancias de bloc de notas de Neptune en Amazon SageMaker AI. Puede usar cuadernos para consultar y visualizar datos en un gráfico de Neptune Analytics.

- Cuando cree un gráfico de Neptune Analytics a partir de una fuente existente, como una instantánea o un clúster de base de datos de Neptune, o archivos de datos almacenados en Amazon Simple Storage Service (Amazon S3), supervise la tarea de importación masiva.
- Automatice los procedimientos operativos de Neptune Analytics, como cambiar el tamaño del gráfico, eliminar y crear una instantánea del gráfico, restaurar el gráfico a partir de una instantánea y restablecer y volver a cargar el gráfico. Utilice la API de Neptune Analytics, que está disponible a través de AWS Command Line Interface (AWS CLI) o SDK s.
- Evalúe el tiempo de actividad requerido de su gráfico. La analítica suele ser efímera; el gráfico solo es necesario durante el tiempo que se necesita para ejecutar los algoritmos. Si este es el caso, utilice AWS CLI o SDKs para hacer una instantánea del gráfico y eliminarlo cuando ya no lo necesite. A continuación, puede restaurarlo a partir de una instantánea más adelante, si es necesario.
- Almacene las cadenas de conexión de forma externa a su cliente. Puede almacenar las cadenas de conexión en AWS Secrets ManagerAmazon DynamoDB o en cualquier ubicación en la que se puedan cambiar de forma dinámica.
- Utilice etiquetas para añadir metadatos a sus recursos de Neptune Analytics y realice un seguimiento del uso en función de las etiquetas. Las etiquetas ayudan a organizar los recursos. Por ejemplo, puede aplicar una etiqueta común a los recursos de un entorno o aplicación específicos. También puede usar etiquetas para analizar la facturación del uso de los recursos; para obtener más información, consulte Organización y seguimiento de los costos mediante etiquetas de asignación de AWS costos en la Guía del usuario de AWS facturación. Además, puede utilizar condiciones en sus políticas AWS Identity and Access Management (de IAM) para controlar el acceso a AWS los recursos en función de las etiquetas utilizadas en esos recursos. Para ello, utilice la clave de condición global `aws:ResourceTag/tag-key`. Para obtener más información, consulte Controlar el acceso a AWS los recursos en la Guía del usuario de IAM.

Diseño de operaciones

Adopte enfoques para mejorar el funcionamiento de los gráficos de Neptune Analytics:

- Mantenga gráficos de Neptune Analytics separados para su uso en desarrollo, pruebas y producción. Estos gráficos pueden tener conjuntos de datos, usuarios y controles operativos diferentes.

- Mantenga gráficos de Neptune Analytics separados para diferentes usos. Por ejemplo, si dos grupos de usuarios analíticos requieren gráficos separados con plazos, modelos, rendimiento y disponibilidad SLAs y patrones de uso diferentes, mantenga gráficos separados para cada grupo.
- Prepare a los usuarios y al personal operativo para las actualizaciones de mantenimiento de Neptune Analytics.

Realice cambios frecuentes, pequeños y reversibles

Las siguientes recomendaciones se centran en los cambios pequeños y reversibles que puede realizar para minimizar la complejidad y reducir la probabilidad de que se interrumpa la carga de trabajo:

- Guarde las plantillas y los scripts de IaC en un servicio de control de código fuente como GitHub o GitLab.

> **Important**
>
> No almacene AWS las credenciales en el control de código fuente.

- Exija que las implementaciones de iAC utilicen un servicio de integración y entrega continuas (CI/CD), como o. AWS CodeDeployAWS CodeBuild Compila, prueba e implementa código en un entorno de Neptune Analytics que no sea de producción antes de promocionarlo a un gráfico de producción.

Implemente la observabilidad para obtener información procesable

Obtenga una comprensión integral del comportamiento, el rendimiento, la confiabilidad, el costo y el estado de las cargas de trabajo. Las siguientes recomendaciones le ayudarán a obtener ese nivel de comprensión de Neptune Analytics:

- Supervise CloudWatch las métricas de Amazon para Neptune Analytics. A partir de estas métricas, puede determinar el tamaño de un gráfico (número de nodos, bordes y vectores, más el tamaño total de bytes), el uso de la CPU y las tasas de solicitudes y errores de consulta.
- Cree CloudWatch paneles y alarmas para métricas clave como`NumQueuedRequestsPerSec`,, `NumOpenCypherRequestsPerSec` `GraphStorageUsagePercentGraphSizeBytes`, y

`CPUUtilization` también para las respuestas de los clientes de Neptune que se encuentran en los registros de sus aplicaciones.

- Configure las notificaciones para supervisar el estado del gráfico de Neptune Analytics, por ejemplo, cuando el tamaño del gráfico, la tasa de solicitudes o el uso de la CPU superen su umbral. Por ejemplo, si `GraphStorageUsagePercent` ha subido al 90 por ciento en un gráfico y tiene intención de crecer significativamente, decida si desea aumentar la capacidad de la Unidad de Capacidad de Neptuno (m-NCU) optimizada para la memoria. Si la m-NCU actual es de 128, aumentarla a 256 reducirá el almacenamiento en aproximadamente un 45 por ciento. Si `NumQueuedRequestsPerSec` suele ser superior a cero, considere la posibilidad de aumentar la capacidad de la m-NCU para ofrecer más capacidad de cómputo. Como alternativa, puede reducir la simultaneidad del lado del cliente.

Aprenda de todos los fallos operativos

Una infraestructura que se recupere automáticamente es un esfuerzo a largo plazo que se desarrolla de forma iterativa a medida que se producen problemas poco frecuentes o las respuestas no son tan eficaces como se desearía. La adopción de las siguientes prácticas impulsa la concentración hacia ese objetivo:

- Impulse la mejora aprendiendo de todos los fracasos.
- Comparta lo aprendido entre los equipos y la organización. Si varios equipos de su organización utilizan Neptune, cree una sala de chat o un grupo de usuarios común para compartir los aprendizajes y las mejores prácticas.

Utilice las funciones de registro para supervisar la actividad no autorizada o anómala

Utilice el registro para observar patrones anómalos de rendimiento y actividad. Tenga en cuenta las siguientes prácticas recomendadas:

- Neptune Analytics admite el registro de las acciones del plano de control mediante el uso de. AWS CloudTrail Para obtener más información, consulte Registrar las llamadas a la API de Neptune Analytics mediante. AWS CloudTrail A través de esta función, puede realizar un seguimiento de la creación, actualización y eliminación de los recursos de Neptune Analytics. Para una supervisión y alertas sólidas, también puede integrar CloudTrail eventos con Amazon CloudWatch Logs. Para

mejorar el análisis de la actividad del servicio Neptune Analytics e identificar los cambios en las actividades de un servidor Cuenta de AWS, puede consultar CloudTrail los registros mediante Amazon Athena. Por ejemplo, puede ejecutar consultas que identifiquen tendencias y aislar la actividad por atributos, como el usuario o la dirección IP de origen.

- También se puede utilizar CloudTrail para habilitar el registro de las actividades del plano de datos de Neptune Analytics, como las ejecuciones de consultas. Puede ver qué consultas se están ejecutando, su frecuencia y su origen. De forma predeterminada, CloudTrail no registra los eventos de datos. Se aplican cargos adicionales a los eventos de datos. Para obtener más información, consulte Precios de AWS CloudTrail.
- También puede registrar las llamadas de las aplicaciones a Neptune Analytics en el plano de control o en el plano de datos. Por ejemplo, si lo usa AWS SDK para Python (Boto3)para realizar consultas, puede habilitar el registro a nivel de depuración para obtener un seguimiento de las consultas en la consola o el archivo. Esto es útil durante el desarrollo. También le recomendamos que capture y registre las excepciones de su aplicación.

Pilar de seguridad

La seguridad en la nube es la máxima prioridad en AWS. Como AWS cliente, usted se beneficia de una arquitectura de centro de datos y red diseñada para cumplir con los requisitos de las organizaciones más sensibles a la seguridad. La seguridad es una responsabilidad compartida entre usted y AWS. El modelo de responsabilidad compartida la describe como seguridad de la nube y seguridad en la nube:

- Seguridad de la nube: AWS es responsable de proteger la infraestructura que se ejecuta Servicios de AWS en la. Nube de AWS AWS también le proporciona servicios que puede utilizar de forma segura. Los auditores externos prueban y verifican periódicamente la eficacia de la AWS seguridad como parte de los programas de AWS cumplimiento. Para obtener más información sobre los programas de conformidad que se aplican a Neptune, consulte AWS Servicios incluidos en el ámbito de aplicación por programa de conformidad.
- Seguridad en la nube: su responsabilidad viene determinada por lo Servicio de AWS que utilice. También es responsable de otros factores, incluida la confidencialidad de los datos, los requisitos de la empresa y la legislación y los reglamentos aplicables. Para obtener más información sobre la privacidad de los datos, consulte la sección Privacidad de datos FAQs. Para obtener información sobre la protección de datos en Europa, consulte el modelo de responsabilidad AWS compartida y la entrada del blog sobre el RGPD.

El pilar de seguridad de AWS Well-Architected Framework le ayuda a entender cómo aplicar el modelo de responsabilidad compartida cuando utiliza Neptune Analytics. En los temas siguientes se explica cómo configurar Neptune Analytics para cumplir sus objetivos de seguridad y conformidad. También aprenderá a utilizar otros Servicios de AWS que le ayuden a supervisar y proteger sus recursos de Neptune Analytics. El pilar de seguridad incluye las siguientes áreas de enfoque clave:

- Seguridad de los datos
- Seguridad de la red
- Autenticación y autorización

Implementar la seguridad de los datos

Las filtraciones y filtraciones de datos ponen en riesgo a sus clientes y pueden tener un impacto negativo sustancial en su empresa. Las siguientes prácticas recomendadas ayudan a proteger los datos de sus clientes de una exposición inadvertida o malintencionada:

- Los nombres de los gráficos, las etiquetas, las funciones de IAM y otros metadatos no deben contener información confidencial o delicada, ya que esos datos pueden aparecer en los registros de facturación o diagnóstico.
- URIs o los enlaces a servidores externos almacenados como datos en Neptune no deben contener información sobre credenciales para validar las solicitudes.
- Un gráfico de Neptune Analytics está cifrado en reposo. Puede utilizar la clave predeterminada o una clave AWS Key Management Service (AWS KMS) de su elección para cifrar el gráfico. También puede cifrar las instantáneas y los datos que se exportan a Amazon S3 durante la importación masiva. Puede eliminar el cifrado cuando se complete la importación.
- Cuando utilice el lenguaje OpenCypher, practique las técnicas adecuadas de validación y parametrización de las entradas para evitar la inyección de código SQL y otras formas de ataques. Evite crear consultas que utilicen la concatenación de cadenas con entradas proporcionadas por el usuario. Utilice consultas parametrizadas o sentencias preparadas para pasar de forma segura los parámetros de entrada a la base de datos de gráficos. Para obtener más información, consulte Ejemplos de consultas parametrizadas de OpenCypher en la documentación de Neptune.

Proteja sus redes

Puede habilitar un gráfico de Neptune Analytics para la conectividad pública, de modo que se pueda acceder a él desde fuera de una nube privada virtual (VPC). Esta conectividad está deshabilitada de forma predeterminada. El gráfico requiere la autenticación de IAM. La persona que llama debe obtener una identidad y tener permisos para usar el gráfico. Por ejemplo, para ejecutar una consulta de OpenCypher, la persona que llama debe tener permisos de lectura, escritura o eliminación en el gráfico específico.

También puede crear puntos finales privados para que el gráfico acceda al gráfico desde una VPC. Al crear el punto final, se especifican la VPC, las subredes y los grupos de seguridad para restringir el acceso para llamar al gráfico.

Para proteger los datos en tránsito, Neptune Analytics aplica conexiones SSL a través de HTTPS al gráfico. Para obtener más información, consulte Protección de datos en Neptune Analytics en la documentación de Neptune Analytics.

Implemente la autenticación y la autorización

Las llamadas a un gráfico de Neptune Analytics requieren autenticación de IAM. La persona que llama debe obtener una identidad y poseer los permisos suficientes para realizar la acción en el gráfico. Para obtener descripciones de las acciones de la API y sus permisos necesarios, consulte la documentación de la API de Neptune Analytics. Puede aplicar controles de estado para restringir el acceso por etiqueta.

La autenticación de IAM utiliza el protocolo AWS Signature versión 4 (SiGv4). Para simplificar el uso desde su aplicación, le recomendamos que utilice un AWS SDK. Por ejemplo, en Python, utilice el cliente Boto3 para Neptune Graph, que abstrae SigV4.

Al cargar datos en el gráfico, la carga por lotes utiliza las credenciales de IAM de la persona que llama. La persona que llama debe tener permisos para descargar datos de Amazon S3 con la relación de confianza configurada para que Neptune Analytics pueda asumir la función de cargar los datos en el gráfico desde los archivos de Amazon S3.

La importación masiva puede realizarse durante la creación del gráfico (por el equipo de infraestructura) o en un gráfico vacío existente (por el equipo de ingeniería de datos que tiene permisos para iniciar las tareas de importación). En ambos casos, Neptune Analytics asume la función de IAM que la persona que llama proporciona como entrada. Esta función le da permiso para leer y enumerar el contenido de la carpeta Amazon S3 en la que se almacenan los datos de entrada.

Pilar de fiabilidad

El AWS pilar de confiabilidad de Well-Architected Framework abarca la capacidad de una carga de trabajo para realizar la función prevista de manera correcta y coherente cuando se espera que lo haga. Esto incluye la capacidad de operar y probar la carga de trabajo durante todo su ciclo de vida.

Una carga de trabajo fiable comienza por tomar decisiones de diseño anticipadas tanto para el software como para la infraestructura. Sus elecciones de arquitectura afectan al comportamiento de la carga de trabajo en todos los pilares de Well-Architected. Para garantizar la confiabilidad, hay patrones específicos que debe seguir, como se explica en esta sección.

El pilar de confiabilidad se centra en las siguientes áreas clave:

- Arquitectura de carga de trabajo, incluidas las cuotas de servicio y los patrones de implementación
- Administración de cambios
- Administración de errores

Comprenda las cuotas de servicio de Neptune

Su AWS cuenta tiene cuotas predeterminadas (antes denominadas límites) para cada una Servicio de AWS de ellas. A menos que se indique lo contrario, cada cuota es específica de la región. Puedes solicitar aumentos para algunas cuotas, pero no para todas.

Para buscar las cuotas de Neptune Analytics, abra la consola Service Quotas. En el panel de navegación, elija y Servicios de AWS, a continuación, seleccione Amazon Neptune Analytics. Preste atención a las cuotas en cuanto al número de gráficos e instantáneas, a la memoria máxima aprovisionada para un gráfico y a las tasas de solicitudes de API.

Si la memoria máxima aprovisionada no es suficiente para su conjunto de datos, evalúe qué tipos de nodos y bordes son esenciales para el uso analítico previsto. Cargue un subconjunto de los datos para que los análisis sean posibles dentro de la capacidad aprovisionada permitida. Muchas cargas de trabajo de análisis, especialmente las que ejecutan algoritmos de gráficos, solo necesitan la topología con un conjunto limitado de propiedades en lugar del gráfico transaccional completo. (Para ver un análisis de las diferencias entre las cargas de trabajo transaccionales y analíticas, consulte la sección sobre el pilar de la eficiencia del rendimiento).

Si el número máximo de gráficos no es suficiente para el uso previsto:

- Considere la posibilidad de combinar gráficos que tengan usos similares.
- Evalúe cuántos gráficos deben ejecutarse en un momento dado. Si tiene un caso práctico de análisis efímero, capture un gráfico y elimínelo cuando ya no lo necesite. Esto reduce el número de gráficos en comparación con la cuota.
- Considere la posibilidad de aprovisionar los gráficos de forma diferente. Cuentas de AWS

Comprenda los patrones de despliegue de Neptune

Comprenda los siguientes puntos de decisión cuando planee implementar un gráfico de Neptune Analytics:

- Siembra: decida si desea crear un gráfico vacío o cargar datos en él en el momento de la creación con datos de Amazon S3, un clúster de base de datos de Neptune existente o una instantánea de base de datos de Neptune existente.

  Recomendación: Si la fuente es un cúmulo o una instantánea de Neptuno, debe cargar sus datos en el momento de la creación del gráfico. Si la fuente es Amazon S3, cargue los datos en el momento de la creación si el esfuerzo de carga es significativo y se realiza mejor como actividad de aprovisionamiento de infraestructura. Si prefiere cargar datos como una actividad de ingeniería de datos o de aplicación, cree un gráfico vacío y cargue los datos desde Amazon S3 más adelante.
- Capacidad: calcule la capacidad aprovisionada necesaria para un gráfico, teniendo en cuenta el tamaño de los datos y el uso esperado de la aplicación.

  Recomendación: En el momento de la creación, especifique la cantidad máxima de memoria aprovisionada para limitar el tamaño del gráfico. Esta configuración es obligatoria. Si es necesario, puede cambiar la capacidad más adelante.
- Disponibilidad y tolerancia a errores: decida si se requieren réplicas para garantizar la disponibilidad. Una réplica actúa como un dispositivo de reserva para la recuperación en caso de que se produzca un error en el gráfico. Un gráfico con réplicas se recupera más rápido que un gráfico sin réplicas. También considere cuánto tiempo se necesita el gráfico, si es solo para análisis efímeros y, de ser así, cuándo se eliminará.

  Recomendación: Determine los requisitos de disponibilidad (por ejemplo, cuánto tiempo puede no estar disponible el gráfico y cuándo puede eliminarse) antes de crearlo.
- Redes y seguridad: determine si necesita conectividad pública, privada o ambas, y si desea cifrar sus datos.

Recomendación: Antes de crear un gráfico, comprenda los requisitos organizativos (por ejemplo, si se permite la conectividad pública y dónde se implementarán las aplicaciones cliente de gráficos).

- Copias de seguridad y recuperación: determine si se deben crear instantáneas y, de ser así, cuándo y en qué condiciones. Considere si su organización tiene requisitos de recuperación ante desastres (DR).

  Recomendación: La creación de instantáneas es una actividad manual. Decida cuándo crear las instantáneas y tenga en cuenta sus requisitos de recuperación ante desastres antes de crear un gráfico.

Gestione y escale los clústeres de Neptune

Un gráfico de Neptune Analytics consta de una única instancia optimizada para la memoria. La capacidad (m-NCU) de la instancia se establece en el momento de la creación. La instancia se puede escalar verticalmente aumentando la capacidad aprovisionada mediante una acción administrativa; también se puede reducir la capacidad aprovisionada. Las réplicas son objetivos de conmutación por error pasiva, por lo que no aumentan la escala de un gráfico. En este sentido, una réplica de un gráfico difiere de una réplica de lectura de una base de datos de Neptuno, que es una instancia activa en un clúster de Neptuno que puede procesar las operaciones de lectura de las aplicaciones.

Las réplicas conllevan costes. El precio de la réplica se basa en la tasa m-NCU del gráfico. Por ejemplo, si un gráfico está aprovisionado para 128 millones de NCU y tiene una sola réplica, el costo es el doble que el de un gráfico equivalente sin réplicas.

En el ámbito del análisis, hay dos motivos principales para ampliar la escala:

- Para proporcionar más memoria y CPU para las consultas y los algoritmos analíticos, dado que la consulta individual es cara, el algoritmo gráfico que se ejecuta es intrínsecamente complejo y requiere más recursos debido a su entrada, o bien la tasa de solicitudes simultáneas es alta. Si estas consultas presentan out-of-memory errores, la ampliación es una solución razonable.
- Para admitir un tamaño de gráfico mayor del previsto. Por ejemplo, si la capacidad aprovisionada actualmente es de 128 M-NCU para admitir 60 GB de datos de origen y necesita 40 GB de datos de origen adicionales, está justificado aumentarla a 256 M-NCU.

Supervise CloudWatch las métricas de Neptune Analytics, como`NumQueuedRequestsPerSec`,`NumOpenCypherRequestsPerSec`,`GraphStorageUsagePercent`, y `GraphSizeBytesCPUUtilization`, para determinar si es necesario escalar. Puede actualizar la configuración de un gráfico a través de la consola AWS CLI, o SDKs. (Para ver ejemplos y prácticas recomendadas, consulte la sección sobre el pilar de la excelencia operativa).

Gestione las copias de seguridad y los eventos de conmutación por error

Utilice réplicas para garantizar que haya un gráfico disponible en caso de fallo. Un gráfico utiliza la persistencia basada en registros para confirmar los cambios en todas las zonas de disponibilidad de un. Región de AWS La réplica actúa como un dispositivo de reserva y tiene acceso a estos datos. Si se produce un error, el gráfico reanuda las operaciones en la réplica. La aplicación sigue utilizando el mismo punto final para conectarse al gráfico. Las solicitudes en curso durante el fallo generan errores, con la excepción de que el servicio no esté disponible. Considere la posibilidad de utilizar un patrón de reintento con retroceso en el código de la aplicación para detectar el error e inténtelo de nuevo tras un breve intervalo. Las nuevas solicitudes realizadas durante la conmutación por error se ponen en cola y es posible que experimenten una latencia más prolongada.

Si no se configura ninguna réplica y el gráfico falla, Neptune Analytics se recupera del almacenamiento duradero, pero la recuperación tarda más porque Neptune tiene que reinicializar los recursos.

Cree instantáneas del gráfico. (Neptune Analytics no toma instantáneas automáticas). Si el gráfico se modifica periódicamente después de crearlo, tome instantáneas frecuentes para capturar su estado actual. Elimine las instantáneas antiguas si no es necesario restaurarlas a un momento anterior.

Puedes compartir las instantáneas con otras cuentas y entre ellas. Regiones de AWS Si tiene requisitos de recuperación ante desastres, considere si restaurar el gráfico en una región diferente a partir de una instantánea cumple con sus requisitos de objetivo de tiempo de recuperación (RTO) y objetivo de punto de recuperación (RPO).

Pilar de eficiencia de rendimiento

El pilar de eficiencia del rendimiento del AWS Well-Architected Framework se centra en cómo optimizar el rendimiento al ingerir o consultar datos. La optimización del rendimiento es un proceso gradual y continuo que consiste en lo siguiente:

- Confirmación de los requisitos empresariales
- Medición del rendimiento de la carga
- Identificar los componentes de bajo rendimiento
- Ajustando los componentes para que se adapten a las necesidades de su empresa

El pilar de la eficiencia del rendimiento proporciona pautas específicas para cada caso de uso que pueden ayudarlo a identificar el modelo de datos gráfico y los lenguajes de consulta correctos que debe utilizar. También incluye las mejores prácticas que se deben seguir a la hora de ingerir y consumir datos de Neptune Analytics.

El pilar de la eficiencia del rendimiento se centra en las siguientes áreas clave:

- Modelado gráfico
- Optimización de las consultas
- Dimensionamiento correcto de gráficos
- Optimización de escritura

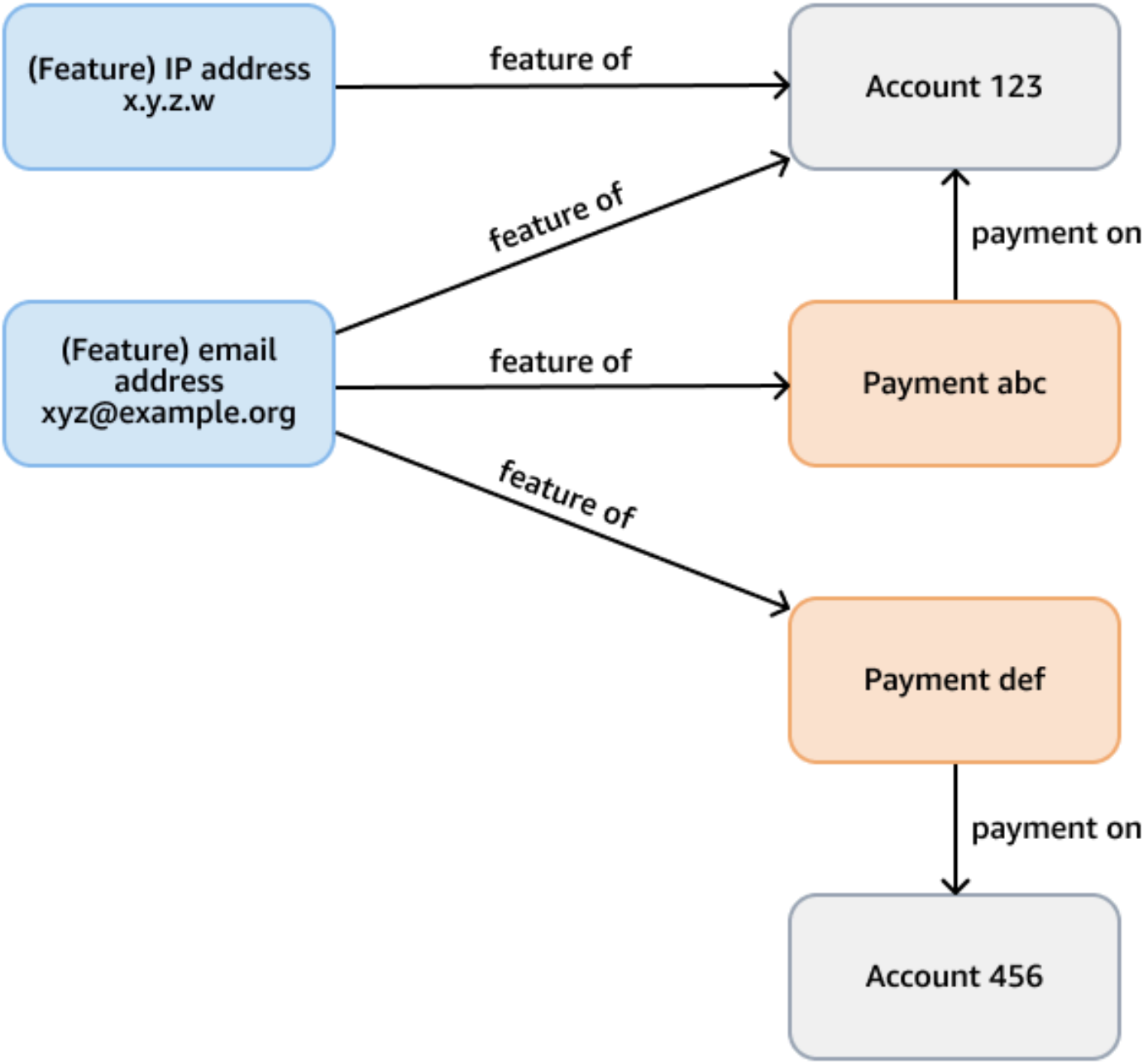
Comprenda el modelado de gráficos para el análisis

En la guía Applying the AWS Well-Architected Framework for Amazon Neptune se analiza el modelado de gráficos para mejorar la eficiencia del rendimiento. Las decisiones de modelado que afectan al rendimiento incluyen la elección de los nodos y bordes necesarios, sus etiquetas y propiedades IDs, la dirección de los bordes, si las etiquetas deben ser genéricas o específicas y, en general, la eficacia con la que el motor de consultas puede navegar por el gráfico para procesar las consultas comunes.

Estas consideraciones también se aplican a Neptune Analytics; sin embargo, es importante distinguir entre los patrones de uso transaccionales y analíticos. Es posible que sea necesario cambiar la

forma de un modelo gráfico que sea eficaz para las consultas en una base de datos transaccional, como una base de datos de Neptune, para fines de análisis.

Por ejemplo, consideremos un gráfico de fraude en una base de datos de Neptune cuyo propósito es comprobar si hay patrones fraudulentos en los pagos con tarjeta de crédito. Este gráfico puede tener nodos que representen las cuentas, los pagos y las características (como la dirección de correo electrónico, la dirección IP o el número de teléfono) tanto de la cuenta como del pago. Este gráfico conectado admite consultas como recorrer una ruta de longitud variable que comienza con un pago determinado y realiza varios saltos para encontrar funciones y cuentas relacionadas. En la siguiente figura se muestra un gráfico de este tipo.



El requisito analítico puede ser más específico, como encontrar comunidades de cuentas que estén vinculadas mediante una función. Para ello, puede utilizar el algoritmo de componentes débilmente conectados (WCC). Ejecutarlo con el modelo del ejemplo anterior es ineficiente, ya que necesita atravesar varios tipos diferentes de nodos y bordes. El modelo del siguiente diagrama es más eficiente. Vincula `account` los nodos con una `shares feature` ventaja si las propias cuentas (o los pagos de las cuentas) comparten una función. Por ejemplo, `Account 123` tiene la función `xyz@example.org` de correo electrónico y `Account 456` usa ese mismo correo electrónico para un pago (). `Payment def`

La complejidad computacional del WCC es `O(|E|logD)` dónde `|E|` está el número de aristas del gráfico y `D` es el diámetro (la longitud de la ruta más larga) que conecta los nodos. Como el modelo transaccional omite los nodos y aristas no esenciales, optimiza tanto el número de aristas como el diámetro y reduce la complejidad del algoritmo del WCC.

Cuando utilice Neptune Analytics, trabaje a partir de los algoritmos y consultas analíticas necesarios. Si es necesario, modifique la forma del modelo para optimizar estas consultas. Puede cambiar la forma del modelo antes de cargar datos en el gráfico o escribir consultas que modifiquen los datos existentes en el gráfico.

Optimización de consultas

Siga estas recomendaciones para optimizar las consultas de Neptune Analytics:

- Utilice consultas parametrizadas y la caché del plan de consultas, que está habilitada de forma predeterminada. Al utilizar la memoria caché del plan, el motor prepara la consulta para su uso posterior, siempre que la consulta se complete en 100 milisegundos o menos, lo que ahorra tiempo en las siguientes invocaciones.
- En el caso de consultas lentas, ejecuta un plan explicativo para detectar los cuellos de botella y realizar las mejoras correspondientes.
- Si utilizas la búsqueda de similitudes vectoriales, decide si las incrustaciones más pequeñas producen resultados de similitud precisos. Puede crear, almacenar y buscar incrustaciones más pequeñas de forma más eficiente.
- Siga las prácticas recomendadas documentadas para usar OpenCypher en Neptune Analytics. Por ejemplo, utilice mapas aplanados en una cláusula UNWIND y especifique las etiquetas de los bordes siempre que sea posible.
- Cuando utilice un algoritmo gráfico, comprenda las entradas y salidas del algoritmo, su complejidad computacional y, en términos generales, cómo funciona.

- Antes de utilizar un algoritmo gráfico, utilice una `MATCH` cláusula para minimizar el conjunto de nodos de entrada. Por ejemplo, para limitar los nodos desde los que realizar búsquedas prioritarias por amplitud (BFS), siga los ejemplos que se proporcionan en la documentación de Neptune Analytics.
- Si es posible, filtre las etiquetas de los nodos y los bordes. Por ejemplo, BFS tiene parámetros de entrada para filtrar el recorrido hasta una etiqueta de nodo específica (`vertexLabel`) o etiquetas de borde específicas (`edgeLabels`).
- Utilice parámetros delimitadores, por ejemplo, `maxDepth` para limitar los resultados.
- Experimente con el `concurrency` parámetro. Pruébelo con un valor de 0, que utiliza todos los subprocesos del algoritmo disponibles para paralelizar el procesamiento. Compárelo con la ejecución de un solo subproceso configurando el parámetro en 1. Un algoritmo puede completarse más rápido en un solo subproceso, especialmente en entradas más pequeñas, como las búsquedas con poco espacio, en las que el paralelismo no ofrece una reducción apreciable del tiempo de ejecución y puede generar una sobrecarga.
- Elige entre tipos de algoritmos similares. Por ejemplo, Bellman-Ford y delta-Stepping son algoritmos de ruta más corta de una sola fuente. Cuando realices pruebas con tu propio conjunto de datos, prueba ambos algoritmos y compara los resultados. El delta suele ser más rápido que el de Bellman-Ford debido a su menor complejidad computacional. Sin embargo, el rendimiento depende del conjunto de datos y de los parámetros de entrada, en particular del parámetro. `delta`

Optimice las escrituras

Siga estas prácticas para optimizar las operaciones de escritura en Neptune Analytics:

- Busque la forma más eficiente de cargar datos en un gráfico. Cuando cargue datos en Amazon S3, utilice la importación masiva si los datos tienen un tamaño superior a 50 GB. Para datos más pequeños, utilice la carga por lotes. Si se out-of-memory producen errores al ejecutar la carga por lotes, considere la posibilidad de aumentar el valor de m-NCU o dividir la carga en varias solicitudes. Una forma de lograrlo consiste en dividir los archivos en varios prefijos en el bucket de S3. En ese caso, llame a batch load por separado para cada prefijo.
- Utilice la importación masiva o el cargador por lotes para rellenar el conjunto inicial de datos del gráfico. Utilice las operaciones transaccionales de creación, actualización y eliminación de OpenCypher únicamente para cambios pequeños.

- Utilice la importación masiva o el cargador por lotes con una simultaneidad de 1 (subproceso único) para incorporar las incrustaciones en el gráfico. Intente cargar las incrustaciones por adelantado mediante uno de estos métodos.
- Evalúe la dimensión de las incrustaciones vectoriales necesarias para una búsqueda de similitud precisa en los algoritmos de búsqueda de similitudes vectoriales. Utilice una dimensión más pequeña si es posible. Esto se traduce en una velocidad de carga más rápida para las incrustaciones.
- Utilice algoritmos de mutación para recordar los resultados algorítmicos si es necesario. Por ejemplo, el algoritmo de centralidad con grado de mutación busca el grado de cada nodo de entrada y escribe ese valor como una propiedad del nodo. Si las conexiones que rodean esos nodos no cambian posteriormente, la propiedad contiene el resultado correcto. No es necesario volver a ejecutar el algoritmo.
- Utilice la acción administrativa de restablecimiento gráfico para borrar todos los nodos, bordes e incrustaciones si necesita empezar de nuevo. Si el gráfico es grande, no es posible eliminar todos los nodos, bordes e incrustaciones mediante una consulta de OpenCypher. Se puede agotar el tiempo de espera de una sola consulta en un conjunto de datos grande. A medida que aumenta el tamaño, el conjunto de datos tarda más en eliminarse y el tamaño de la transacción aumenta. Por el contrario, el tiempo necesario para completar el restablecimiento de un gráfico es prácticamente constante y la acción ofrece la opción de crear una instantánea antes de ejecutarla.

Gráficos del tamaño correcto

El rendimiento general depende de la capacidad aprovisionada de un gráfico de Neptune Analytics. La capacidad se mide en unidades denominadas unidades de capacidad de Neptune optimizadas para la memoria (m -). NCUs Asegúrese de que el gráfico tenga el tamaño suficiente para admitir el tamaño del gráfico y las consultas. Ten en cuenta que el aumento de la capacidad no mejora necesariamente el rendimiento de una consulta individual.

Si es posible, cree el gráfico importando datos de una fuente existente, como Amazon S3 o un clúster o una instantánea de Neptune existente. Puede establecer límites en la capacidad mínima y máxima. También puede cambiar la capacidad aprovisionada en un gráfico existente.

Supervise CloudWatch métricas como`NumQueuedRequestsPerSec`, `NumOpenCypherRequestsPerSec` `GraphStorageUsagePercentGraphSizeBytes`, y `CPUUtilization` evalúe si el gráfico tiene el tamaño correcto. Determine si se necesita más

capacidad para soportar el tamaño y la carga del gráfico. Para obtener más información sobre cómo interpretar algunas de estas métricas, consulte la sección sobre el pilar de la excelencia operativa.

Pilar de optimización de costos

El pilar de optimización de costes del AWS Well-Architected Framework se centra en evitar costes innecesarios. Las siguientes recomendaciones pueden ayudarle a cumplir los principios de diseño de optimización de costes y las mejores prácticas arquitectónicas de Neptune Analytics.

El pilar de optimización de costes se centra en las siguientes áreas clave:

- Comprender el gasto a lo largo del tiempo y controlar la asignación de fondos
- Seleccionar recursos del tipo y la cantidad correctos
- Escalar para satisfacer las necesidades empresariales sin gastar de más

Comprenda los patrones de uso y los servicios necesarios

Antes de adoptar Neptune Analytics, evalúe si su caso de uso es adecuado para el análisis de gráficos.

- Bases de datos de gráficos: una base de datos de gráficos como Neptune es adecuada para su carga de trabajo si su modelo de datos tiene una estructura de gráficos discernible y sus consultas necesitan explorar relaciones y recorrer varios saltos. Una base de datos de gráficos no es adecuada para los siguientes patrones:
  - Principalmente consultas de salto único. En este caso de uso, considere si sus datos podrían representarse mejor como atributos de un objeto.
  - Datos JSON o binarios de objetos grandes (blob) almacenados como propiedades.

- Análisis de gráficos: Neptune Analytics es un motor de base de datos de análisis de gráficos que puede analizar rápidamente grandes cantidades de datos de gráficos en la memoria para obtener información y encontrar tendencias. Puede almacenar y consultar datos de gráficos tanto en una base de datos de Neptune como en un gráfico de Neptune Analytics. Una base de datos Neptune es la más adecuada para las necesidades de procesamiento transaccional en línea (OLTP) escalable. Neptune Analytics es ideal para cargas de trabajo de análisis efímeras. Puede usar ambos en combinación cargando datos de su base de datos de Neptuno orientada a transacciones en un gráfico de Neptune Analytics para ejecutar el análisis de esos datos. Cuando se complete el análisis, puede eliminar el gráfico de Neptune Analytics. Para obtener una comparación más

detallada, consulte Cuándo usar Neptune Analytics y Cuándo usar Neptune Database en la documentación de Neptune Analytics.

Determine, prestando atención al costo, la mejor manera de completar su gráfico de Neptune Analytics.

- Importe datos gráficos de forma masiva agrupados en un bucket de S3. Recomendamos esta opción si sus datos se prepararon previamente para su carga masiva en una base de datos de Neptune, o si ya tiene, o puede producir fácilmente, los datos para analizarlos en CSV u otros formatos compatibles que requiera la importación masiva. Puede ejecutar la importación masiva como parte del procedimiento de creación de gráficos. Puede establecer límites en la capacidad mínima y máxima. También puede ejecutar la importación en un gráfico vacío creado anteriormente y supervisar la tarea de importación mientras se ejecuta.
- Puede crear un gráfico vacío y, a continuación, rellenarlo mediante una consulta de OpenCypher mediante la carga por lotes. Esta opción es ideal si los datos que se van a cargar se almacenan en Amazon S3 y tienen un tamaño inferior a 50 GB.
- Puede rellenar el gráfico a partir de los datos del clúster de base de datos de Neptuno (compatible con la versión 1.3.0 o posterior de Neptune Database). La intención de este patrón es realizar análisis de los datos que se encuentran actualmente en su base de datos de gráficos. Incluso si la base de datos se rellenó inicialmente mediante carga masiva, es posible que haya cambiado significativamente desde entonces. Para importar desde la base de datos, Neptune Analytics clona la base de datos y exporta los datos del clon a un bucket de S3. Este procedimiento conlleva costes: en particular, los costes de la base de datos de Neptune para ejecutar el clon y los costes de Amazon S3 para almacenar y consumir los datos exportados. El clon se elimina cuando se completa la exportación. Puede eliminar los datos exportados en Amazon S3.
- Puede rellenar el gráfico a partir de la instantánea de un clúster de base de datos de Neptune. Es similar a la opción anterior, excepto que la fuente es una instantánea de la base de datos. Para importar desde una instantánea, Neptune Analytics primero restaura la instantánea en un nuevo clúster de base de datos y, a continuación, exporta los datos a un bucket de S3. Este procedimiento conlleva costes: en particular, los costes de la base de datos de Neptune para ejecutar el clúster restaurado y los costes de Amazon S3 para almacenar y consumir los datos exportados.
- También puede realizar consultas de OpenCypher para crear, actualizar o eliminar datos utilizando transacciones en el gráfico que cumplan con la atomicidad, la coherencia, el aislamiento y

la durabilidad (ACID). Recomendamos este enfoque como una forma de realizar pequeñas actualizaciones, pero no como una forma de sembrar el gráfico.

Si los datos necesarios para el análisis ya están almacenados en Amazon S3, recomendamos importarlos de forma masiva o cargarlos por lotes. Son más rentables que rellenar el gráfico desde un clúster o una instantánea de la base de datos de Neptune.

Seleccione los recursos prestando atención al costo

Los precios de Neptune Analytics utilizan una unidad conocida como Unidad de capacidad de Neptune optimizada para memoria (m-NCU). La ejecución de un gráfico con una m-NCU determinada conlleva un coste fijo por hora. Un gráfico puede tener réplicas para la conmutación por error, y estas réplicas también conllevan un coste de m-NCU por hora.

Recomendamos las siguientes prácticas recomendadas para estimar la capacidad, limitar los costes y supervisar los costes en relación con el rendimiento:

- Si es posible, cree el gráfico importando datos de una fuente existente: datos almacenados en Amazon S3 o un clúster o una instantánea de Neptune existente. Esto le ahorra esfuerzo, ya que Neptune Analytics realiza la ardua tarea de sembrar el gráfico y puede especificar una capacidad máxima limitada.
- Puede cambiar la capacidad aprovisionada en un gráfico existente.
- Cuando el gráfico ya no sea necesario, puede crear una instantánea y eliminarlo. Si necesita volver a utilizarla, puede restaurar la gráfica a partir de la instantánea.
- Puede elegir el número de réplicas al crear el gráfico. Establezca el valor de acuerdo con sus requisitos de disponibilidad de análisis. Ahorre costes minimizando esta configuración. El valor máximo de 2 permite dos instancias de réplica en zonas de disponibilidad independientes. El valor mínimo de 0 significa que Neptune Analytics no ejecutará una réplica. Sin embargo, la recuperación es más rápida cuando hay una réplica disponible. Para obtener una explicación de las fallas y la recuperación de los gráficos, consulte la sección sobre el pilar de confiabilidad.
- Supervise los gastos de Neptune Analytics para los períodos de facturación actuales y pasados mediante. Administración de facturación y costos de AWS
- Supervise las métricas de Neptune Analytics CloudWatch `NumQueuedRequestsPerSec` `NumOpenCypherRequestsPerSec` `GraphStorageUsagePercentGraphSizeBytes`, especialmente para evaluar si la capacidad aprovisionada tiene el tamaño adecuado para el

gráfico. `CPUUtlization` Determine si una capacidad más pequeña puede adaptarse a la tasa de solicitudes observada, al uso de la CPU y al tamaño del gráfico.

- Si necesita un punto final privado para su gráfico, preste atención a los costos de los puntos finales elásticos de nube privada virtual (VPC) IPs, las pasarelas de NAT u otros costos relacionados con la VPC. Para obtener más información, consulte los precios de Amazon VPC y los precios de Amazon EC2.
- Es posible que desee ejecutar una o más instancias de Neptune notebook para proporcionar una interfaz de cliente que ayude a los desarrolladores y analistas a consultar y visualizar el gráfico (consulte los precios de Neptune Workbench). Para minimizar los costes, comparta la instancia entre los usuarios y cree carpetas de bloc de notas independientes para cada usuario. Cierre la instancia cuando no esté en uso. Para obtener información sobre cómo automatizar el cierre, consulte la entrada del AWS blog Automatice la detención y el inicio de los recursos del entorno Amazon Neptune mediante etiquetas de recursos.

Pilar de sostenibilidad

El pilar de sostenibilidad del AWS Well-Architected Framework se centra en minimizar los impactos ambientales de la ejecución de cargas de trabajo en la nube. Los temas clave incluyen un modelo de responsabilidad compartida para la sostenibilidad, comprender el impacto y maximizar el uso para minimizar los recursos necesarios y reducir los impactos posteriores.

El pilar de la sostenibilidad contiene las siguientes áreas de enfoque clave:

- ¿Su impacto
- Objetivos de sostenibilidad
- Uso maximizado
- Anticipar y adoptar ofertas de software nuevas y más eficientes
- Uso de servicios gestionados
- Reducción del impacto descendente

Esta guía se centra en comprender su impacto. Para obtener más información sobre los demás principios de diseño de sostenibilidad, consulte AWS Well-Architected Framework.

Sus elecciones y requisitos tienen un impacto en el medio ambiente. Si puede elegir Regiones de AWS que tengan una menor intensidad de carbono y si sus requisitos reflejan las necesidades reales de carga de trabajo en lugar de maximizar únicamente el tiempo de actividad y la durabilidad, la sostenibilidad de la carga de trabajo aumenta. En las siguientes secciones se analizan las mejores prácticas y las consideraciones que, si se adoptan en el diseño de la carga de trabajo y en las operaciones continuas, tendrán un impacto ambiental positivo

Tenga en cuenta su Región de AWS selección

Algunos Regiones de AWS están cerca de los proyectos de energía renovable de Amazon o ubicados donde la red tiene una intensidad de carbono publicada inferior a la de otros. Tenga en cuenta el impacto en la sostenibilidad de las regiones que podrían ser viables para su carga de trabajo y compare su lista con las regiones en las que Neptune Analytics está disponible.

Optimice el consumo

Minimice el consumo de Neptune Analytics practicando lo siguiente:

- La analítica suele ser efímera. El gráfico solo es necesario durante el tiempo necesario para ejecutar los algoritmos y registrar los resultados. Si este es el caso, tome una instantánea del gráfico y elimínelo cuando ya no lo necesite. Si es necesario, puede restaurarlo a partir de una instantánea más adelante.
- Si la carga de trabajo es efímera y tiene la flexibilidad de decidir cuándo ejecutar los análisis, tenga en cuenta day-to-day las tendencias del consumo de energía. La demanda de electricidad es mayor en determinados momentos. Si se encuentra en los Estados Unidos, consulte las métricas del consumo diario de electricidad en el sitio web de la Administración de Información Energética (EIA) de los Estados Unidos. Si es posible, ejecute las cargas de trabajo durante los períodos de menor actividad en su región.
- Si la carga de trabajo no es efímera, sino que solo debe estar disponible durante períodos limitados, elimine el gráfico y restáurelo a partir de una instantánea cuando sea necesario. Si su disponibilidad sigue un cronograma, automatice el proceso de restauración mediante scripts para que el gráfico esté listo a la hora programada.
- Si los datos son de solo lectura o no han cambiado desde la última instantánea, no vuelva a capturarlos antes de eliminarlos.
- Detenga los cuadernos Neptune cuando no estén en uso.
- Supervise CloudWatch métricas como`NumQueuedRequestsPerSec`,`NumOpenCypherRequestsPerSec`, `GraphStorageUsagePercentGraphSizeBytes`, y `CPUUtilization` evalúe si el gráfico está sobredimensionado. Determine si una capacidad de instancia más pequeña puede adaptarse a la tasa de solicitudes observada, el uso de la CPU y el tamaño del gráfico.

Optimice los patrones de desarrollo y arquitectura del software

Para evitar el desperdicio, optimice sus modelos y consultas, y comparta los recursos de cómputo para utilizar todos los recursos disponibles en las instancias y los clústeres de Neptune. Entre las prácticas recomendadas específicas se incluyen las siguientes:

- Optimice las consultas y las invocaciones de algoritmos gráficos. Utilice consultas parametrizadas y utilice la caché del plan de consultas, que está habilitada de forma predeterminada. En el caso de consultas lentas, ejecuta un plan explicativo para realizar mejoras. Si utilizas la búsqueda por similitud vectorial, decide si las incrustaciones más pequeñas producen resultados de similitud precisos, ya que las incrustaciones más pequeñas se pueden crear, almacenar y buscar de forma

más eficiente. Antes de utilizar un algoritmo gráfico, utilice una `MATCH` cláusula para minimizar el conjunto de nodos de entrada. Si es posible, filtra las etiquetas de los nodos y los bordes.

- Busque la forma más eficiente de cargar datos en el gráfico. Si carga desde datos en Amazon S3, utilice la importación masiva si los datos tienen un tamaño superior a 50 GB. Utilice la carga por lotes para datos más pequeños.
- Pide a los desarrolladores que compartan las instancias del bloc de notas de Neptune en lugar de que cada uno cree su propia instancia. Crea carpetas de bloc de notas independientes para cada desarrollador en una sola instancia de Jupyter. Cierre la instancia cuando no esté en uso.

Recursos

Referencias

- AWS Well-Architected
- AWS Documentación de Well-Architected Framework
- Aplicación del marco AWS Well-Architected para Amazon Neptune
- Mejores prácticas de Neptune Analytics

Publicaciones de blog y vídeos

- Publicaciones del blog de Neptune (blog de AWS base de datos)
- Automatice la detención y el inicio de los recursos del entorno Amazon Neptune mediante etiquetas de recursos (blog de AWS bases de datos)
- Aperitivos de Amazon Neptune (vídeos cortos) YouTube

Formación

- Introducción a Amazon Neptune
- Cree con Amazon Neptune
- Modelado de datos para Amazon Neptune
- Taller de análisis de Amazon Neptune

Historial de documentos

En la siguiente tabla, se describen cambios significativos de esta guía. Si quiere recibir notificaciones de futuras actualizaciones, puede suscribirse a las notificaciones RSS.

Cambio	Descripción	Fecha
Publicación inicial	—	20 de diciembre de 2024

AWS Glosario de orientación prescriptiva

Los siguientes son términos de uso común en las estrategias, guías y patrones proporcionados por la Guía AWS prescriptiva. 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 EC2 instancia 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. Algunos ejemplos de funciones agregadas 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 AIOps se utiliza en la estrategia de AWS migración, 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.

branch

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 está seguro, despliega la nueva versión y reemplaza la versión actual en su totalidad.

CCoE

Consulte Cloud Center of Excellence.

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 CCo E 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: realizar inversiones fundamentales para escalar su adopción de la nube (p. ej., crear una landing zone, definir una CCo E, establecer un modelo de operaciones)

- Migración: migración de aplicaciones individuales
- Reinvención: 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 en el 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 oBitbucket Cloud. 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, Amazon SageMaker AI 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, puesta en escena y producción del proceso de publicación del software. CI/CD is commonly described as a pipeline. CI/CDpuede 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

Vea la visión artificial.

D

datos en reposo

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

clasificación de datos

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

del AWS Well-Architected Framework. Para obtener más información, consulte Clasificación de datos.

desviación de datos

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

datos en tránsito

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

malla de datos

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

minimización de datos

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

perímetro de datos

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

preprocesamiento de datos

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

procedencia de los datos

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

titular de los datos

Persona cuyos datos se recopilan y procesan.

almacenamiento de datos

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

lenguaje de definición de datos (DDL)

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

lenguaje de manipulación de datos (DML)

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

DDL

Consulte el lenguaje de definición de bases de datos.

conjunto profundo

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

aprendizaje profundo

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

defense-in-depth

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

administrador delegado

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

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

Implementación

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

entorno de desarrollo

Consulte entorno.

control de detección

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

asignación de flujos de valor para el desarrollo (DVSM)

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

gemelo digital

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

tabla de dimensiones

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

desastre

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

recuperación de desastres (DR)

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

DML

Consulte el lenguaje de manipulación de bases de datos.

diseño basado en el dominio

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

DR

Consulte recuperación ante desastres.

detección de deriva

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.

EDI

Véase intercambio electrónico 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.

intercambio electrónico de datos (EDI)

El intercambio automatizado de documentos comerciales entre organizaciones. Para obtener más información, consulte Qué es el intercambio electrónico de datos.

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).

entorno

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.

ERP

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.

indicaciones de unos pocos pasos

Proporcionar a un LLM un pequeño número de ejemplos que demuestren la tarea y el resultado deseado antes de pedirle que realice una tarea similar. Esta técnica es una aplicación del aprendizaje contextual, en el que los modelos aprenden a partir de ejemplos (planos) integrados en las instrucciones. Las indicaciones con pocas tomas pueden ser eficaces para tareas que requieren un formato, un razonamiento o un conocimiento del dominio específicos. Consulte también el apartado de mensajes sin intervención.

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.

FM

Consulte el modelo básico.

modelo de base (FM)

Una gran red neuronal de aprendizaje profundo que se ha estado entrenando con conjuntos de datos masivos de datos generalizados y sin etiquetar. FMs son capaces de realizar una amplia variedad de tareas generales, como comprender el lenguaje, generar texto e imágenes y conversar en lenguaje natural. Para obtener más información, consulte Qué son los modelos básicos.

G

IA generativa

Un subconjunto de modelos de IA que se han entrenado con grandes cantidades de datos y que pueden utilizar un simple mensaje de texto para crear contenido y artefactos nuevos, como imágenes, vídeos, texto y audio. Para obtener más información, consulte Qué es la IA generativa.

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.

imagen dorada

Instantánea de un sistema o software que se utiliza como plantilla para implementar nuevas instancias de ese sistema o software. Por ejemplo, en la fabricación, una imagen dorada se puede utilizar para aprovisionar software en varios dispositivos y ayuda a mejorar la velocidad, la escalabilidad y la productividad de las operaciones de fabricación de dispositivos.

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 el cumplimiento en todas las unidades organizativas (OUs). 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

HA

Consulte la 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.

datos retenidos

Parte de los datos históricos etiquetados que se ocultan de un conjunto de datos que se utiliza para entrenar un modelo de aprendizaje automático. Puede utilizar los datos de reserva para evaluar el rendimiento del modelo comparando las predicciones del modelo con los datos de reserva.

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 versiones. DevOps

periodo de hiperatención

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

I

IaC

Vea la infraestructura como código.

políticas basadas en identidad

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

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.

IIoT

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 AWS de referencia de seguridad recomienda configurar la cuenta de red con entradas, salidas e inspección VPCs para proteger la interfaz bidireccional entre la aplicación y el resto de Internet.

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 (T) IIo

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 de la Internet de las cosas (IIoT) industrial.

VPC de inspección

En una arquitectura de AWS cuentas múltiples, una VPC centralizada que gestiona las inspecciones del tráfico de red VPCs entre Internet y las redes locales (en una misma o Regiones de AWS diferente). La arquitectura AWS de referencia de seguridad recomienda configurar su cuenta de red con entrada, salida e inspección VPCs para proteger la interfaz bidireccional entre la 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 obtener más información, consulte Interpretabilidad del modelo de aprendizaje automático 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.

modelo de lenguaje grande (LLM)

Un modelo de IA de aprendizaje profundo que se entrena previamente con una gran cantidad de datos. Un LLM puede realizar múltiples tareas, como responder preguntas, resumir documentos, traducir textos a otros idiomas y completar oraciones. Para obtener más información, consulte Qué son. LLMs

migración grande

Migración de 300 servidores o más.

LBAC

Consulte 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.

LLM

Véase un modelo de lenguaje amplio.

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 una red bien definida APIs 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 AWS servicios sin servidor.

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 un uso ligero. APIs 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: realoje la migración a Amazon EC2 con AWS 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.

ORR

Consulte la revisión de la preparación operativa.

OT

Consulte la 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 AWS de referencia de seguridad recomienda configurar la cuenta de red con entradas, salidas e inspección VPCs para proteger la interfaz bidireccional entre la 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.

policy

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 desarrollo.

zonas alojadas privadas

Un contenedor que contiene información sobre cómo desea que Amazon Route 53 responda a las consultas de DNS de un dominio y sus subdominios dentro de uno o más VPCs. 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.

encadenamiento rápido

Utilizar la salida de una solicitud de LLM como entrada para la siguiente solicitud para generar mejores respuestas. Esta técnica se utiliza para dividir una tarea compleja en subtareas o para refinar o ampliar de forma iterativa una respuesta preliminar. Ayuda a mejorar la precisión y la relevancia de las respuestas de un modelo y permite obtener resultados más detallados y personalizados.

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.

publish/subscribe (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).

RAG

Consulte Retrieval Augmented Generation.

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.

réplica de lectura

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.

Generación aumentada de recuperación (RAG)

Tecnología de inteligencia artificial generativa en la que un máster hace referencia a una fuente de datos autorizada que se encuentra fuera de sus fuentes de datos de formación antes de generar una respuesta. Por ejemplo, un modelo RAG podría realizar una búsqueda semántica en la base de conocimientos o en los datos personalizados de una organización. Para obtener más información, consulte Qué es el RAG.

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.

seguridad desde el diseño

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

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. Existen cuatro tipos principales de controles de seguridad: preventivos, de detección, con capacidad 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 EC2 instancia de Amazon 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)

Política que proporciona un control centralizado de los permisos de todas las cuentas de una organización en AWS Organizations. SCPs defina barreras o establezca límites a las acciones que un administrador puede delegar en usuarios o roles. Puede utilizarlas SCPs como listas de permitidos o rechazados para especificar qué servicios o acciones están 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.

SLO

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.

SPOF

Consulte el punto único de falla.

esquema en forma de estrella

Estructura organizativa de una base de datos que utiliza una tabla de hechos grande para almacenar datos medidos o transaccionales 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.

indicador del sistema

Una técnica para proporcionar contexto, instrucciones o pautas a un LLM para dirigir su comportamiento. Las indicaciones del sistema ayudan a establecer el contexto y las reglas para las interacciones con los usuarios.

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

Un centro de tránsito de red que puede usar para interconectar sus VPCs redes con las locales. 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

Una conexión entre dos VPCs que le 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 AWS de calificación de la carga de trabajo.

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.

aviso de tiro cero

Proporcionar a un LLM instrucciones para realizar una tarea, pero sin ejemplos (imágenes) que puedan ayudar a guiarla. El LLM debe utilizar sus conocimientos previamente entrenados para realizar la tarea. La eficacia de las indicaciones cero depende de la complejidad de la tarea y de la calidad de las indicaciones. Consulte también las indicaciones de pocos pasos.

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 version original de inglés, prevalecerá la version en inglés.