Resilience analysis framework

AWS Guía prescriptiva

AWS Guía prescriptiva: Resilience analysis framework

Table of Contents

Marco de análisis de resiliencia

John Formento, Bruno Emer, Steven Hooper, Jason Barto y Michael Haken, de Amazon Web Services (AWS)

Septiembre de 2023(historial de documentos)

Los estándares y procesos consistentes y repetibles son una parte importante de la mejora continua. Esto también es válido para la resiliencia de los sistemas distribuidos. El propósito de esta guía es introducir un marco de análisis de la resiliencia que proporcione una forma coherente de analizar los modos de falla y cómo podrían afectar a sus cargas de trabajo. El uso de este marco durante todo el ciclo de vida de su carga de trabajo, desde el diseño hasta la operación, le ayuda a mejorar continuamente la resiliencia de sus cargas de trabajo ante una gama más amplia de posibles modos de fallo de forma coherente y repetible. Esto ayuda a garantizar que cumpla sus objetivos de resiliencia y mantenga las propiedades de resiliencia deseadas de sus cargas de trabajo.

Este marco se desarrolló a partir de la experiencia de los equipos de campo de arquitectura de soluciones de AWS en su trabajo con clientes de todos los sectores. Está dirigido a desarrolladores que pueden ocupar varios puestos de trabajo, como gerentes de producto, desarrolladores de software, ingenieros de sistemas, equipos de operaciones y arquitectos. Estas son las personas que más saben sobre el sistema, servicio o producto que se está analizando. El uso del marco en ejercicios continuos puede ayudarle a progresar gradualmente y a cumplir sus objetivos de resiliencia a largo plazo.

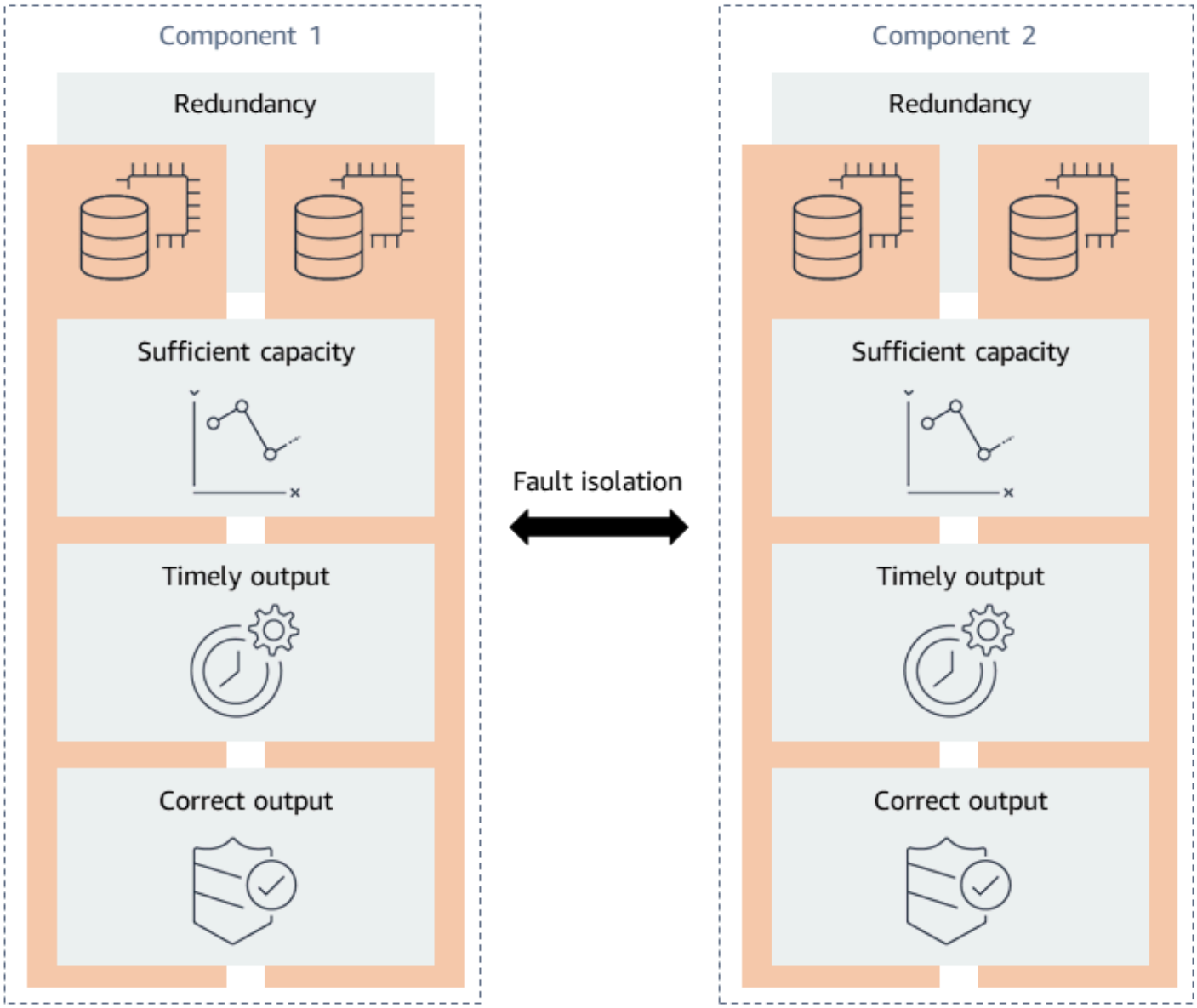
El objetivo del marco es identificar los posibles modos de fallo y los controles preventivos y correctivos que puede utilizar para mitigar su impacto. Incluso si las fallas se producen en componentes que no están directamente bajo su control, como el aumento de las tasas de error en una dependencia, debe tener en cuenta cómo esas fallas pueden afectar a su carga de trabajo y cómo diseñar esa carga de trabajo para responder a estas fallas. En última instancia, debe centrarse enfallas a las que puede respondermediante el uso de una mitigación que esté bajo su control.

Esta guía describe el marco y, a continuación, analiza cómo identificar y documentar una carga de trabajo, cómo aplicar el marco a esa carga de trabajo y cómo evaluar las estrategias de mitigación para detectar cualquier posible fallo que se detecte.

Contenido

Información general del marco de

El marco de análisis de la resiliencia se desarrolló identificando las propiedades de resiliencia deseadas de una carga de trabajo. Las propiedades deseadas son las cosas que usted quiere que sean ciertas acerca del sistema. Por lo general, la resiliencia se mide en función de la disponibilidad, por lo que cinco propiedades son las características de un sistema distribuido de alta disponibilidad: redundancia, capacidad suficiente, producción puntual, salida correcta y aislamiento de fallas. Estas propiedades se muestran en el siguiente diagrama.



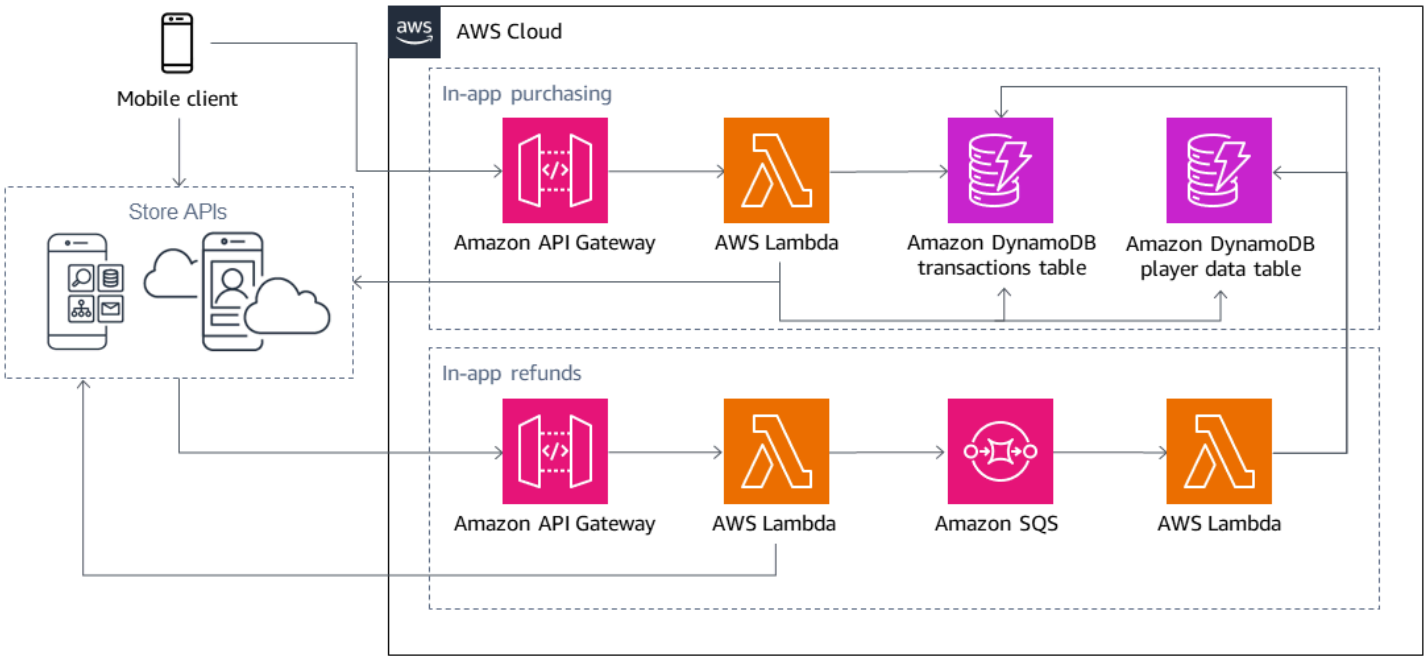
- Redundancia— La tolerancia a los fallos se logra mediante la redundancia, que elimina los puntos únicos de fallo (SPOF). La redundancia puede abarcar desde componentes de repuesto de la carga de trabajo hasta réplicas completas de todo el conjunto de aplicaciones. Al considerar la redundancia para sus aplicaciones, es importante tener en cuenta el nivel de redundancia que proporcionan la infraestructura, los almacenes de datos y las dependencias que utiliza. Por ejemplo, Amazon DynamoDB y Amazon Simple Storage Service (Amazon S3) proporcionan redundancia al replicar los datos en varias zonas de disponibilidad de una región, yAWS Lambdaejecuta sus funciones en varios nodos de trabajo en varias zonas de disponibilidad. Para cada servicio que utilice, tenga en cuenta lo que proporciona el servicio y lo que debe diseñar.
- Capacidad suficiente— Su carga de trabajo requiere recursos suficientes para funcionar según lo previsto. Los recursos incluyen memoria, ciclos de CPU, subprocesos, almacenamiento, rendimiento, cuotas de servicio y muchos otros.
- Salida puntual— Cuando los clientes utilizan su carga de trabajo, esperan que desempeñe la función prevista en un período de tiempo razonable. A menos que el servicio incluya un acuerdo de nivel de servicio (SLA) en materia de latencia, sus expectativas se basan generalmente en datos empíricos, es decir, en su propia experiencia. Estoexperiencia media del clientepor lo general, se considera la latencia media (P50) del sistema. Si su carga de trabajo tarda más de lo esperado, esta latencia puede afectar a la experiencia de sus clientes.
- Salida correcta— Se requiere la salida correcta del software de la carga de trabajo para que proporcione la funcionalidad prevista. Un resultado incorrecto o incompleto puede ser peor que no recibir respuesta alguna.
- Aislamiento de fallas— El aislamiento de fallas restringe el alcance del impacto al contenedor de fallas previsto cuando se produce una falla. Garantiza que los componentes específicos de la carga de trabajo fallen juntos y, al mismo tiempo, evita que el fallo se propague en cascada a otros componentes no deseados. También ayuda a limitar el alcance del impacto de su carga de trabajo en los clientes. El aislamiento de errores es algo diferente de las cuatro propiedades anteriores, ya que acepta que ya se ha producido un error, pero debe contenerse. Puede crear un aislamiento de errores en la infraestructura, las dependencias y las funciones de software.

Cuando se infringe una propiedad deseada, se puede provocar que una carga de trabajo no esté disponible o se perciba que no está disponible. Basándonos en estas propiedades de resiliencia deseadas y en nuestra experiencia trabajando con muchasAWSclientes, hemos identificado cinco categorías de fallas comunes: puntos de falla únicos, carga excesiva, latencia excesiva, errores de configuración y errores, y destino compartido, que abreviamos como SEEMS. Estas proporcionan un método coherente para clasificar los posibles modos de fallo y se describen en la siguiente tabla.

Categoría de fallo	Viola	Definición
Puntos únicos de falla (SPOF)	Redundancia	Una falla en un solo componente interrumpe el sistema debido a la falta de redundancia del componente.
Carga excesiva	Capacidad suficiente	El consumo excesivo de un recurso debido a una demanda o un tráfico excesivos impide que el recurso desempeñe la función esperada. Esto puede incluir alcanzar límites y cuotas, lo que provoca la limitación y el rechazo de las solicitudes.
Latencia excesiva	Salida puntual	El procesamiento del sistema o la latencia del tráfico de red superan el tiempo esperado, los objetivos de nivel de servicio (SLO) o los acuerdos de nivel de servicio (SLA) esperados.
Errores y errores de configuración	Salida correcta	Los errores de software o una mala configuración del sistema provocan un resultado incorrecto.
Fecha compartida	Aislamiento de fallas	Un fallo provocado por cualquiera de las categorías de fallos anteriores sobrepasa los límites de aislamiento de fallos previstos y se extiende en cascada a otras partes del sistema o a otros clientes.

Comprensión de la carga de trabajo

Para aplicar el marco, comience por comprender la carga de trabajo que desea analizar. Un diagrama de la arquitectura del sistema proporciona un punto de partida para documentar los detalles más relevantes del sistema. Sin embargo, tratar de analizar una carga de trabajo completa puede resultar complejo, ya que muchos sistemas tienen numerosos componentes e interacciones. En su lugar, le recomendamos que se centre enhistorias de usuarios, que sonexplicaciones informales y generales de las características del software escritas desde la perspectiva del usuario final. Su propósito es articular cómo una función de software proporciona valor al cliente. A continuación, puede modelar estas historias de usuario con diagramas de arquitectura y diagramas de flujo de datos para facilitar la evaluación de los componentes técnicos que proporcionan la funcionalidad empresarial descrita. Por ejemplo, una solución de compra de juegos móviles integrada en una aplicación puede tener dos historias de usuario: «comprar créditos integrados en la aplicación» y «obtener reembolsos integrados en la aplicación», tal y como se muestra en el siguiente diagrama. (Este ejemplo de arquitectura destaca cómo se puede descomponer un sistema en historias de usuario; no pretende representar una aplicación muy resistente).



Cada historia de usuario consta de cuatro componentes comunes: código y configuración, infraestructura, almacenes de datos y dependencias externas. Los diagramas deben incluir todos estos componentes y reflejar las interacciones entre los componentes. Por ejemplo, si hay una carga excesiva en su punto de enlace de Amazon API Gateway, considere cómo esa carga se transfiere en cascada a otros componentes del sistema, como suAWS Lambdafunciones o tablas de Amazon

DynamoDB. El seguimiento de estas interacciones le ayuda a comprender cómo el modo de fallo puede afectar a la historia del usuario. Puede capturar este flujo de forma visual con un diagrama de flujo de datos o utilizando flechas de flujo sencillas en un diagrama de arquitectura, como en la ilustración anterior. Para cada componente, considere la posibilidad de capturar detalles como el tipo de información que se transmite, la información que se recibe, si la comunicación es sincrónica o asíncrona y qué límites de falla se están cruzando. En el ejemplo, las tablas de DynamoDB se comparten en ambas historias de usuario, como se puede ver en las flechas que indican que el componente Lambda del historial de reembolsos integrado en la aplicación accede a las tablas de DynamoDB del historial de compras integradas en la aplicación. Esto significa que un error provocado por la historia de un usuario que realiza compras dentro de la aplicación podría repercutir en cascada en la historia del usuario que realiza devoluciones integradas en la aplicación, como consecuencia de una suerte compartida.

Además, es importante entender la configuración básica de cada componente. La configuración básica identifica restricciones como el número medio y máximo de transacciones por segundo, el tamaño máximo de una carga útil, el tiempo de espera del cliente y las cuotas de servicio predeterminadas o actuales del recurso. Si va a modelar un diseño nuevo, le recomendamos que documente los requisitos funcionales del diseño y tenga en cuenta los límites. Esto le ayuda a comprender cómo se pueden manifestar los modos de fallo en el componente.

Por último, debe priorizar las historias de los usuarios en función del valor empresarial que proporcionan. Esta priorización le ayuda a centrarse primero en las funciones más importantes de su carga de trabajo. A continuación, puede centrar su análisis en los componentes de la carga de trabajo que forman parte de la ruta crítica para lograr esa funcionalidad y aprovechar el valor al utilizar el marco con mayor rapidez. A medida que avance en el proceso, podrá examinar otras historias de usuarios con diferentes prioridades.

Aplicación del marco

La mejor manera de aplicar el marco de análisis de la resiliencia es empezar con un conjunto estándar de preguntas, organizadas por categoría de error, que debes formular sobre cada componente de la historia de usuario que estás analizando. Si algunas preguntas no se aplican a todos los componentes de su carga de trabajo, utilice las preguntas que sean más aplicables.

Puedes abordar la forma de pensar en los modos de fallo desde dos perspectivas:

- ¿Cómo afecta la falla a la capacidad del componente para respaldar la historia del usuario?
- ¿Cómo afecta la falla a las interacciones del componente con los demás componentes?

Por ejemplo, si tenemos en cuenta los almacenes de datos y la carga excesiva, puede pensar en los modos de error en los que la base de datos está sometida a una carga excesiva y se agota el tiempo de espera de las consultas. También podría pensar en cómo su cliente de base de datos podría sobrecargar la base de datos con reintentos o no cerrar las conexiones de la base de datos, agotando así el conjunto de conexiones. Otro ejemplo es un proceso de autenticación, que puede constar de varios pasos. Debe pensar en cómo el fallo de una aplicación de autenticación multifactorial (MFA) o de un proveedor de identidad (IdP) externo podría afectar a la historia de un usuario de este sistema de autenticación.

Al responder a las siguientes preguntas, debe tener en cuenta el origen del error. Por ejemplo, ¿la sobrecarga se debió a un aumento de clientes o a un operador humano que dejó demasiados nodos fuera de servicio durante una actividad de mantenimiento? Es posible que puedas identificar varias fuentes de error en cada pregunta, lo que podría requerir distintas mitigaciones. Al hacer las preguntas, lleve un registro de los posibles modos de falla que descubra, a qué componentes se aplican y la fuente de cada falla.

Puntos únicos de fallo

- ¿El componente está diseñado para ser redundante?
- ¿Qué ocurre si el componente falla?
- ¿Puede su aplicación tolerar la pérdida parcial o total de una única zona de disponibilidad?

Latencia excesiva

- ¿Qué ocurre si este componente experimenta un aumento de la latencia o si un componente con el que interactúa tiene un aumento de la latencia (o si se producen interrupciones de la red, como el restablecimiento del TCP)?
- ¿Ha configurado adecuadamente los tiempos de espera con una estrategia de reintentos?
- ¿Fallas rápido o despacio? ¿Se producen efectos en cascada, como enviar involuntariamente todo el tráfico a un recurso dañado porque falla rápidamente?
- ¿Cuáles son las solicitudes más caras que se hacen a este componente?

Carga excesiva

- ¿Qué puede abrumar a este componente? ¿Cómo puede este componente superar a otros componentes?
- ¿Cómo puede evitar desperdiciar recursos en un trabajo que nunca tendrá éxito?
- ¿Tiene un disyuntor configurado para el componente?
- ¿Puede algo crear un atraso insuperable?
- ¿Dónde puede este componente experimentar un comportamiento bimodal?
- ¿Qué límites o cuotas de servicio se pueden superar (incluida la capacidad de almacenamiento)?
- ¿Cómo se escala el componente bajo carga?

Configuración errónea y errores

- ¿Cómo se evita que las configuraciones incorrectas y los errores se implementen en la producción?
- ¿Se puede revertir automáticamente una implementación defectuosa o desviar el tráfico del contenedor de errores en el que se implementó la actualización o el cambio?
- ¿Qué barandas tiene instaladas para evitar los errores del operador?
- ¿Qué elementos (como credenciales o certificados) pueden caducar?

¿Tarifa compartida

- ¿Cuáles son sus límites de aislamiento de fallas?
- ¿Se han realizado cambios en las unidades de despliegue al menos tan pequeñas como las previstaslímites de aislamiento de fallaspero idealmente más pequeño, como un entorno de una sola caja (una sola instancia dentro del límite de aislamiento de fallas)?

- ¿Se comparte este componente entre las historias de los usuarios u otras cargas de trabajo?
- ¿Qué otros componentes están estrechamente relacionados con este componente?
- ¿Qué ocurre si este componente o sus dependencias sufren una falla parcial o gris?

Tras formular estas preguntas, también puede utilizar SEEMS para desarrollar otras preguntas específicas para su carga de trabajo y para cada componente. La mejor forma de utilizar SEEMS es como una forma estructurada de pensar en los modos de fallo y como fuente de inspiración al realizar un análisis de resiliencia. No es una taxonomía rígida. No pierda tiempo preocupándose por la categoría a la que pertenece un modo de falla en particular; no es importante. ¿Quéeslo importante es que pensaste en el fracaso y lo escribiste. No hay respuestas incorrectas; ser creativo y pensar de forma innovadora es beneficioso. Además, no dé por sentado que un modo de fallo ya está mitigado; incluya todos los posibles modos de fallo que se le ocurran.

Es poco probable que anticipe todos los posibles modos de falla en su primer ejercicio. Las múltiples iteraciones del marco le ayudan a generar un modelo más completo, por lo que no tiene que intentar resolverlo todo en la primera pasada. Puede ejecutar el análisis con una cadencia regular, semanal o quincenal. En cada sesión, concéntrese en un modo o componente de fallo específico. Esto puede ayudar a lograr un progreso constante e incremental en la mejora de la resiliencia de su carga de trabajo. Tras recopilar una lista de posibles modos de fallo para una historia de usuario, podrás decidir qué hacer al respecto.

Mitigar los posibles fallos

Ahora que tiene posibles fallos en los componentes de una historia de usuario, puede centrarse en las mitigaciones. En primer lugar, analice las posibles compensaciones en relación con el impacto potencial y la probabilidad de cada fallo que descubra. A continuación, determine el nivel de observabilidad requerido y seleccione una estrategia de mitigación. Las compensaciones deberían incluir el esfuerzo por instrumentar el nivel correcto de observabilidad y la estrategia de mitigación. Por último, determine la cadencia adecuada para realizar revisiones periódicas de los análisis de resiliencia.

Secciones

Comprender las compensaciones y los riesgos

Las arquitecturas resilientes deben usar un puñado de mecanismos probados, simples y confiables para responder a las fallas. Para lograr los niveles más altos de resiliencia, las cargas de trabajo deben detectar y recuperarse automáticamente de tantos modos de falla como sea posible. Hacerlo requiere una gran inversión en la realización de un análisis de resiliencia. Esto significa que lograr niveles más altos de resiliencia implica hacer concesiones. Sin embargo, a medida que sigas haciendo concesiones, llegarás a un punto de rentabilidad decreciente en relación con tus objetivos de resiliencia. Estas son las compensaciones más habituales:

- Costo: los componentes redundantes, la mejora de la observabilidad, las herramientas adicionales o el aumento de la utilización de los recursos se traducirán en un aumento de los costos.
- Complejidad del sistema: detectar los modos de falla y responder a ellos, incluidas las soluciones de mitigación, y, posiblemente, no utilizar servicios gestionados aumentan la complejidad del sistema.
- Esfuerzo de ingeniería: los desarrolladores necesitan más horas de trabajo para crear soluciones que detecten los modos de fallo y respondan a ellos.

- Sobrecarga operativa: monitorear y operar un sistema que maneja más modos de falla puede aumentar la sobrecarga operativa, especialmente cuando no se pueden usar servicios administrados para mitigar modos de falla específicos.
- Latencia y coherencia: la creación de sistemas distribuidos que favorezcan la disponibilidad requiere sacrificar coherencia y latencia, como se describe en el teorema de PACELC.

Al considerar las mitigaciones de los modos de falla identificados en la historia de usuario, considere las compensaciones que debe hacer. Al igual que ocurre con la seguridad, la resiliencia es un problema de optimización. Debe tomar la decisión de evitar, mitigar, transferir o aceptar los riesgos que plantea la falla identificada. Puede que haya algunos modos de fallo que puedas evitar, un conjunto que aceptes y algunos que puedas transferir. Puede optar por mitigar muchos de los modos de error que identifique. Para determinar qué enfoque adoptar, realice una evaluación planteándose dos preguntas: ¿Cuál es la probabilidad de que se produzca la falla? ¿Cuál es el impacto en la carga de trabajo si se produce?

La probabilidad es qué tan plausible es que ocurra un evento. Por ejemplo, si la historia de usuario tiene un componente que funciona en una sola instancia de Amazon Elastic Compute Cloud (Amazon EC2), es posible que el componente se interrumpa en algún momento del funcionamiento del sistema, tal vez debido a procedimientos de aplicación de parches o errores del sistema

operativo. Como alternativa, una base de datos gestionada por Amazon Relational Database Service (Amazon RDS) que sincroniza datos entre sus instancias principal y secundaria tiene una baja probabilidad de dejar de estar completamente disponible.

El impacto es una estimación del daño que puede causar un suceso. Debe evaluarse tanto desde una perspectiva financiera como de reputación, y es relativa al valor de las historias de los usuarios a las que afecta. Por ejemplo, una base de datos sobrecargada podría tener un impacto significativo en la capacidad de un sistema de comercio electrónico para aceptar nuevos pedidos. Sin embargo, la pérdida de una sola instancia de una flota de 20 instancias detrás de un balanceador de carga probablemente tendría muy poco impacto.

Puede comparar las respuestas a estas preguntas con el costo de las compensaciones que debe hacer para mitigar el riesgo. Si tiene en cuenta esta información a la vista de su umbral de riesgo y sus objetivos de resiliencia, le servirá de base para decidir qué modos de falla planea mitigar activamente.

Observabilidad del modo de fallo

Para mitigar un modo de fallo, primero debe detectar si está afectando o está a punto de afectar a su carga de trabajo. Una mitigación solo es efectiva si hay una señal de que se debe tomar una acción. Esto significa que parte de la creación de cualquier mitigación incluye, como mínimo, verificar que se tiene o se está creando la observabilidad necesaria para detectar el impacto de la falla.

Debe considerar los síntomas observables del modo de falla en dos dimensiones:

- ¿Cuáles son los principales indicadores que indican que el sistema se acerca a una situación en la que podría producirse un impacto en breve?
- ¿Cuáles son los indicadores de retraso que pueden mostrar el impacto del modo de fallo lo más rápido posible una vez que se ha producido?

Por ejemplo, un error de carga excesiva que se aplica a un elemento de la base de datos podría tener un recuento de conexiones como indicador principal. Puede ver el aumento constante del número de conexiones como un indicador principal de que la base de datos podría superar pronto el límite de conexiones, por lo que puede tomar medidas, como cancelar las conexiones utilizadas menos recientemente, para reducir el recuento de conexiones. El indicador de retraso indica cuándo se ha superado el límite de conexión a la base de datos y si los errores de conexión a la base de datos aumentan. Además de recopilar métricas de aplicaciones e infraestructuras, considere

la posibilidad de recopilar indicadores clave de rendimiento (KPI) para detectar cuándo las fallas afectan a la experiencia del cliente.

Siempre que sea posible, le recomendamos que incluya ambos tipos de indicadores en su estrategia de observabilidad. En algunos casos, es posible que no puedas crear indicadores principales, pero siempre debes tener un indicador rezagado para cada fallo que desees mitigar. Para elegir la mitigación adecuada, también debes considerar si un indicador adelantado o retrasado detectó el fallo. Por ejemplo, piensa en un aumento repentino del tráfico a tu sitio web. Es probable que solo veas un indicador de retraso. En este caso, el escalado automático por sí solo puede no ser la mejor forma de mitigar el problema, ya que la implementación de nuevos recursos lleva tiempo, mientras que la limitación podría evitar la sobrecarga casi de inmediato y dar tiempo a la aplicación para escalar o reducir la carga. Por el contrario, si se trata de un aumento gradual del tráfico, aparecerá un indicador principal. En este caso, la regulación no sería adecuada porque tienes tiempo para responder escalando automáticamente el sistema.

Estrategias de mitigación comunes

Para empezar, piense en utilizar mitigaciones preventivas para evitar que el modo de fallo afecte a la historia del usuario. Entonces deberías pensar en las mitigaciones correctivas. Las mitigaciones correctivas ayudan al sistema a recuperarse automáticamente o a adaptarse a las condiciones cambiantes. Esta es una lista de las mitigaciones más comunes para cada categoría de fallo que se ajustan a las propiedades de resiliencia.

Categoría de error	Propiedades de resiliencia deseadas	Mitigaciones
Puntos únicos de fallo (SPOF)	Redundancia y tolerancia a fallos	• Implemente la redundancia, por ejemplo, mediante el uso de varias instancias de EC2 detrás de Elastic Load Balancing (ELB). • Elimine las dependencias del plano de control del servicio AWS global y asuma las dependencias únicamente en los planos de datos del servicio global.

		• Utilice una degradación adecuada cuando un recurso no esté disponible, de modo que su sistema se mantenga estable estáticamente ante un único punto de fallo.
Carga excesiva	Capacidad suficiente	• Las principales estrategias de mitigación son la limitación de la velocidad, la reducción de la carga y la priorización del trabajo, el trabajo constante, el retraso exponencial y el reintento con fluctuaciones o sin reintentos, el control del servicio más pequeño, la gestión de la profundidad de las colas, elescalado automático, la evitación de las cachés inactivas y los disyuntores. • También debe tener en cuenta su plan de capacidad y pensar en los límites futuros de capacidad y escalado, tanto relacionados con los recursos de AWS como con los límites de su sistema, que podría alcanzar.

Latencia excesiva	Salida puntual	• Implemente tiempos de espera configurados adecuadamente o tiempos de espera adaptables (modifique los valores de los tiempos de espera en función de las condicion es de latencia actuales y previstas para permitir que una dependencia lenta progrese en lugar de renunciar a las solicitudes lentas). • Implemente el retroceso exponencial y vuelva a intentarlo con fluctuaciones y coberturas, utilizando tecnologías como el TCP multiruta cuando se conecte a servicios en la nube desde entornos locales y experimente latencia en rutas específicas, mediante interacciones asincróni cas con sistemas poco acoplados, almacenamiento en caché y sin desperdiciar trabajo.

Configuración errónea y errores	Salida correcta	• La principal forma de detectar errores funcionales y repetibles en el software es realizar pruebas rigurosas mediante mecanismos como el análisis estático, las pruebas unitarias, las pruebas de integración, las pruebas de regresión, las pruebas de carga y las pruebas de resiliencia. • Implemente estrategias como la infraestructura como código (IaC) y la automatización de la integración y la entrega continuas (CI/CD) para ayudar a mitigar las amenazas de mala configuración. • Utilice técnicas de despliegue, como los despliegues unidireccionales,los despliegues fraccionados que estén alineados con los límites de aislamiento de fallas o los despliegues azul/verde para reducir los errores y las configuraciones incorrectas.

Fecha compartida	Aislamiento de fallas	• Implemente la tolerancia a errores en su sistema y utilice límites de aislamiento de errores lógicos y físicos, como varios clústeres de procesamiento o de contenedores, varias cuentas de AWS, múltiples AWS Identity and Access Management entidades principales (IAM), múltiples zonas de disponibilidad y, quizás, varias. Regiones de AWS • Técnicas como las arquitecturas basadas en celdas y la fragmentación aleatoria también pueden mejorar el aislamiento de las fallas. • Tenga en cuenta patrones como el acoplamiento flexible y la degradación gradual para evitar fallos en cascada. Al priorizar las historias de usuario, también puede utilizar esa priorización para distinguir entre las historias de usuarios que son esenciales para la función empresarial principal y las historias de usuario que pueden degradarse fácilmente. Por ejemplo, en un sitio de comercio electrónico, no querrás que el deterioro

	del widget de promocion es del sitio web afecte a la capacidad de procesar nuevos pedidos.

Si bien la implementación de algunas de estas medidas de mitigación requiere un esfuerzo mínimo, otras (como la adopción de una arquitectura basada en celdas para aislar los fallos de forma predecible y reducir al mínimo los fallos de destino compartido) podrían requerir un rediseño de toda la carga de trabajo y no solo de los componentes de una historia de usuario concreta. Como se mencionó anteriormente, es importante sopesar la probabilidad y el impacto del modo de falla con las ventajas y desventajas que se deben hacer para mitigarlo.

Además de las técnicas de mitigación que se aplican a cada categoría de modo de falla, debes pensar en las mitigaciones necesarias para recuperar la historia del usuario o todo el sistema. Por ejemplo, un error podría detener un flujo de trabajo e impedir que los datos se escriban en los destinos previstos. En este caso, es posible que necesite herramientas operativas para reimpulsar el flujo de trabajo o corregir los datos manualmente. Es posible que también tenga que incorporar un mecanismo de puntos de control a su carga de trabajo para evitar la pérdida de datos en caso de que se produzcan errores. O puede que tengas que crear un cable adicional para pausar el flujo de trabajo y dejar de aceptar nuevos trabajos para evitar más daños. En estos casos, debe pensar en las herramientas operativas y las barandillas que necesita.

Por último, siempre debe asumir que los seres humanos van a cometer errores a medida que desarrolle su estrategia de mitigación. Si bien DevOps las prácticas modernas buscan automatizar las operaciones, las personas aún tienen que interactuar con sus cargas de trabajo por varios motivos. Una acción humana incorrecta podría provocar un error en cualquiera de las categorías de SEEMS, como eliminar demasiados nodos durante el mantenimiento y provocar una sobrecarga, o configurar incorrectamente una marca de función. Estos escenarios son, en realidad, un fracaso de las barreras preventivas. Un análisis de la causa raíz nunca debe terminar con la conclusión de que «un humano cometió un error». En cambio, debería abordar las razones por las que era posible cometer errores en primer lugar. Por lo tanto, su estrategia de mitigación debe considerar cómo los operadores humanos pueden interactuar con los componentes de la carga de trabajo y cómo prevenir o minimizar el impacto de los errores de los operadores humanos mediante barreras de seguridad.

Mejora continua

La resiliencia es un proceso continuo. A lo largo del ciclo de vida del sistema, el entorno en el que opera cambiará. Para garantizar que su sistema siga siendo resistente, debe integrar el marco en sus revisiones operativas y arquitectónicas periódicas. Es posible que encuentre nuevos modos de falla que no identificó la primera vez, o que pueda implementar medidas de mitigación nuevas o imprevistas anteriormente. El análisis de la resiliencia debe ser un proceso iterativo y no un ejercicio de una sola vez.

Deberías probar empíricamente tus estrategias de mitigación con procesos como la ingeniería del caos o los días de juego para comprobar que funcionan según lo esperado. Si no dispones de un mecanismo de pruebas riguroso, no estarás seguro de que la mitigación funcionará según lo esperado cuando la necesites. Durante el análisis de resiliencia, puede determinar que un modo de falla ya está controlado por una mitigación específica, pero también es importante poner a prueba esas suposiciones. Debe probar tanto las mitigaciones existentes como las nuevas que se crearon mediante el marco de análisis de resiliencia.

También debes evaluar qué tan bien realizaste el análisis mediante retrospectivas en equipo. ¿Sabían todos en qué estaban trabajando durante el análisis? ¿La cantidad de modos de falla que encontró mediante el análisis de resiliencia se ajustó a las expectativas del equipo? ¿Podría identificar las mitigaciones de todos los modos de falla que descubrió? ¿El equipo consideró útil el proceso? ¿Cree que mejorará la resiliencia de su carga de trabajo?

Cuando se produzcan eventos de fallo reales que afecten a la disponibilidad de la carga de trabajo, registre el modo de fallo específico, los componentes que formaron parte del error y el patrón de mitigación que se utilizó. Haga que estos metadatos se puedan buscar en su herramienta de análisis posterior al incidente para poder determinar en qué componentes y modos de falla centrarse en el futuro. A lo largo de este proceso, puede interactuar con su equipo de AWS cuentas y con los arquitectos de soluciones.

Conclusión y recursos

Esta guía presenta un marco para realizar el análisis de la resiliencia de forma continua y coherente. Este marco le ayuda a identificar cómo los puntos únicos de fallo, la carga excesiva, la latencia excesiva, la mala configuración y los errores, así como el destino compartido, pueden afectar a los componentes de su carga de trabajo. La identificación de estos modos de falla le ayuda a determinar una estrategia de mitigación adecuada como parte de la creación de una arquitectura orientada a la recuperación.

Para obtener información adicional sobre el análisis de la resiliencia, consulte los siguientes enlaces:

- Marco del ciclo de vida de la resiliencia (orientación AWS prescriptiva)
- Soluciones para la resiliencia (biblioteca de AWS soluciones)
- Hacia una resiliencia continua (Adrian Hornsby, The Cloud Architect, 24 de marzo de 2021)

Historial de documentos

En la siguiente tabla se describen los cambios importantes en esta guía. Si desea recibir notificaciones sobre futuras actualizaciones, puede suscribirse a unaFuente RSS.

Cambio	Descripción	Fecha
Publicación inicial	—	5 de septiembre de 2023

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 Postgre SQL 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 (RDSAmazon) 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 un 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

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

IA

Véase inteligencia artificial.

AIOps

Consulte las operaciones de inteligencia artificial.

anonimización

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

antipatrones

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

control de aplicaciones

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

cartera de aplicaciones

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

inteligencia artificial (IA)

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

operaciones de inteligencia artificial (AIOps)

El proceso de utilizar técnicas de machine learning para resolver problemas operativos, reducir los incidentes operativos y la intervención humana, y mejorar la calidad del servicio. Para obtener más información sobre cómo 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 ABACla AWS documentación de 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 con 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 CAForganiza 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 las comunicaciones de las personas a fin de ayudar a la organización a adoptar la nube con éxito. Para obtener más información, consulte el AWS CAFsitio web y el AWS CAFdocumento técnico.

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 WQFse 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 API llamadas sospechosas 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 endianismo.

clasificación binaria

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

filtro de floración

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

implementación azul/verde

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

bot

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

botnet

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

rama

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

acceso con cristales rotos

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

estrategia de implementación sobre infraestructura existente

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

caché de búfer

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

capacidad empresarial

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

planificación de la continuidad del negocio () BCP

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

C

CAF

Consulte el marco AWS de adopción de la nube.

despliegue canario

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

CCoE

Consulte Cloud Center of Excellence.

CDC

Consulte la captura de datos de cambios.

cambiar la captura de datos (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. Se puede utilizar CDC para varios 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 CCoEpublicaciones del 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 un CCoE 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 o Bitbucket 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, AWS Panorama ofrece dispositivos que añaden CV a las redes de cámaras locales, y Amazon SageMaker proporciona algoritmos de procesamiento de imágenes para CV.

desviación de configuración

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

base de datos de gestión de la 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, se utilizan datos CMDB de una etapa de migración de descubrimiento y análisis de la cartera.

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 Cuenta de AWS región o en una organización mediante una YAML plantilla. Para obtener más información, consulte los paquetes de conformidad en la AWS Config documentación.

integración y entrega continuas (CI/CD)

El proceso de automatización de las etapas de origen, compilación, prueba, presentación y producción del proceso de lanzamiento del software. CI/CD is commonly described as a pipeline. CI/CDpueden 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 bases 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 bases 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.

mapeo del flujo de valor de 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. DVSMamplía el proceso de mapeo 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 Modernizar la antigua Microsoft. ASP NET(ASMX) servicios web de forma incremental mediante contenedores y Amazon API Gateway.

DR

Consulte recuperación ante desastres.

detección de desviaciones

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

DVSM

Consulte el mapeo del flujo de valor del desarrollo.

E

EDA

Consulte el análisis exploratorio de datos.

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

Un servicio que puede alojar en una nube privada virtual (VPC) para compartirlo con otros usuarios. Puede crear un servicio de punto final con otros Cuentas de AWS o AWS Identity and Access Management (IAM) principales AWS PrivateLink y conceder permisos a ellos. Estas

cuentas o entidades principales pueden conectarse a su servicio de puntos finales de forma privada mediante la creación de puntos finales de interfazVPC. Para obtener más información, consulte Crear un servicio de punto final en la documentación de Amazon Virtual Private Cloud (AmazonVPC).

planificación de recursos empresariales (ERP)

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

cifrado de sobre

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

environment

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

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

epopeya

En las metodologías ágiles, son categorías funcionales que ayudan a organizar y priorizar el trabajo. Las epopeyas brindan una descripción detallada de los requisitos y las tareas de implementación. Por ejemplo, las cuestiones AWS CAF de seguridad 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 la planificación de recursos empresariales.

análisis exploratorio de datos () 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. EDAse 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, 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 pocos pasos

LLMProporcionando 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 detallado () 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 cimentación (FM)

Una gran red neuronal de aprendizaje profundo que se ha estado entrenando con conjuntos de datos masivos de datos generalizados y sin etiquetar. FMsson 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 IAM permisos. Las barreras de protección de detección detectan las vulneraciones de las políticas y los problemas de conformidad, y generan alertas para su corrección. Se implementan mediante Amazon AWS Config AWS Security Hub GuardDuty AWS Trusted Advisor, Amazon Inspector y AWS Lambda cheques personalizados.

H

JA

Consulte alta disponibilidad.

migración heterogénea de bases de datos

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

alta disponibilidad (HA)

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

modernización histórica

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

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 for 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, una revisión suele realizarse fuera del flujo de trabajo de DevOps publicación habitual.

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 IAM directores que define sus permisos en el Nube de AWS entorno.

aplicación inactiva

Aplicación que tiene un uso medio CPU de memoria entre el 5 y el 20 por ciento durante un período 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 .

entrante (ingreso) VPC

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 e Internet en general.

migración gradual

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

Industria 4.0

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

infraestructura

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

infraestructura como código (IaC)

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

Internet de las cosas industrial () IIoT

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

inspección VPC

En una arquitectura de AWS múltiples cuentas, una arquitectura centralizada VPC que gestiona las inspecciones del tráfico de red entre Internet y las redes locales VPCs (en una misma o diferente Regiones de AWS). La arquitectura AWS de referencia de seguridad recomienda configurar la 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. ITILproporciona la base paraITSM.

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 ITSM las herramientas, 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. An 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 el control de acceso basado en etiquetas.

privilegio mínimo

La práctica recomendada de seguridad que consiste en conceder los permisos mínimos necesarios para realizar una tarea. Para obtener más información, consulte Aplicar permisos con privilegios mínimos en la documentación. IAM

migrar mediante lift-and-shift

Consulte 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 de telemetría y cola de mensajes () 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

Migration Acceleration Program (MAP)

Un AWS programa que brinda soporte de consultoría, capacitación y servicios para ayudar a las organizaciones a construir una base operativa sólida para migrar a la nube y para ayudar a compensar el costo inicial de las migraciones. MAPincluye 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 habituales.

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 estar compuestos por analistas y propietarios de operaciones, 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.

Evaluación de la cartera de migración () MPA

Una herramienta en línea que proporciona información para validar el argumento empresarial para migrar a. Nube de AWS MPAproporciona una evaluación detallada de la cartera (tamaño

correcto de los servidores, precios, TCO comparaciones y 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 la oleada). La MPAherramienta (requiere iniciar sesión) está disponible de forma gratuita para todos los consultores y AWS consultores de los socios. APN

Evaluación de la preparación para la migración (MRA)

El proceso que consiste en obtener información sobre el estado de preparación de una organización para la nube, identificar los puntos fuertes y débiles y elaborar un plan de acción para cerrar las brechas identificadas, utilizando la AWS CAF. Para obtener más información, consulte la Guía de preparación para la migración. MRAes la primera fase de la estrategia de AWS migración.

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.

MPA

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

Consulte 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. OPC-UA proporciona un estándar de interoperabilidad con esquemas de cifrado, autenticación y autorización de datos.

acuerdo a nivel operativo () OLA

Un acuerdo que aclara lo que los grupos de TI funcionales se prometen ofrecer 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 AWS CloudTrail que registra 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.

gestión del cambio organizacional (OCM)

Marco para administrar las transformaciones empresariales importantes y disruptivas desde la perspectiva de las personas, la cultura y el liderazgo. OCMayuda a las organizaciones a prepararse para los nuevos sistemas y estrategias y a realizar la transición a ellos acelerando la adopción del cambio, abordando los problemas de la transición e impulsando los cambios culturales y organizacionales. En la estrategia de AWS migración, este marco se denomina aceleración de las personas, debido a la velocidad de cambio que requieren los proyectos de adopción de la nube. Para obtener más información, consulte la OCMguía.

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). OACadmite todos los depósitos de S3 Regiones de AWS, el cifrado del lado del servidor con AWS KMS (SSE-KMS) y el cifrado dinámico `PUT` y `DELETE` las solicitudes al depósito de S3.

identidad de acceso de origen () OAI

En CloudFront, una opción para restringir el acceso y proteger el contenido de Amazon S3. Cuando lo usaOAI, 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 OAC, que proporciona un control de acceso mejorado y más detallado.

ORR

Consulte la revisión de la preparación operativa.

NO

Consulte tecnología operativa.

saliente (salida) VPC

En una arquitectura AWS multicuenta, VPC que gestiona las conexiones de red que se inician desde una aplicación. La arquitectura de referencia de AWS 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.

P

límite de permisos

Una política IAM de administración asociada a IAM los directores para establecer los permisos máximos que puede tener el usuario o el rol. Para obtener más información, consulte los límites de los permisos en la IAM documentación.

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 PII incluyen nombres, direcciones e información de contacto.

PII

Consulte la información de identificación personal.

manual de estrategias

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

PLC

Consulte controlador lógico programable.

PLM

Consulte la gestión del ciclo de vida del producto.

política

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

persistencia políglota

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

evaluación de cartera

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

predicate

Una condición de consulta que devuelve `true` o`false`, normalmente, 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 de un Cuenta de AWS, un IAM rol o un usuario. Para obtener más información, consulte los términos y conceptos de Principal in Roles en la IAM documentación.

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 DNS las consultas de un dominio y sus subdominios dentro de uno o másVPCs. 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 industria manufacturera, una computadora adaptable y altamente confiable que monitorea las máquinas y automatiza los procesos de fabricación.

encadenamiento rápido

Utilizar la salida de un LLMmensaje como entrada para el siguiente mensaje 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 microservicio basado en microservicios MES, 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 SQL relacional.

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

RACImatriz

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.

RASCImatriz

Véase responsable, responsable, consultado, informado (RACI).

RCAC

Consulte el control de acceso por filas y columnas.

read replica

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

rediseñar

Ver 7 Rs.

objetivo de punto de recuperación (RPO)

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

objetivo de tiempo de recuperación (RTO)

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

refactorizar

Ver 7 Rs.

Región

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

regresión

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

volver a alojar

Consulte 7 Rs.

versión

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

trasladarse

Ver 7 Rs.

redefinir la plataforma

Ver 7 Rs.

recompra

Ver 7 Rs.

resilencia

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, responsable, consultada, 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 RASCImatriz y, si la excluye, se denomina RACImatriz.

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

Una tecnología de IA generativa en la que, antes de generar una respuesta, LLMhace referencia a una fuente de datos autorizada que se encuentra fuera de sus fuentes de datos de entrenamiento. Por ejemplo, un RAG modelo puede realizar una búsqueda semántica en la base de conocimientos o en los datos personalizados de una organización. Para obtener más información, consulte Qué es RAG.

Rotation

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 SQL expresiones básicas y flexibles que tienen reglas de acceso definidas. RCACconsta 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

SAML2.0

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 AWS API operaciones sin tener que crear un registro de usuario IAM para todos los miembros de la organización. Para obtener más información sobre la federación SAML basada en 2.0, consulte Acerca de la federación basada SAML en 2.0 en la documentación. 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 de seguridad y gestión de eventos (SIEM)

Herramientas y servicios que combinan los sistemas de gestión de la información de seguridad (SIM) y de gestión de eventos de seguridad (SEM). Un SIEM sistema recopila, monitorea y

analiza datos de servidores, redes, dispositivos y otras fuentes para detectar amenazas y brechas de seguridad y generar alertas.

automatización de las respuestas 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 automática incluyen la modificación de un grupo VPC de seguridad, 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 Servicio de AWS quien los recibe.

política de control de servicios (SCP)

Una política que proporciona un control centralizado de los permisos de todas las cuentas de una organización en AWS Organizations. SCPsdefina 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

El 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 fallo.

esquema en forma de estrella

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

patrón de higo estrangulador

Un enfoque para modernizar los sistemas monolíticos mediante la reescritura y el reemplazo gradual de las funciones del sistema hasta que se pueda desmantelar el sistema heredado. Este patrón utiliza la analogía de una higuera que crece hasta convertirse en un árbol estable y, finalmente, se apodera y reemplaza a su host. El patrón fue presentado por Martin Fowler como una forma de gestionar el riesgo al reescribir sistemas monolíticos. Para ver un ejemplo de cómo aplicar este patrón, consulta Modernizar la versión antigua de MicrosoftASP. NET(ASMX) servicios web de forma incremental mediante contenedores y Amazon API Gateway.

subred

Un rango de direcciones IP en su. VPC Una subred debe residir en una sola zona de disponibilidad.

control de supervisión 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

Técnica para proporcionar contexto, instrucciones o pautas LLMa un comportamiento y dirigirlo. 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.

VPCmirando

Una conexión entre dos VPCs que permite enrutar el tráfico mediante direcciones IP privadas. Para obtener más información, consulta Qué es el VPC peering en la VPC documentación de Amazon.

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

SQLFunción 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.

WORM

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

Proporciona instrucciones para realizar una LLMtarea, pero no proporciona ejemplos (imágenes) que puedan ayudar a guiarla. LLMDebe utilizar sus conocimientos previamente entrenados para

realizar la tarea. La eficacia de las indicaciones rápidas depende de la complejidad de la tarea y de la calidad de las mismas. Consulte también las indicaciones de pocos pasos.

aplicación zombi

Una aplicación que tiene un uso medio CPU de memoria inferior 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.