Marco de Buena Arquitectura de AWS

Marco de Buena Arquitectura de AWS

Marco de Buena Arquitectura de AWS: Marco de Buena Arquitectura de AWS

Table of Contents

Resumen 1
Resumen 1
Introducción 2
Definiciones 2
En la arquitectura 4
Principios generales de diseño 5
Los cinco pilares del marco 6
Excelencia operativa 6
Principios de diseño 6
Definición 7
Prácticas recomendadas 7
Recursos 13
Seguridad 14
Principios de diseño 14
Definición 15
Prácticas recomendadas 15
Recursos 20
Fiabilidad 20
Principios de diseño 21
Definición 21
Prácticas recomendadas 21
Recursos 25
Eficiencia de rendimiento 26
Principios de diseño 26
Definición 27
Prácticas recomendadas 27
Recursos 32
Optimización de costos 32
Principios de diseño 33
Definición 33
Prácticas recomendadas 34
Recursos 38
Proceso de revisión 39
Conclusión 41
Colaboradores 42
Documentación adicional 43
Revisiones del documento 44
Apéndice: Preguntas y prácticas recomendadas 46
Excelencia operativa 46
Organización 46
Preparación 49
Operación 52
Evolución 54
Seguridad 54
Seguridad 55
Administración de identidades y accesos 56
Detección 57
Protección de la infraestructura 58
Protección de los datos 59
Respuesta ante incidentes 61
Fiabilidad 61
Bases 62
Arquitectura de las cargas de trabajo 63
Administración de los cambios 65

Administración de los errores 67
Eficiencia de rendimiento 69
Selección 70
Revisión 73
Monitoreo 74
Compensaciones 74
Optimización de costos 75
Práctica de la administración financiera en la nube 75
Concientización sobre los gastos y el uso 76
Recursos rentables 78
Administración de los recursos de oferta y demanda 80
Optimización con el paso del tiempo 80
Avisos 82

Marco de Buena Arquitectura de AWS

Fecha de publicación: Julio de 2020 (Revisiones del documento (p. 44))

Resumen

El Marco de Buena Arquitectura de AWS lo ayuda a comprender las ventajas y desventajas de las decisiones que toma cuando crea sistemas en AWS. Mediante el uso del marco, aprenderá las prácticas recomendadas de arquitectura para diseñar y operar sistemas en la nube seguros, fiables, eficientes y rentables.

Introducción

El Marco de Buena Arquitectura de AWS lo ayuda a comprender las ventajas y desventajas de las decisiones que toma cuando crea sistemas en AWS. Mediante el uso del marco, aprenderá las prácticas recomendadas de arquitectura para diseñar y operar sistemas en la nube seguros, fiables, eficientes y rentables. Ofrece una forma para que pueda medir de manera constante sus arquitecturas en función de las prácticas recomendadas e identificar las áreas de mejora. El proceso para revisar una arquitectura es una conversación constructiva sobre decisiones arquitectónicas y no es un mecanismo de auditoría. Creemos que tener sistemas de buena arquitectura aumenta considerablemente la probabilidad del éxito empresarial.

Los arquitectos de soluciones de AWS tienen mucha experiencia en la arquitectura de soluciones en una amplia variedad de negocios verticales y casos de uso. Hemos ayudado a diseñar y revisar las arquitecturas de miles de clientes en AWS. A partir de esta experiencia, identificamos las prácticas recomendadas y las estrategias básicas para la arquitectura de sistemas en la nube.

El Marco de Buena Arquitectura de AWS documenta un conjunto de preguntas básicas para que comprenda si una arquitectura específica cumple con los requisitos de las prácticas recomendadas en la nube. El marco le ofrece un enfoque coherente para evaluar los sistemas en relación con las cualidades que se esperan de los sistemas modernos basados en las nubes, así como la reparación que se requeriría para alcanzar esas cualidades. A medida que AWS continúa evolucionando y nosotros continuamos obteniendo más información del trabajo con nuestros clientes, seguiremos perfeccionando la definición de buena arquitectura.

Este marco está destinado a aquellos que tienen roles de tecnología, como directores de tecnología (CTO), arquitectos, desarrolladores y miembros de equipos operativos. Describe las prácticas recomendadas y estrategias de AWS para utilizarlas al diseñar y operar una carga de trabajo en la nube y ofrece vínculos a más detalles de implementación y patrones arquitectónicos. Para obtener más información, consulte la página de inicio de AWS Well-Architected.

AWS también ofrece un servicio gratuito para revisar sus cargas de trabajo. El AWS Well-Architected Tool (AWS WA Tool) es un servicio en la nube que proporciona un proceso consistente para que revise y mida su arquitectura con base en el AWS Well-Architected Framework. AWS WA Tool ofrece recomendaciones para que las cargas de trabajo sean más fiables, seguras, eficientes y rentables.

Para facilitar la aplicación de las prácticas recomendadas, hemos creado los laboratorios de AWS Well-Architected, que ofrecen un repositorio de código y documentación para brindar experiencia práctica en la aplicación de las prácticas recomendadas. También nos unimos a los socios selectos de la red de socios de AWS (APN), que son miembros del programa para socios de AWS Well-Architected. Estos socios de APN cuentas con vastos conocimientos sobre AWS y pueden ayudarlo a revisar y mejorar sus cargas de trabajo.

Definiciones

Todos los días, los expertos de AWS ayudan a los clientes para diseñar la arquitectura de los sistemas y así aprovechar las prácticas recomendadas en la nube. Trabajamos con usted para analizar los pros y los contras relacionados con la arquitectura a medida que sus diseños evolucionan. A medida que implementa estos sistemas en los entornos en vivo, conocemos si el rendimiento de los sistemas es óptimo y las consecuencias de esos pros y contras.

Utilizamos nuestros conocimientos para crear el AWS Well-Architected Framework, que ofrece un conjunto consistente de prácticas recomendadas para que los clientes y socios evalúen las arquitecturas. Además, proporciona una serie de preguntas que se pueden utilizar para evaluar cómo una arquitectura se ajusta a las prácticas recomendadas de AWS.

AWS Well-Architected Framework se basa en cinco pilares: excelencia operativa, seguridad, fiabilidad, eficacia de rendimiento y optimización de costos.

Tabla 1 Pilares del Marco de Buena Arquitectura de AWS

Nombre	Descripción
Excelencia operativa	La capacidad para admitir el desarrollo y ejecutar cargas de trabajo de manera eficaz, obtener información acerca de las operaciones y mejorar continuamente admitiendo procesos y procedimientos para ofrecer valor de negocio.
Seguridad	El pilar de la seguridad abarca la capacidad para proteger los datos, sistemas y activos y aprovecha las tecnologías de la nube a fin de mejorar la seguridad.
Fiabilidad	El pilar de la fiabilidad incluye la capacidad de una carga de trabajo para llevar a cabo la función prevista de forma correcta y consistente en el momento esperado. Esto incluye la capacidad de operar y probar la carga de trabajo a través de su ciclo de vida completo. Este documento ofrece orientación exhaustiva sobre las prácticas recomendadas para implementar cargas de trabajo fiables en AWS.
Eficiencia de rendimiento	La habilidad de utilizar recursos informáticos de manera eficiente para cumplir con los requisitos del sistema y mantener esa eficiencia a medida que la demanda cambia y la tecnología evoluciona.
Optimización de costos	La capacidad para ejecutar sistemas para entregar valor de negocio al menor precio.

Utilizamos los siguientes términos en AWS Well-Architected Framework:

- A componente es el código, la configuración y los recursos de AWS que conjuntamente satisfacen un requisito. Un componente suele ser la unidad de propiedad técnica que se desacopla de otros componentes.
- El término carga de trabajo se utiliza para identificar un conjunto de componentes que conjuntamente entregan valor de negocio. Por lo general, una carga de trabajo es el nivel de detalle que comunican los líderes empresariales y tecnológicos.
- Concebimos la arquitectura como la manera en que los componentes trabajan juntos en una carga de trabajo. Los diagramas de arquitectura suelen centrarse en la forma en la que se comunican e interactúan los componentes.
- Hitos marcan los cambios clave en la arquitectura a medida que evoluciona a lo largo del ciclo de vida del producto (diseño, implementación, pruebas, puesta en marcha y producción).
- Dentro de una organización, la cartera tecnológica es el conjunto de cargas de trabajo que se necesita para que funcione la empresa.

Al diseñar la arquitectura de las cargas de trabajo, se analizan los pros y los contras y se hacen compensaciones entre los pilares en función del contexto empresarial. Estas decisiones empresariales pueden guiar las prioridades de su ingeniería. Puede llevar a cabo una optimización para reducir los costos en los entornos de desarrollo a costa de la fiabilidad o, en el caso de soluciones críticas, puede optimizar la fiabilidad a mayor costo. En soluciones de comercio electrónico, el rendimiento puede afectar los ingresos y la tendencia a que el cliente compre. La seguridad y la excelencia operativa no suelen contraponerse a los demás pilares.

En la arquitectura

En los entornos en las instalaciones, los clientes suelen tener un equipo central para la arquitectura de la tecnología que actúa como una superposición con otros equipos de productos o características para asegurarse de que cumplen con las prácticas recomendadas. Los equipos de arquitectura tecnológica generalmente incluyen personas como: el arquitecto técnico (infraestructura), el arquitecto de soluciones (software), el arquitecto de datos, el arquitecto de redes y el arquitecto de seguridad. Con frecuencia, estos equipos utilizan el esquema de arquitectura del Open Group (TOGAF) o el marco de Zachman como parte de la capacidad de arquitectura empresarial.

En AWS preferimos distribuir las capacidades en equipos y no tener un equipo centralizado en esa capacidad. Hay riesgos cuando se elige distribuir la autoridad de la toma de decisiones, p. ej., asegurar que los equipos cumplan con las normas internas. Mitigamos estos riesgos de dos maneras. En primer lugar, contamos con las prácticas que se enfocan en que cada equipo tenga esa capacidad. Además, recurrimos a expertos que se aseguran de que estos equipos aumenten el nivel de los estándares con los que necesitan cumplir. En segundo lugar, implementamos mecanismos que realizan comprobaciones automatizadas para garantizar que se cumpla con los estándares. El enfoque distribuido se basa en los principios de liderazgo de Amazony establece una cultura que abarca todas las funciones que funciona con el cliente como punto de partida. Los equipos obsesionados con el cliente construyen productos en respuesta a una necesidad del cliente.

En el caso de la arquitectura, eso significa que esperamos que todos los equipos tengan la capacidad de crear arquitecturas y cumplir con las prácticas recomendadas. Para ayudar a los nuevos equipos a obtener estas capacidades o a los equipos actuales a aumentar el nivel, permitimos el acceso a una comunidad virtual de ingenieros principales que pueden revisar sus diseños y ayudarle a entender cuáles son las prácticas recomendadas de AWS. La comunidad de ingeniería principal trabaja para que las prácticas recomendadas sean visibles y accesibles. Una forma de hacerlo, p. ej., es a través de charlas a la hora del almuerzo que se centran en la aplicación de las prácticas recomendadas a ejemplos reales. Estas charlas se graban y se pueden utilizar como parte de los materiales de incorporación para los nuevos miembros del equipo.

Las prácticas recomendadas de AWS surgen de nuestra experiencia en el manejo de miles de sistemas a escala de Internet. Preferimos utilizar datos para definir las prácticas recomendadas, pero también solicitamos ayuda a expertos en el tema, como ingenieros principales, para establecerlas. A medida que los ingenieros principales contemplan el surgimiento de nuevas prácticas recomendadas, trabajan como una comunidad para asegurarse de que los equipos las cumplan. Con el tiempo, estas prácticas recomendadas se formalizan en nuestros procesos de revisión internos, así como en mecanismos que garantizan el cumplimiento. Well-Architected es la implementación orientada al cliente de nuestro proceso de revisión interna, donde hemos codificado nuestro principal pensamiento de ingeniería a través de funciones de campo como la arquitectura de soluciones y los equipos internos de ingeniería. Well-Architected Framework es un mecanismo escalable que permite aprovechar estos aprendizajes.

Si sigue el enfoque de una comunidad de ingeniería principal con propiedad distribuida de la arquitectura, creemos que puede surgir una empresa Well-Architected que esté impulsada por la necesidad del cliente. Los líderes tecnológicos (como CTO o gerentes de desarrollo) que llevan a cabo las revisiones de Well-Architected en todas sus cargas de trabajo le ayudarán a que comprenda mejor los riesgos en su cartera tecnológica. Utilice este enfoque para poder identificar ejes temáticos en los equipos que su organización podría abordar mediante mecanismos, capacitación o charlas a la hora del almuerzo, donde sus ingenieros principales puedan compartir su opinión sobre áreas específicas con varios equipos.

Principios generales de diseño

El Marco de Buena Arquitectura identifica un conjunto de principios generales de diseño para permitir el buen diseño en la nube:

- Deje de sacar conclusiones sobre sus necesidades de capacidad: si toma una decisión de capacidad deficiente al implementar una carga de trabajo, es posible que termine optando por costosos recursos inactivos o que se tenga que enfrentar a las implicaciones de rendimiento de una capacidad limitada. Elimine estos problemas con la informática en la nube. Puede utilizar tanta o tan poca capacidad como necesite y escalar de manera vertical y horizontal automáticamente.
- Pruebe sistemas a escala de producción: en la nube puede crear un entorno de prueba en la escala de producción bajo demanda, realizar las pruebas y, a continuación, retirar los recursos. Debido a que solo debe pagar por el entorno de prueba cuando está en funcionamiento, puede simular un entorno en vivo por una fracción del costo que supondría realizar las pruebas en las instalaciones.
- Automatice para facilitar la experimentación arquitectónica: la automatización permite crear y replicar las cargas de trabajo a bajo costo y evitar los gastos del esfuerzo manual. Puede rastrear los cambios en su automatización, auditar el impacto y volver a los parámetros anteriores cuando sea necesario.
- Permita arquitecturas evolutivas: en un entorno tradicional, las decisiones relacionadas con la arquitectura suelen implementarse como eventos estáticos y puntuales, con unas pocas versiones importantes de un sistema durante su vida. A medida que un negocio y su contexto continúan evolucionando, estas decisiones iniciales podrían obstaculizar la capacidad del sistema para satisfacer las cambiantes necesidades comerciales. En la nube la capacidad para automatizar y probar a demanda reduce el riesgo de impacto de los cambios de diseño. De esta manera, se permite que los sistemas evolucionen con el tiempo para que los negocios puedan aprovechar las innovaciones como una práctica estándar.
- Impulse arquitecturas con datos: en la nube, puede recopilar datos sobre la manera en que sus decisiones relacionadas con la arquitectura afectan el comportamiento de la carga de trabajo. Esto permite tomar decisiones basadas en los hechos sobre cómo mejorar la carga de trabajo. Su infraestructura en la nube está codificada, por lo que puede utilizar esos datos para informar sus opciones de arquitectura y mejoras a lo largo del tiempo.
- Mejore mediante días de prueba: pruebe el funcionamiento de la arquitectura y los procesos. Para ello, programe días de prueba con regularidad para simular los eventos de la producción. Con esto podrá comprender dónde se pueden realizar las mejoras y puede ayudar a desarrollar la experiencia de la organización para hacer frente a los eventos.

Los cinco pilares del marco

Crear un sistema de software es como construir un edificio. Si los cimientos no son resistentes, los problemas estructurales pueden socavar la integridad y la función del edificio. Cuando se diseñan soluciones tecnológicas, si se descuidan los cinco pilares: la excelencia operativa, la seguridad, la fiabilidad, la eficiencia del rendimiento y la optimización de los costos, puede ser difícil construir un sistema que cumpla con sus expectativas y requisitos. Incorpore estos pilares en su arquitectura para producir sistemas estables y eficientes. De esta manera, podrá enfocarse en otros aspectos de diseño, como los requisitos de funcionamiento.

Temas

- Excelencia operativa (p. 6)
- Seguridad (p. 14)
- Fiabilidad (p. 20)
- Eficiencia de rendimiento (p. 26)
- Optimización de costos (p. 32)

Excelencia operativa

La excelencia operativa comprende la capacidad para dar soporte al desarrollo y ejecutar cargas de trabajo de manera eficaz, obtener información acerca de las operaciones y mejorar continuamente el soporte a los procesos y los procedimientos para ofrecer valor de negocio.

El pilar de la excelencia operativa proporciona una descripción general de los principios de diseño, las prácticas recomendadas y las preguntas. Puede encontrar orientación normativa acerca de la implementación en el documento técnico sobre el pilar de la excelencia operativa.

Temas

- Principios de diseño (p. 6)
- Definición (p. 7)
- Prácticas recomendadas (p. 7)
- Recursos (p. 13)

Principios de diseño

Existen cinco principios de diseño para la excelencia operativa en la nube:

- Realizar operaciones como código: en la nube, puede aplicar la misma disciplina de ingeniería que utiliza para el código de aplicaciones en todo el entorno. Puede definir toda la carga de trabajo (aplicaciones, infraestructura) como código y actualizarla con código. Puede implementar sus procedimientos operativos como código y automatizar la ejecución si los activa en respuesta a eventos. Si realiza operaciones como código, limita la posibilidad de error humano y habilita respuestas coherentes a los eventos.
- Realizar cambios pequeños, reversibles y frecuentes: diseñe cargas de trabajo para permitir que los componentes se actualicen de forma regular. Realice cambios en incrementos pequeños que puedan revertirse si se producen errores (sin afectar a los clientes cuando sea posible).
- Mejorar los procedimientos operativos con frecuencia: a medida que utilice los procedimientos operativos, busque oportunidades para mejorarlos. Mientras su carga de trabajo evoluciona, haga que

sus procedimientos también lo hagan de forma adecuada. Configure días de práctica regulares para revisar todos los procedimientos y validar que sean efectivos y que los equipos los conozcan.

- Anticipar los errores: realice ejercicios “premortem” para identificar los posibles orígenes de errores de manera que se puedan eliminar o mitigar. Pruebe las situaciones de error y compruebe que entiende sus efectos. Pruebe los procedimientos de respuesta para asegurarse de que sean efectivos y que los equipos conozcan su ejecución. Configure días de práctica con regularidad para probar las respuestas de la carga de trabajo y del equipo a eventos simulados.
- Aprender de todos los errores operativos: impulse las mejoras a partir de las lecciones aprendidas de todos los eventos y los errores operativos. Comparta lo aprendido con los equipos y toda la organización.

Definición

Existen cuatro áreas de prácticas recomendadas que se deben tener en cuenta para lograr la excelencia operativa en la nube:

- Organización
- Preparación
- Operación
- Evolución

Los líderes de su organización definen los objetivos empresariales. Su organización debe comprender los requisitos y las prioridades, además de utilizarlos para organizar y realizar trabajos que respalden el logro de los resultados empresariales. Su carga de trabajo debe brindar la información necesaria para poder respaldarla. Implementar servicios para habilitar la integración, la implementación y la entrega de la carga de trabajo permitirá aumentar el flujo de cambios beneficiosos en la fase de producción mediante la automatización de los procesos repetitivos.

Pueden existir riesgos inherentes a la operación de la carga de trabajo. Debe comprender esos riesgos y tomar una decisión con fundamentos para avanzar a la fase de producción. Sus equipos deben ser capaces de brindar soporte a su carga de trabajo. Las métricas comerciales y operativas que derivan de los resultados empresariales deseados le permitirán comprender el estado de la carga de trabajo y las actividades operativas, además de responder a incidentes. Sus prioridades cambiarán a medida que se modifiquen las necesidades empresariales y el entorno de negocio. Utilice estos aspectos como un bucle de retroalimentación para mejorar de manera continua la organización y el funcionamiento de su carga de trabajo.

Prácticas recomendadas

Temas

- Organización (p. 7)
- Preparación (p. 10)
- Operación (p. 12)
- Evolución (p. 13)

Organización

Los equipos deben comprender de la misma manera toda la carga de trabajo, su rol en ella y los objetivos empresariales compartidos para establecer las prioridades que permitirán el éxito empresarial. Las prioridades claras maximizan los beneficios de sus esfuerzos. Evalúe las necesidades internas y externas de los clientes que involucran a las partes interesadas clave, incluidos los equipos de negocio, desarrollo y operaciones, para determinar dónde se deben concentrar los esfuerzos. La evaluación de las necesidades

de los clientes garantizará que comprenda por completo el respaldo que se necesita para alcanzar los resultados empresariales. Asegúrese de conocer las directrices o las obligaciones definidas por la gobernanza de su organización, así como los factores externos, como los requisitos de conformidad normativa y los estándares del sector, que pueden exigir o enfatizar un enfoque específico. Compruebe que cuenta con los mecanismos necesarios para identificar los cambios en los requisitos de gobernanza interna y de conformidad externa. Si no se identifican requisitos, asegúrese de haber aplicado la debida diligencia a esta tarea. Revise sus prioridades con regularidad de manera que se puedan actualizar a medida que cambian las necesidades.

Evalúe las amenazas a su negocio (por ejemplo, riesgos y obligaciones empresariales y amenazas a la seguridad de la información) y guarde esta información en un registro de riesgos. Evalúe el impacto de los riesgos y las compensaciones entre intereses opuestos o enfoques alternativos. Por ejemplo, se puede enfatizar la aceleración de la comercialización de características nuevas por encima de la optimización de costos, o puede elegir una base de datos relacional para datos no relacionales con el fin de simplificar el esfuerzo de migración de un sistema sin refactorización. Administre los beneficios y los riesgos para tomar decisiones con fundamentos al momento de determinar dónde concentrar los esfuerzos. Algunos riesgos u opciones pueden ser aceptables por un tiempo. Tal vez sea posible mitigar los riesgos asociados o quizás se vuelva inaceptable permitir que un riesgo permanezca, en cuyo caso tendrá que tomar medidas para abordarlo.

Los equipos deben comprender el rol que juegan en el logro de los resultados empresariales. Los equipos deben comprender el rol que tienen en el éxito de otros equipos, conocer el rol de los demás equipos en su propio éxito y tener objetivos en común. Comprender la responsabilidad, la propiedad, la manera en que se toman las decisiones y quién tiene la autoridad para hacerlo ayudará a concentrar los esfuerzos y a maximizar los beneficios de sus equipos. Las necesidades de un equipo dependerán del cliente al que brinden soporte, la organización, la conformación del equipo y las características de su carga de trabajo. Es poco razonable esperar que un solo modelo operativo pueda respaldar a todos los equipos y las cargas de trabajo en la organización.

Asegúrese de que haya propietarios identificados para cada aplicación, carga de trabajo, plataforma y componente de infraestructura, y que cada proceso y procedimiento tenga un propietario definido responsable de su definición y propietarios responsables de su rendimiento.

Comprender el valor de negocio de cada componente, proceso y procedimiento, el motivo por el que se establecieron esos recursos o se realizan las actividades y la razón por la que esa propiedad existe informará las acciones de los miembros de su equipo. Defina claramente las responsabilidades de los miembros del equipo de manera que actúen de forma adecuada y tengan mecanismos para identificar la responsabilidad y la propiedad. Cuente con mecanismos para solicitar incorporaciones, cambios y excepciones de manera que no restrinja la innovación. Defina acuerdos entre los equipos donde se describa cómo trabajan juntos para respaldarse entre sí y respaldar los resultados empresariales.

Apoye a los miembros del equipo para que puedan ser más eficaces a la hora de actuar y de respaldar los resultados empresariales. Los líderes principales comprometidos deben establecer expectativas y medir el éxito. Deben ser los patrocinadores, los defensores y los impulsores de la adopción de las prácticas recomendadas y de la evolución de la organización. Permita a los miembros del equipo actuar cuando los resultados estén en riesgo para minimizar el impacto y aliéntelos a realizar escalamientos hacia los responsables de la toma de decisiones y las partes interesadas cuando crean que exista un riesgo, de manera que pueda abordarse y se eviten los incidentes. Proporcione comunicaciones oportunas, claras y factibles sobre los riesgos conocidos y los eventos planificados para que los miembros del equipo puedan actuar de manera oportuna y adecuada.

Fomente la experimentación para acelerar el aprendizaje y mantener a los miembros del equipo interesados y comprometidos. Los equipos deben mejorar sus conjuntos de habilidades para adoptar nuevas tecnologías y admitir cambios en la demanda y las responsabilidades. Proporcione tiempo de estructura dedicado para el aprendizaje con el objetivo de apoyar y respaldar este aspecto. Asegúrese de que los miembros de su equipo tengan los recursos, tanto herramientas como miembros del equipo, para tener éxito y realizar escalamientos con el fin de respaldar los resultados empresariales. Aproveche la diversidad entre las organizaciones para buscar varias perspectivas únicas. Utilice esta perspectiva para aumentar el nivel de innovación, desafiar sus suposiciones y reducir el riesgo de sesgo de confirmación.

Aumente los niveles de inclusión, diversidad y accesibilidad dentro de sus equipos para obtener perspectivas beneficiosas.

Si existen requisitos regulatorios o de conformidad externos que se aplican a la organización, debería utilizar los recursos suministrados en la sección de Conformidad en la nube de AWS para facilitar la educación de los equipos para que puedan determinar el impacto en las prioridades. El Marco de Buena Arquitectura hace énfasis en el aprendizaje, la medición y la mejora. Ofrece un enfoque uniforme para evaluar arquitecturas e implementar diseños que se puedan escalar con el paso del tiempo. AWS ofrece AWS Well-Architected Tool para ayudar a revisar el enfoque antes del desarrollo, el estado de las cargas de trabajo antes de la producción y el estado de las cargas de trabajo durante la producción. Puede comparar las cargas de trabajo con las prácticas recomendadas sobre arquitectura de AWS, monitorear el estado general de estas y obtener información sobre riesgos potenciales. AWS Trusted Advisor es una herramienta que proporciona acceso a un conjunto principal de comprobaciones que recomienda optimizaciones útiles para organizar las prioridades. Los clientes de Business Support y Enterprise Support tienen acceso a comprobaciones adicionales centradas en la seguridad, la fiabilidad, el rendimiento y la optimización de costos, que pueden ayudar a definir aún más sus prioridades.

AWS puede ayudarlo a instruir a sus equipos acerca de AWS y sus servicios para que entiendan mejor de qué manera sus elecciones pueden afectar a la carga de trabajo. Debe utilizar los recursos que ofrece AWS Support (el Centro de conocimiento de AWS, los foros de debate de AWS y el Centro de AWS Support) y la documentación de AWS para instruir a sus equipos. Póngase en contacto con AWS Support a través del Centro de AWS Support para obtener respuestas a sus preguntas sobre AWS. AWS también comparte los patrones y las prácticas recomendadas que hemos aprendido a través del funcionamiento de AWS en la Biblioteca de creadores de Amazon. Existe una gran variedad de más información útil disponible en el blog de AWS y el podcast oficial de AWS. AWS Training and Certification proporciona formación técnica gratuita a través de cursos digitales autoguiados acerca de los aspectos fundamentales de AWS. También puede registrarse para obtener formación técnica impartida por instructores a fin de respaldar aún más el desarrollo de las habilidades en AWS de sus equipos.

Debe usar herramientas o servicios que permitan controlar de manera centralizada los entornos en todas las cuentas, como AWS Organizations, para ayudar a administrar los modelos operativos. Los servicios similares a AWS Control Tower amplían esta capacidad de administración al permitir la definición de diseños (que respalden los modelos operativos) para la configuración de cuentas, la aplicación de gobernanza continua con AWS Organizations y la automatización del aprovisionamiento de nuevas cuentas. Los proveedores de servicios administrados, como AWS Managed Services, los socios de AWS Managed Services o los proveedores de servicios administrados en la red de socios de AWS, ofrecen experiencia en la implementación de entornos en la nube y son útiles para satisfacer los requisitos de seguridad y conformidad, además de los objetivos empresariales. Agregar servicios administrados a su modelo operativo puede ayudarlo a ahorrar tiempo y recursos. Además, permite que sus equipos internos no carguen con tantas responsabilidades y permanezcan centrados en los resultados estratégicos que destacarán su negocio, en lugar de seguir concentrados en desarrollar nuevas habilidades y capacidades.

Las siguientes preguntas se enfocan en estas consideraciones para la excelencia operativa. (Para ver la lista de las preguntas y prácticas recomendadas relacionadas con la excelencia operativa, consulte el Apéndice (p. 46)).

OPS 1 ¿Cómo se determina cuáles son las prioridades?
Todos deben entender su rol en el proceso que permite alcanzar el éxito empresarial. Cuente con objetivos compartidos a fin de establecer prioridades para los recursos. Esto maximizará los beneficios de sus esfuerzos.

OPS 2 ¿Cómo se estructura la organización de manera que respalde los resultados empresariales?
Los equipos deben comprender el rol que juegan en el logro de los resultados empresariales. Los equipos deben comprender el rol que tienen en el éxito de otros equipos, conocer el rol de los demás

OPS 2 ¿Cómo se estructura la organización de manera que respalde los resultados empresariales?
equipos en su propio éxito y tener objetivos en común. Comprender la responsabilidad, la propiedad, la manera en que se toman las decisiones y quién tiene la autoridad para hacerlo ayudará a concentrar los esfuerzos y a maximizar los beneficios de sus equipos.

OPS 3 ¿Cómo la cultura organizativa respalda los resultados empresariales?
Brinde soporte a los miembros de su equipo para que puedan ser más eficaces a la hora de tomar medidas y de respaldar los resultados empresariales.

Es posible que desee destacar un pequeño subconjunto de prioridades en algún momento. Utilice un enfoque equilibrado a largo plazo para garantizar el desarrollo de las capacidades necesarias y la administración de riesgos. Revise sus prioridades con regularidad y actualícelas a medida que cambien las necesidades. Cuando la responsabilidad y la propiedad no están definidas o no se conocen, se corre el riesgo de no tomar las medidas necesarias a tiempo y de que surjan esfuerzos redundantes y potencialmente contradictorios a la hora de abordar esas necesidades. La cultura organizativa tiene un efecto directo en la satisfacción laboral y la retención de los miembros del equipo. Facilite el compromiso y las capacidades de los miembros de su equipo para lograr el éxito de su negocio. Es necesario experimentar para dar lugar a la innovación y para que las ideas se transformen en resultados. Reconozca que un resultado no deseado es un experimento exitoso que identificó un camino que no conduce al éxito.

Preparación

Si desea prepararse para la excelencia operativa, debe comprender las cargas de trabajo y sus comportamientos esperados. Luego, podrá diseñarlas para que ofrezcan información sobre su estado y podrá crear procedimientos para respaldarlas.

Diseñe su carga de trabajo de manera que brinde la información necesaria para comprender su estado interno (por ejemplo, métricas, registros, eventos y seguimientos) en todos los componentes a fin de respaldar los problemas de investigación y observación. Itere a fin de desarrollar la telemetría necesaria para monitorear el estado de la carga de trabajo, identificar el momento en que los resultados corren riesgo y habilitar respuestas efectivas. Cuando instrumente su carga de trabajo, capture una gran cantidad de información que le permita conocer la situación (por ejemplo, cambios de estado, actividad del usuario, acceso con privilegios, contadores del uso) y tenga en cuenta que puede utilizar filtros para seleccionar la información más útil con el paso del tiempo.

Adopte enfoques que mejoren el flujo de los cambios en la fase de producción y que permitan la refactorización, la retroalimentación rápida sobre la calidad y la corrección de errores. Estos enfoques aceleran los cambios beneficiosos que se aplican a la fase de producción, limitan los problemas implementados y permiten una rápida identificación y solución de los problemas que acarrearon las actividades de implementación o se detectaron en sus entornos.

Adopte enfoques que ofrezcan una rápida valoración acerca de la calidad y permitan una rápida recuperación de aquellos cambios que no tengan los resultados deseados. La aplicación de estas prácticas mitiga el impacto de los problemas que surgen como consecuencia de la implementación de cambios. Planifique los cambios incorrectos de manera que pueda responder más rápido si es necesario, y evalúe y valide los cambios que haga. Tenga conocimiento de las actividades planeadas en sus entornos de manera que pueda administrar el riesgo de cambios que tengan un impacto en dichas actividades planeadas. Destaque los cambios reversibles, pequeños y frecuentes para limitar su alcance. Esto permite que la resolución de problemas sea más sencilla y que las correcciones sean más rápidas, además de la posibilidad de revertir el cambio. Esto también implica que pueda obtener el beneficio de cambios valiosos con mayor frecuencia.

Evalúe la disposición operativa de sus cargas de trabajo, procesos, procedimientos y personal con el fin de comprender los riesgos operativos relacionados con su carga de trabajo. Debe utilizar un proceso

consistente (que incluya listas de verificación manuales o automatizadas) a fin de saber cuándo estará listo para trabajar con su carga de trabajo o un cambio. Esto también permitirá encontrar algunas áreas que necesitan planificación para poder abordarse. Cuente con manuales de procedimientos que documenten sus actividades de rutina y con manuales de estrategias que lo guíen en los procesos de resolución de problemas. Comprenda los beneficios y los riesgos para tomar decisiones con fundamentos que permitan que los cambios avancen a la fase de producción.

AWS le permite ver toda su carga de trabajo (aplicaciones, infraestructura, política, gobernanza y operaciones) como código. Esto significa que puede aplicar la misma disciplina de ingeniería que se utiliza para el código de aplicaciones en todos los elementos de su pila y compartirlos con los equipos o las organizaciones con el fin de aumentar los beneficios de los esfuerzos de desarrollo. Use las operaciones como código en la nuble y la capacidad de experimentar de manera segura para desarrollar la carga de trabajo, los procedimientos operativos y los errores de prueba. Usar AWS CloudFormation permite tener entornos consistentes, con plantillas, de desarrollo en un entorno de pruebas, de prueba y de producción con un crecimiento de los niveles de control de operaciones.

Las siguientes preguntas se enfocan en estas consideraciones para la excelencia operativa.

OPS 4 ¿Cómo se diseña la carga de trabajo de manera que sea posible comprender su estado?
Diseñe su carga de trabajo de manera que brinde la información necesaria de todos los componentes (por ejemplo, métricas, registros y rastreos) y pueda comprender su estado interno. Esto le permite ofrecer respuestas efectivas cuando sea necesario.

OPS 5 ¿Cómo se reducen los defectos, se facilita la corrección y se mejora el flujo en la producción?
Adopte enfoques que mejoren el flujo de los cambios en la producción y que permitan la refactorización, la retroalimentación rápida sobre la calidad y la corrección de errores. Estos enfoques aceleran los cambios beneficiosos que se aplican a la fase de producción, limitan los problemas implementados y permiten una rápida identificación y solución de los problemas que acarrearon las actividades de implementación.

OPS 6 ¿Cómo se mitigan los riesgos de implementación?
Adopte enfoques que ofrezcan una rápida valoración acerca de la calidad y permitan una rápida recuperación de aquellos cambios que no tengan los resultados deseados. La aplicación de estas prácticas mitiga el impacto de los problemas que surgen como consecuencia de la implementación de cambios.

OPS 7 ¿Cómo saber que se está listo para dar respaldo a una carga de trabajo?
Evalúe la disposición operativa de sus cargas de trabajo, procesos y procedimientos y personal con el fin de comprender los riesgos operativos relacionados con su carga de trabajo.

Invierta en la implementación de actividades de operaciones como código para maximizar la productividad del personal de operaciones, minimizar las tasas de error y habilitar las respuestas automáticas. Realice análisis “pre-mortem” para anticipar los errores y crear procedimientos cuando sea adecuado. Aplique metadatos con etiquetas de recursos y AWS Resource Groups mediante una estrategia de etiquetado consistente para permitir la identificación de los recursos. Etiquete sus recursos para la organización, la contabilidad de costos y los controles de accesos, con el objetivo de ejecutar actividades de operaciones

automatizadas. Adopte prácticas de implementación que aprovechen la elasticidad de la nube para facilitar las actividades de desarrollo y la implementación previa de sistemas con el fin de lograr implementaciones más rápidas. Cuando realice cambios en las listas de verificación que utiliza para evaluar sus cargas de trabajo, planifique lo que hará con los sistemas activos que ya no presentan conformidad.

Operación

El funcionamiento correcto de una carga de trabajo se mide a través del logro de los resultados de la empresa y de los clientes. Defina los resultados esperados, determine cómo se medirá el éxito e identifique las métricas que se usarán en esos cálculos con el fin de determinar si la carga de trabajo y el funcionamiento son correctos. El estado operativo incluye el estado de la carga de trabajo y el estado y el éxito de las actividades operativas realizadas para admitir la carga de trabajo (por ejemplo, la implementación y la respuesta a incidentes). Establezca puntos de referencia para las métricas respecto de las mejoras, la investigación y la intervención, recopile y analice las métricas y, luego, valide la comprensión del éxito de las operaciones y cómo cambia con el paso del tiempo. Use las métricas recopiladas para determinar si satisface las necesidades del cliente y empresariales e identifique las áreas que necesitan mejoras.

Se requiere una administración eficaz y efectiva de los eventos operativos para lograr la excelencia operativa. Esto se aplica a los eventos operativos planificados y no planificados. Use manuales de procedimientos para eventos ya conocidos y use manuales de estrategias para ayudar en la investigación y la resolución de problemas. Priorice las respuestas a los eventos basados en el impacto de la empresa y del cliente. Asegúrese de que si se genera una alerta en respuesta a un evento, exista un proceso asociado para ejecutar, con un propietario identificado de forma específica. Defina con anticipación el personal que se requiere para resolver un evento e incluya desencadenadores de escalamiento para involucrar a personal adicional, según sea necesario, en función de la urgencia y el impacto. Identifique e involucre a personas con la autoridad de tomar decisiones sobre procedimientos a seguir donde habrá un impacto empresarial a partir de una respuesta a un evento que no se abordó anteriormente.

Comunique el estado operativo de las cargas de trabajo a través de paneles y notificaciones adaptadas a la audiencia de destino (por ejemplo, clientes, empresas, desarrolladores, operaciones) para que se puedan tomar las medidas adecuadas, se puedan administrar las expectativas y se los informe cuando se reanuden las operaciones normales.

En AWS, puede generar vistas del panel de las métricas recopiladas de las cargas de trabajo y de manera nativa de AWS. Puede aprovechar CloudWatch o aplicaciones de terceros para agregar y presentar vistas de niveles de empresas, cargas de trabajo y operaciones de las actividades operativas. AWS proporciona información sobre la carga de trabajo a través de funciones de registro, incluidas AWS X-Ray, CloudWatch, CloudTrail y registros de flujo de VPC, lo que permite la identificación de problemas de la carga de trabajo para el análisis y la corrección de la causa raíz.

Las siguientes preguntas se enfocan en estas consideraciones para la excelencia operativa.

OPS 8 ¿Cómo se comprende el estado de la carga de trabajo?
Defina, registre y analice las métricas de las cargas de trabajo para obtener visibilidad en los eventos de carga de trabajo y poder tomar las medidas adecuadas.

OPS 9 ¿Cómo se comprende el estado de las operaciones?
Defina, registre y analice las métricas de las operaciones para obtener visibilidad en los eventos operativos y poder tomar las medidas adecuadas.

OPS 10 ¿Cómo se administran los eventos de operaciones y carga de trabajo?
Prepare y valide procedimientos para responder a los eventos con el fin de minimizar la interrupción de su carga de trabajo.

Todas las métricas recopiladas se deben alinear con una necesidad empresarial y con los resultados que respaldan. Desarrolle respuestas con scripts a los eventos ya conocidos y automatice su rendimiento en respuesta al reconocimiento del evento.

Evolución

Debe aprender, compartir y mejorar continuamente para mantener la excelencia operativa. Dedique los ciclos de trabajo a hacer mejoras graduales continuas. Realice análisis posteriores a los incidentes de todos los eventos que afecten a los clientes. Identifique los factores que contribuyeron a los incidentes y las acciones preventivas para limitar o prevenir que se repitan. Comunique los factores que contribuyeron a los incidentes a las comunidades afectadas según corresponda. Evalúe de forma regular y priorice las oportunidades de mejora (por ejemplo, solicitudes de características, corrección de problemas y requisitos de conformidad), que incluye los procedimientos de la carga de trabajo y de las operaciones.

Incluya bucles de retroalimentación en sus procedimientos para identificar rápidamente áreas que requieren mejora y capture los aprendizajes de la ejecución de las operaciones.

Comparta las lecciones aprendidas con los equipos para compartir los beneficios de dichas lecciones. Analice las tendencias en las lecciones aprendidas y realice análisis retrospectivo entre equipos de las métricas de las operaciones con el fin de identificar las oportunidades y los métodos para lograr mejoras. Implemente cambios diseñados para producir mejoras y evaluar los resultados para determinar el éxito.

En AWS, puede exportar sus datos de registro a Amazon S3 o enviar registros directamente a Amazon S3 para su almacenamiento a largo plazo. Mediante AWS Glue, puede detectar y preparar los datos de registro en Amazon S3 para el análisis. También, puede almacenar los metadatos asociados en AWS Glue Data Catalog de AWS. Amazon Athena, mediante su integración nativa con AWS Glue, se puede utilizar para analizar los datos de registro, consultándolos mediante SQL estándar. Con una herramienta de inteligencia empresarial como Amazon QuickSight, puede visualizar, explorar y analizar los datos. Descubrimiento de tendencias y eventos de interés que pueden implementar mejoras.

La siguiente pregunta se enfoca en estas consideraciones para la excelencia operativa.

OPS 11 ¿Cómo se impulsa el progreso de las operaciones?
Dedique tiempo y recursos a la mejora gradual y continua a fin de desarrollar la efectividad y la eficiencia de sus operaciones.

La evolución exitosa de las operaciones está fundamentada en lo siguiente: las mejoras pequeñas y frecuentes; el suministro de entornos seguros y tiempo para experimentar, desarrollar y probar las mejoras; y entornos en los que se alienta al aprendizaje a partir de los errores. El respaldo de operaciones de los entornos de pruebas, de desarrollo y producción, con un aumento del nivel de controles operativos, facilita el desarrollo y aumenta la capacidad de predicción de los resultados exitosos de los cambios implementados en fase de producción.

Recursos

Consulte los siguientes recursos para obtener más información acerca de nuestras prácticas recomendadas para la excelencia operativa.

Documentación

- DevOps y AWS

Documento técnico

- Pilar de la excelencia operativa

Video

- DevOps en Amazon

Seguridad

El pilar de la seguridad abarca la capacidad para proteger los datos, sistemas y activos para aprovechar las tecnologías en la nube a fin de mejorar la seguridad.

El pilar de la seguridad ofrece una descripción general de los principios de diseño, prácticas recomendadas y preguntas. Puede encontrar orientación normativa acerca de la implementación en el Documento técnico sobre el pilar de la seguridad.

Temas

- Principios de diseño (p. 14)
- Definición (p. 15)
- Prácticas recomendadas (p. 15)
- Recursos (p. 20)

Principios de diseño

Existen siete principios de diseño para la seguridad en la nube:

- Implemente una base de identidad sólida: implemente el principio de mínimo privilegio y aplique la segregación de tareas con la autorización indicada para cada interacción con los recursos de AWS. Centralice la administración de la identidad y elimine la dependencia de las credenciales estáticas y duraderas.
- Facilite la trazabilidad: monitoree, alerte y audite acciones y cambios del entorno en tiempo real. Integre la recopilación de registros y métricas con los sistemas para investigar y tomar medidas automáticamente.
- Aplique la seguridad en todos los niveles: aplique un enfoque de defensa en profundidad con múltiples controles de seguridad. Aplíquelo en todos los niveles (por ejemplo: al extremo de la red, VPC, balanceo de carga, todas las instancias y servicios informáticos, sistema operativo, aplicaciones y código).
- Automatice las prácticas recomendadas de seguridad: los mecanismos de seguridad automatizados basados en software mejoran la capacidad para escalar de manera más rápida y rentable de forma segura. Cree arquitecturas seguras mediante la implementación de controles que se definen y administran como código en plantillas de versión controlada.
- Proteja los datos en tránsito y en reposo: clasifique los datos en niveles de confidencialidad y utilice mecanismos como el cifrado, la tokenización y el control de acceso, según corresponda.
- Aleje a las personas de los datos: utilice mecanismos y herramientas para reducir o eliminar la necesidad de acceso directo o del procesamiento manual de datos. Esto reduce los riesgos de uso incorrecto, modificación o error humano durante la manipulación de datos confidenciales.

- Prepárese para eventos de seguridad: esté listo para los incidentes mediante el establecimiento de procesos y una política de administración e investigación de incidentes que se ajusten a las necesidades de la organización. Ejecute simulaciones de respuesta ante incidentes y utilice herramientas con automatización a fin de aumentar la velocidad de detección, investigación y recuperación.

Definición

Existen seis áreas de prácticas recomendadas para la seguridad en la nube:

- Seguridad
- Administración de identidades y accesos
- Detección
- Protección de la infraestructura
- Protección de los datos
- Respuesta ante incidentes

Antes de diseñar cualquier carga de trabajo, implemente prácticas que fomenten la seguridad. Usted querrá controlar qué acciones pueden realizar usuarios específicos. Además, necesitará poder identificar incidentes de seguridad, proteger sus sistemas y servicios y mantener la confidencialidad e integridad de los datos mediante la protección de datos. Debe contar con un proceso bien definido y practicado para responder a los incidentes de seguridad. Estas herramientas y técnicas son importantes porque respaldan objetivos como la prevención de la pérdida económica o la conformidad con las obligaciones normativas.

El modelo de responsabilidad compartida de AWS permite a las organizaciones que adoptan la nube alcanzar sus objetivos en torno a la seguridad y conformidad. Dado que AWS protege físicamente la infraestructura que respalda nuestros servicios en la nube, como cliente de AWS, usted puede concentrarse en usar los servicios para lograr sus objetivos. La nube de AWS también ofrece un mayor acceso a los datos de seguridad y un enfoque automatizado para responder a eventos de seguridad.

Prácticas recomendadas

Temas

- Seguridad (p. 15)
- Administración de identidades y accesos (p. 16)
- Detección (p. 17)
- Protección de la infraestructura (p. 18)
- Protección de los datos (p. 18)
- Respuesta ante incidentes (p. 19)

Seguridad

A fin de operar la carga de trabajo de forma segura, debe aplicar prácticas recomendadas generales en todas las áreas de la seguridad. Tome los requisitos y los procesos que ha definido en la excelencia operativa a nivel de la organización y carga de trabajo y aplíquelos en todas las áreas.

Mantenerse al día con las recomendaciones del sector y AWS y la inteligencia de amenazas facilita la evolución del modelo de amenazas y los objetivos de control. La automatización de los procesos de seguridad, las pruebas y la validación permiten escalar las operaciones de seguridad.

La siguiente pregunta se enfoca en estas consideraciones para la seguridad. (Para ver una lista con las preguntas y las prácticas recomendadas sobre la seguridad, consulte el Apéndice (p. 54)).

SEGURIDAD 1 ¿Cómo se opera la carga de trabajo de manera segura?
A fin de operar la carga de trabajo de forma segura, debe aplicar prácticas recomendadas generales en todas las áreas de la seguridad. Tome los requisitos y los procesos que ha definido en la excelencia operativa a nivel de la organización y la carga de trabajo y aplíquelos en todas las áreas. Mantenerse al día con las recomendaciones de AWS y las fuentes del sector, así como la inteligencia de amenazas, facilita la evolución del modelo de amenazas y los objetivos de control. La automatización de los procesos de seguridad, las pruebas y la validación permiten escalar las operaciones de seguridad.

AWS recomienda segregar las diferentes cargas de trabajo por cuenta, según su función y los requisitos de conformidad o de confidencialidad de los datos.

Administración de identidades y accesos

Identity and Access Management es una pieza clave de un programa de seguridad de la información, que garantiza que únicamente los usuarios y componentes autorizados y autenticados puedan acceder a los recursos y solo de la forma prevista. Por ejemplo, debe definir los elementos principales (es decir, las cuentas, usuarios, roles y servicios que pueden realizar acciones en la cuenta), crear políticas que se ajusten a esos elementos principales e implementar una administración de credenciales sólida. Estos elementos de administración de privilegios son el núcleo de la autenticación y autorización.

En AWS, la administración de privilegios está respaldada principalmente por el servicio AWS Identity and Access Management (IAM), que permite controlar el acceso programático y de los usuarios a los servicios y recursos de AWS. Debe aplicar políticas detalladas, que asignen permisos a un usuario, grupo, rol o recurso. También tiene la capacidad de exigir prácticas de contraseña seguras, como el nivel de complejidad, evitando la reutilización y aplicando la autenticación multifactor (MFA). Puede usar la federación con su servicio de directorio existente. Para las cargas de trabajo que requieren que los sistemas tengan acceso a AWS, IAM permite el acceso seguro a través de roles, perfiles de instancia, identidad federada y credenciales temporales.

Las siguientes preguntas se enfocan en estas consideraciones para la seguridad.

SEGURIDAD 2 ¿Cómo se administran las identidades para las personas y las máquinas?
Hay dos tipos de identidades que necesitará administrar cuando aborde las cargas de trabajo operativas de AWS. Conocer el tipo de identidad que debe administrar y a la cual debe conceder acceso lo ayuda asegurarse de que las identidades correctas tengan acceso a los recursos correctos bajo las condiciones correctas. Identidades humanas: los administradores, los desarrolladores, los operadores y los usuarios finales requieren una identidad para obtener acceso a los entornos y a las aplicaciones de AWS. Estos son miembros de su organización o usuarios externos con los que colabora, que interactúan con sus recursos de AWS mediante un navegador web, una aplicación cliente o herramientas interactivas de línea de comandos. Identidades de máquinas: las aplicaciones de servicios, las herramientas operativas y las cargas de trabajo requieren una identidad para realizar solicitudes a los servicios de AWS, como, por ejemplo, para leer datos. Estas identidades incluyen máquinas que se ejecutan en su entorno de AWS, como las instancias de Amazon EC2 o las funciones de AWS Lambda. También puede administrar las identidades de máquinas para los usuarios externos que necesiten acceso. Además, también puede tener máquinas fuera de AWS que necesiten acceso a su entorno de AWS.

SEGURIDAD 3 ¿Cómo se administran los permisos para las personas y las máquinas?
Administre los permisos para controlar el acceso a las identidades de las personas y de las máquinas que requieran acceso a AWS y a su carga de trabajo. Los permisos controlan a qué se tiene acceso, quién puede acceder y bajo qué condiciones lo hace.

Las credenciales no deben compartirse entre usuarios o sistemas. El acceso del usuario debe otorgarse con el uso de un enfoque de privilegios mínimos con las prácticas recomendadas, incluidos los requisitos de contraseña y MFA. El acceso programático, incluidas las llamadas API a los servicios de AWS, se debe realizar con credenciales temporales y de privilegios limitados, tales como las que otorga AWS Security Token Service.

AWS proporciona recursos que pueden ayudarlo con la gestión de la identidad y el acceso. Para facilitar el aprendizaje de las prácticas recomendadas, consulte nuestros laboratorios prácticos sobre la administración de credenciales y autenticación, el control del acceso humano, y el control del acceso programático.

Detección

Puede usar los controles de detección para identificar un posible incidente o amenaza de seguridad. Son una parte esencial de los marcos de gestión y se pueden utilizar para apoyar un proceso de calidad, una obligación legal o de conformidad o para identificar amenazas y responder a ellas. Existen diferentes tipos de controles de detección. Por ejemplo, realizar el inventario de los activos y sus atributos detallados fomenta una toma de decisiones más eficaz (y controles del ciclo de vida) para ayudar a establecer líneas de base operativas. También puede utilizar la auditoría interna, un examen de los controles relacionados con los sistemas de información, para garantizar que las prácticas cumplan con las políticas y los requisitos y que haya configurado las notificaciones de alerta automáticas correctas en función de las condiciones definidas. Estos controles son factores reactivos importantes que pueden ayudar a la organización a identificar y comprender el alcance de la actividad anómala.

En AWS, puede implementar controles de detección mediante el procesamiento de registros, eventos y monitoreo que permite realizar auditorías, análisis automatizados y alarmas. Los registros de CloudTrail, las llamadas a la API de AWS y CloudWatch proporcionan monitoreo de métricas con alarmas, y AWS Config proporciona un historial de configuración. Amazon GuardDuty es un servicio de detección de amenazas administrado que monitorea continuamente la actividad maliciosa o el comportamiento no autorizado para proteger sus cuentas y cargas de trabajo de AWS. Los registros de nivel de servicio también están disponibles, por ejemplo, puede usar Amazon Simple Storage Service (Amazon S3) para registrar solicitudes de acceso.

La siguiente pregunta se enfoca en estas consideraciones para la seguridad.

SEGURIDAD 4 ¿Cómo se detectan e investigan los eventos de seguridad?
Capture y analice los eventos a partir de registros y métricas para obtener visibilidad. Tome medidas con respecto a los eventos de seguridad y las amenazas potenciales a fin de ayudar a asegurar su carga de trabajo.

La administración de registros es importante para una carga de trabajo Well-Architected por razones que van desde seguridad o análisis forense hasta requisitos normativos o legales. Es fundamental que analice los registros y responda a ellos para poder identificar posibles incidentes de seguridad. AWS ofrece funciones que facilitan la implementación de la administración de registros mediante la capacidad para definir un ciclo de vida de conservación de los datos o para definir el lugar donde estos se conservarán, archivarán o posiblemente eliminarán. Esto permite que el manejo de los datos confiables y predecibles sea más simple y rentable.

Protección de la infraestructura

La protección de la infraestructura abarca las metodologías de control, como la defensa en profundidad, que son necesarias para aplicar las prácticas recomendadas y cumplir las obligaciones organizativas o normativas. El uso de estas metodologías es fundamental para el éxito de las operaciones en desarrollo en la nube o las instalaciones.

En AWS, puede implementar la inspección de paquetes con estado y sin estado, ya sea con el uso de tecnologías nativas en AWS o productos y servicios de socios disponibles a través de AWS Marketplace. Utilice Amazon Virtual Private Cloud (Amazon VPC) para crear un entorno privado, seguro y escalable en el que pueda definir la topología, incluidas las gateways, las tablas de enrutamiento y las subredes públicas y privadas.

Las siguientes preguntas se enfocan en estas consideraciones para la seguridad.

SEGURIDAD 5 ¿Cómo se protegen los recursos de red?
Cualquier carga de trabajo que tenga alguna forma de conectividad de red, ya sea de Internet o una red privada, requiere varios niveles de defensa para ayudar a protegerse de las amenazas externas e internas relacionadas con la red.

SEGURIDAD 6 ¿Cómo se protegen los recursos informáticos?
Los recursos informáticos de la carga de trabajo requieren varios niveles de defensa para facilitar la protección contra las amenazas internas y externas. Los recursos informáticos incluyen instancias de EC2, contenedores, funciones de AWS Lambda, servicios de base de datos, dispositivos de IoT y más.

En cualquier tipo de entorno se recomiendan múltiples capas de defensa. En el caso de la protección de la infraestructura, muchos de los conceptos y métodos son válidos en los modelos en la nube y las instalaciones. El cumplimiento de la protección de límites, el monitoreo de los puntos de entrada y salida y la implementación exhaustiva de registros, monitoreo y alertas es esencial para un efectivo plan de seguridad de la información.

Los clientes de AWS pueden personalizar, o reforzar, la configuración de una Amazon Elastic Compute Cloud (Amazon EC2), un contenedor de Amazon Elastic Container Service (Amazon ECS) o una instancia de AWS Elastic Beanstalk, y esta configuración puede perdurar en una imagen de Amazon Machine (AMI). Luego, ya sea que se activen por Auto Scaling o se lancen manualmente, todos los servidores (instancias) virtuales nuevos lanzados con esta AMI reciben la configuración reforzada.

Protección de los datos

Antes de diseñar la arquitectura de cualquier sistema, se deben establecer prácticas fundamentales que incidan en la seguridad. Por ejemplo, la clasificación de los datos permite categorizarlos en función del nivel de confidencialidad, así como el cifrado protege los datos al impedir el acceso no autorizado. Estas herramientas y técnicas son importantes porque respaldan objetivos como la prevención de la pérdida económica o la conformidad con las obligaciones normativas.

En AWS, las siguientes prácticas facilitan la protección de los datos:

- Como cliente de AWS, mantiene el control total sobre los datos.
- AWS facilita el cifrado de los datos y la administración de claves, incluida la rotación regular de estas, tarea que AWS puede automatizar fácilmente o que usted puede realizar.
- Está disponible el registro detallado con contenido importante, como el acceso a archivos y cambios.

- AWS ha diseñado sistemas de almacenamiento que ofrecen una resistencia excepcional. Por ejemplo, Amazon S3 Estándar, Estándar - Acceso poco frecuente de S3, Única zona - Acceso poco frecuente de S3 y Amazon Glacier están diseñados para proporcionar una durabilidad del 99,999999999 % de los objetos durante un año determinado. Este nivel de durabilidad corresponde a una pérdida esperada anual promedio de 0,000000001 % de los objetos.
- El control de versiones, que puede ser parte de un proceso más amplio de gestión del ciclo de vida de datos, puede proteger los datos contra sobreescrituras accidentales, eliminaciones y daños similares.
- AWS nunca inicia el movimiento de datos entre Regiones. El contenido colocado en una región permanecerá en esa región a menos que habilite explícitamente una función o aproveche un servicio que ofrezca esa funcionalidad.

Las siguientes preguntas se enfocan en estas consideraciones para la seguridad.

SEGURIDAD 7 ¿Cómo se clasifican los datos?
La clasificación de datos proporciona una forma de categorizar los datos en función de la criticidad y la confidencialidad, a fin de determinar los controles de protección y retención adecuados.

SEGURIDAD 8 ¿Cómo se protegen los datos en reposo?
Proteja sus datos en reposo mediante la implementación de varios controles a fin de reducir el riesgo de acceso no autorizado o de manipulación indebida.

SEGURIDAD 9 ¿Cómo se protegen los datos en tránsito?
Proteja sus datos en tránsito mediante la implementación de varios controles a fin de reducir el riesgo de acceso no autorizado o pérdida.

AWS proporciona múltiples medios para cifrar datos en reposo y en tránsito. Desarrollamos funciones en nuestros servicios que facilitan el cifrado de sus datos. Por ejemplo, hemos implementado el cifrado del lado del servidor (SSE) para Amazon S3 con el fin de facilitar el almacenamiento de los datos en forma cifrada. También puede hacer que todo el proceso de cifrado y descifrado HTTPS (generalmente conocido como terminación SSL) esté a cargo de Elastic Load Balancing (ELB).

Respuesta ante incidentes

Incluso con controles preventivos y de detección extremadamente sólidos, la organización debe implementar procesos para responder al potencial impacto de los incidentes de seguridad y mitigarlos. El diseño de la arquitectura de su carga de trabajo incide considerablemente en la capacidad de los equipos para operar con eficacia durante un incidente, para aislar o contener los sistemas y para restablecer las operaciones a un buen estado conocido. Establecer las herramientas y el acceso antes de un incidente de seguridad y luego practicar rutinariamente la respuesta ante incidentes durante los días de prueba ayudará a garantizar que la arquitectura pueda dar lugar a una recuperación e investigación oportunas.

En AWS, las siguientes prácticas facilitan la respuesta efectiva ante incidentes:

- Está disponible el registro detallado con contenido importante, como el acceso a archivos y cambios.
- Los eventos se pueden procesar automáticamente y activar herramientas que automaticen las respuestas mediante el uso de las API de AWS.
- Puede aprovisionar anticipadamente herramientas y una “sala limpia” mediante AWS CloudFormation. Esto permite realizar análisis forenses en un entorno seguro y aislado.

La siguiente pregunta se enfoca en estas consideraciones para la seguridad.

SEGURIDAD 10 ¿Cómo se anticipa, responde y recupera ante los incidentes?
La preparación es esencial para la investigación, respuesta y recuperación oportuna y efectiva de incidentes de seguridad a fin de ayudar a minimizar la interrupción en su organización.

Asegúrese de tener una manera de otorgar acceso rápidamente a su equipo de seguridad y automatice el aislamiento de instancias, así como la captura de los datos y estados para el análisis forense.

Recursos

Consulte los siguientes recursos para obtener más información sobre nuestras prácticas recomendadas para la seguridad.

Documentación

- Seguridad en la nube de AWS
- Conformidad en AWS
- Blog de seguridad de AWS

Documento técnico

- Pilar de seguridad
- Información general sobre la seguridad en AWS
- Prácticas recomendadas de seguridad en AWS
- Riesgo y conformidad en AWS

Video

- Estado de madurez de la seguridad en AWS
- Información general sobre la responsabilidad compartida

Fiabilidad

El pilar de la fiabilidad incluye la capacidad de una carga de trabajo para llevar a cabo la función prevista de forma correcta y consistente en el momento esperado. Esto incluye la capacidad de operar y probar la carga de trabajo a través de su ciclo de vida completo. Este documento ofrece orientación exhaustiva sobre las prácticas recomendadas para implementar cargas de trabajo fiables en AWS.

El pilar de fiabilidad ofrece una descripción general de los principios de diseño, las prácticas recomendadas y las preguntas. Puede encontrar orientación normativa acerca de la implementación en el Documento técnico sobre el pilar de fiabilidad.

Temas

- Principios de diseño (p. 21)
- Definición (p. 21)
- Prácticas recomendadas (p. 21)
- Recursos (p. 25)

Principios de diseño

Existen cinco principios de diseño para la fiabilidad en la nube:

- Recuperarse de los errores automáticamente: si monitorea una carga de trabajo para obtener los indicadores clave de rendimiento (KPI), puede activar el proceso de automatización cuando se supera un límite. Estos KPI deben ser una medida del valor comercial, no de los aspectos técnicos de la operación del servicio. Esto permite la notificación automática, el seguimiento de los errores y los procesos de recuperación automatizados que solucionan o reparan el error. Con una automatización más sofisticada, es posible anticipar y corregir los errores antes de que ocurran.
- Probar los procedimientos de recuperación: en un entorno en las instalaciones, a menudo se realizan pruebas para demostrar que la carga de trabajo funciona en una situación particular. Por lo general, las pruebas no se utilizan para validar las estrategias de recuperación. En la nube, puede realizar pruebas para detectar de qué forma se producen errores en su carga de trabajo y puede validar los procedimientos de recuperación. Puede utilizar la automatización para simular diferentes errores o para recrear las situaciones que causaron errores anteriormente. Este enfoque expone las rutas de los errores que puede probar y corregir antes de que ocurra una situación de error real, de manera que se reduce el riesgo.
- Escalar horizontalmente para aumentar la disponibilidad de la carga de trabajo agregada: reemplace un recurso grande por varios recursos pequeños para reducir el impacto de un solo error en toda la carga de trabajo. Distribuya las solicitudes en varios recursos más pequeños para asegurarse de que no compartan un punto común de error.
- Dejar de suponer la capacidad: una causa común de los errores en las cargas de trabajo en las instalaciones es la saturación de recursos cuando las demandas que se le asignan a una carga de trabajo exceden su capacidad (este suele ser el objetivo de los ataques de denegación de servicio). En la nube, puede monitorear la demanda y la utilización de la carga de trabajo. Además, puede automatizar el proceso de incorporación o eliminación de recursos a fin de mantener el nivel óptimo para satisfacer la demanda sin llegar a un aprovisionamiento excesivo o insuficiente. Aún existen límites, pero algunas cuotas se pueden controlar y otras se pueden administrar (consulte la sección Administrar Service Quotas y restricciones de servicio).
- Administrar los cambios en la automatización: los cambios en la infraestructura se deben realizar mediante la automatización. Entre los cambios que deben administrarse se incluyen los cambios en la automatización, que luego se pueden seguir y revisar.

Definición

Existen cuatro áreas de prácticas recomendadas para la fiabilidad en la nube:

- Bases
- Arquitectura de las cargas de trabajo
- Administración de los cambios
- Administración de los errores

Para lograr la fiabilidad, debe comenzar por las bases: un entorno donde todas las cuotas de servicio y la topología de red se adapten a la carga de trabajo. La arquitectura de la carga de trabajo del sistema distribuido debe estar diseñada para prevenir y reducir los errores. La carga de trabajo debe controlar los cambios en la demanda o los requisitos. Además, debe estar diseñada para detectar los errores y recuperarse de forma automática.

Prácticas recomendadas

Temas

- Bases (p. 22)
- Arquitectura de las cargas de trabajo (p. 22)
- Administración de los cambios (p. 23)
- Administración de los errores (p. 24)

Bases

Los requisitos básicos son aquellos cuyo alcance se extiende más allá de una sola carga de trabajo o proyecto. Antes de diseñar cualquier sistema, deben establecerse los requisitos básicos que influyen en la fiabilidad. Por ejemplo, debe tener suficiente ancho de banda de red para su centro de datos.

Con AWS la mayoría de estos requisitos básicos ya están incorporados o se los puede satisfacer según sea necesario. El diseño de la nube hace que esta sea casi ilimitada, de manera que es responsabilidad de AWS satisfacer la necesidad de una capacidad de cómputo y de conexión de red suficiente, lo que le permite cambiar el tamaño del recurso y las asignaciones bajo demanda.

Las siguientes preguntas se enfocan en estas consideraciones para la fiabilidad. (Para ver una lista con las preguntas y las prácticas recomendadas sobre la fiabilidad, consulte el Apéndice (p. 61)).

FIABILIDAD 1 ¿Cómo se administran las cuotas y las restricciones de servicio?
Para las arquitecturas de cargas de trabajo basadas en la nube, existen las cuotas de servicio (que también se denominan límites de servicio). Estas cuotas existen para evitar el aprovisionamiento accidental de más recursos de los que necesita y para limitar la tasa de solicitudes en las operaciones de la API a fin de proteger los servicios de un uso inadecuado. Además, existen restricciones de recursos, por ejemplo, la tasa con la que puede enviar bits por un cable de fibra óptica o la cantidad de almacenamiento en un disco físico.

FIABILIDAD 2 ¿Cómo se planifica la topología de red?
A menudo, las cargas de trabajo se encuentran en varios entornos. Entre ellos se incluyen varios entornos en la nube (de acceso público y privado) y, posiblemente, su infraestructura de centros de datos existente. Los planes deben incluir las consideraciones sobre la red, como la conectividad dentro del sistema y entre sistemas, la administración de direcciones IP públicas y privadas y la resolución de nombres de dominio.

Para las arquitecturas de cargas de trabajo basadas en la nube, existen las cuotas de servicio (que también se denominan límites de servicio). Estas cuotas existen para evitar el aprovisionamiento accidental de más recursos de los que necesita y para limitar la tasa de solicitudes en las operaciones de la API a fin de proteger los servicios de un uso inadecuado. A menudo, las cargas de trabajo se encuentran en varios entornos. Usted debe supervisar y administrar estas cuotas para todos los entornos de carga de trabajo. Entre ellos se incluyen varios entornos en la nube (de acceso público y privado) y pueden incluir la infraestructura de su centro de datos existente. Los planes deben incluir las consideraciones sobre la red, como la conectividad dentro del sistema y entre sistemas, la administración de direcciones IP públicas, la administración de direcciones IP privadas y la resolución de nombres de dominio.

Arquitectura de las cargas de trabajo

Una carga de trabajo confiable comienza con la toma de decisiones de diseño inicial para el software y la infraestructura. Sus opciones de arquitectura afectarán el comportamiento de la carga de trabajo en los cinco pilares de Well-Architected. En cuanto a la fiabilidad, debe seguir determinados patrones.

Con AWS, los desarrolladores de cargas de trabajo pueden elegir los lenguajes y las tecnologías que usarán. Los SDK de AWS eliminan la complejidad de la codificación al proporcionar API específicas del lenguaje para los servicios de AWS. Estos SDK, más la elección de lenguajes, permiten a los desarrolladores implementar las prácticas recomendadas de fiabilidad aquí presentadas. Los desarrolladores también pueden leer y aprender sobre cómo Amazon crea y opera software en Amazon Builders' Library.

Las siguientes preguntas se enfocan en estas consideraciones para la fiabilidad.

FIABILIDAD 3 ¿Cómo se diseña la arquitectura de servicios para la carga de trabajo?
Cree cargas de trabajo sumamente escalables y confiables a través de una arquitectura orientada a servicios (SOA) o una arquitectura de microservicios. La arquitectura orientada a servicios (SOA) es la práctica de crear componentes de software reutilizables a través de las interfaces de servicios. La arquitectura de microservicios ha avanzado en la creación de los componentes proporcionando componentes más pequeños y simples.

FIABILIDAD 4: ¿Cómo se diseñan interacciones en un sistema distribuido para evitar errores?
Los sistemas distribuidos dependen de las redes de comunicación para interconectar los componentes, como servidores o servicios. A pesar de la pérdida de datos o la latencia en estas redes, su carga de trabajo debe operar de manera confiable. Los componentes del sistema distribuido deben funcionar de manera que no afecten negativamente a otros componentes o a la carga de trabajo. Las prácticas recomendadas evitan errores y mejoran el tiempo promedio entre los errores (MTBF).

FIABILIDAD 5 ¿Cómo se diseñan interacciones en un sistema distribuido para mitigar o tolerar errores?
Los sistemas distribuidos dependen de las redes de comunicación para interconectar los componentes (como servidores o servicios). A pesar de la pérdida de datos o la latencia sobre estas redes, su carga de trabajo debe funcionar de manera confiable. Los componentes del sistema distribuido deben funcionar de manera que no afecten negativamente a otros componentes o a la carga de trabajo. Las prácticas recomendadas permiten que las cargas de trabajo toleren errores o presiones, se recuperen más rápido de estos y mitiguen el impacto de dichas dificultades. El resultado es un mejor tiempo promedio de recuperación (MTTR).

Los sistemas distribuidos dependen de las redes de comunicación para interconectar los componentes, como servidores o servicios. A pesar de la pérdida de datos o la latencia en estas redes, su carga de trabajo debe operar de manera confiable. Los componentes del sistema distribuido deben funcionar de manera que no afecten negativamente a otros componentes o a la carga de trabajo.

Administración de los cambios

Los cambios en su carga de trabajo o su entorno se deben anticipar y adaptar a fin de lograr un funcionamiento confiable de la carga de trabajo. Se incluyen los cambios impuestos en su carga de trabajo, como los picos en la demanda, y también los internos, como las implementaciones de características y los parches de seguridad.

Con AWS, puede monitorear el comportamiento de una carga de trabajo y automatizar la respuesta a los KPI. Por ejemplo, la carga de trabajo puede agregar servidores a medida que esta obtiene más usuarios. Puede controlar quién tiene permiso para realizar cambios en la carga de trabajo y auditar el historial de estos cambios.

Las siguientes preguntas se enfocan en estas consideraciones para la fiabilidad.

FIABILIDAD 6 ¿Cómo se monitorean los recursos de las cargas de trabajo?
Los registros y las métricas son herramientas poderosas para obtener información sobre el estado de su carga de trabajo. Puede configurar su carga de trabajo para monitorear los registros y las métricas y enviar notificaciones cuando se superen los límites o se produzcan eventos significativos. El monitoreo permite que su carga de trabajo reconozca cuándo se superan los límites de bajo rendimiento o cuándo se producen errores, de manera que se pueda recuperar automáticamente como respuesta.

FIABILIDAD 7 ¿Cómo se diseña la carga de trabajo para que se adapte a los cambios en la demanda?
Una carga de trabajo escalable proporciona elasticidad para agregar o eliminar recursos de forma automática, de manera que coincidan estrechamente con la demanda actual en cualquier momento específico.

FIABILIDAD 8 ¿Cómo se implementan los cambios?
Los cambios controlados son necesarios para implementar nuevas funcionalidades y para asegurarse de que el entorno operativo, así como también las cargas de trabajo, ejecutan un software conocido, que se puede reemplazar de una manera predecible o que contiene los parches adecuados. Si no se controlan estos cambios, es más difícil predecir los efectos de estos cambios o abordar los problemas que surjan como consecuencia de ellos.

Cuando diseña una carga de trabajo para agregar y eliminar recursos automáticamente en respuesta a los cambios en la demanda, esto no solo aumenta la fiabilidad sino que también asegura que el éxito empresarial no se convierta en una carga. Con el monitoreo, su equipo recibirá una alerta automática cuando los KPI se desvíen de las normas previstas. El registro automático de cambios en su entorno le permite auditar e identificar rápidamente acciones que podrían haber afectado la fiabilidad. Los controles sobre la gestión de cambios aseguran que pueda aplicar las reglas que le brindan la fiabilidad que necesita.

Administración de los errores

En cualquier sistema de complejidad razonable se espera que se produzcan errores. La confiabilidad requiere que la carga de trabajo sea consciente de los errores a medida que ocurren y tome medidas para evitar el impacto en la disponibilidad. Las cargas de trabajo deben ser capaces de resistir errores y reparar problemas automáticamente.

Con AWS, puede aprovechar la automatización para reaccionar a los datos de monitoreo. Por ejemplo, cuando una métrica específica cruza un umbral, puede activar una acción automatizada para remediar el problema. Además, en lugar de tratar de diagnosticar y corregir un recurso con errores que es parte del entorno de producción, puede reemplazarlo por uno nuevo y realizar el análisis del recurso con errores fuera de banda. Dado que la nube le permite instalar versiones temporales de un sistema entero a bajo costo, puede usar pruebas automatizadas para verificar procesos de recuperación completa.

Las siguientes preguntas se enfocan en estas consideraciones para la fiabilidad.

FIABILIDAD 9 ¿Cómo se realizan copias de seguridad de los datos?
Realice copias de seguridad de los datos, las aplicaciones y las configuraciones a fin de cumplir con los requisitos de los objetivos de tiempo de recuperación (RTO) y los objetivos de puntos de recuperación (RPO).

FIABILIDAD 10 ¿Cómo se utiliza el aislamiento de errores para proteger la carga de trabajo?
Los límites del aislamiento de errores restringen los efectos de un error dentro de la carga de trabajo a una cantidad limitada de componentes. Los componentes que se encuentren por fuera de los límites no se ven afectados por el error. La implementación de varios límites de aislamiento de errores le permite restringir el impacto de los errores en su carga de trabajo.

FIABILIDAD 11: ¿Cómo se diseña la carga de trabajo para tolerar errores de componentes?
Las cargas de trabajo que presenten requisitos de alta disponibilidad y tiempo medio de recuperación (MTTR) bajo se deben diseñar de forma que sean resistentes.

FIABILIDAD 12 ¿Cómo se prueba la fiabilidad?
Después de haber diseñado su carga de trabajo para que sea resistente a las presiones de la producción, las pruebas son la única forma de garantizar que funcionará como se diseñó y proporcionará la resistencia que espera.

FIABILIDAD 13 ¿Cómo se planifica la recuperación ante desastres (DR)?
Tener copias de seguridad y componentes de carga de trabajo redundantes en las instalaciones es el primer paso de su estrategia de recuperación de desastres (DR). Los objetivos de tiempo y punto de recuperación son los objetivos que debe cumplir para lograr la restauración de la disponibilidad. Debe establecer estos objetivos en función de las necesidades de la empresa. Implemente una estrategia para cumplir estos objetivos, teniendo en cuenta la ubicación y la función de los recursos y los datos de la carga de trabajo.

Realice regularmente una copia de seguridad de los datos y pruebe los archivos de copia de seguridad para garantizar que puede recuperarse de errores lógicos y físicos. Una clave para administrar los errores es la prueba frecuente y automatizada de las cargas de trabajo para ocasionar errores y luego observar cómo se recuperan. Haga esto regularmente y asegúrese de que tales pruebas también se activen después de cambios significativos en la carga de trabajo. Realice un seguimiento activo de los KPI, como el objetivo de tiempo de recuperación (RTO) y el objetivo de punto de recuperación (RPO), a fin de evaluar la resistencia de una carga de trabajo (especialmente en escenarios de prueba de errores). El seguimiento de los KPI lo ayudará a identificar y mitigar puntos únicos de errores. El objetivo es probar a fondo los procesos de recuperación de la carga de trabajo para que esté seguro de que puede recuperar todos sus datos y continuar prestando servicios a sus clientes, incluso en situaciones de problemas recurrentes. Sus procesos de recuperación deben ejercerse tan bien como sus procesos de producción habituales.

Recursos

Consulte los siguientes recursos para obtener más información sobre las prácticas recomendadas para la fiabilidad.

Documentación

- Documentación de AWS
- Infraestructura global de AWS
- AWS Auto Scaling: cómo funcionan los planes de escalado

- ¿En qué consiste AWS Backup?

Documento técnico

- Pilar de fiabilidad: AWS Well-Architected
- Implementación de microservicios en AWS

Eficiencia de rendimiento

El pilar de eficiencia de rendimiento incluye la habilidad de utilizar recursos informáticos de manera eficiente para cumplir con los requisitos del sistema y mantener esa eficiencia a medida que la demanda cambia y la tecnología evoluciona.

El pilar de eficiencia de rendimiento ofrece una descripción general de los principios de diseño, las prácticas recomendadas y las preguntas. Puede encontrar orientación normativa acerca de la implementación en el documento técnico sobre el pilar de la eficiencia de rendimiento.

Temas

- Principios de diseño (p. 26)
- Definición (p. 27)
- Prácticas recomendadas (p. 27)
- Recursos (p. 32)

Principios de diseño

Existen cinco principios de diseño para la eficiencia de rendimiento en la nube:

- Democratizar las tecnologías avanzadas: facilite la implementación de tecnología avanzada para su equipo mediante la delegación de tareas complejas al proveedor de nube. En lugar de pedirle a su equipo de TI que aprenda sobre el alojamiento y la ejecución de una nueva tecnología, considere consumir la tecnología como un servicio. Por ejemplo, las bases de datos NoSQL, la transcodificación de medios y el aprendizaje automático son tecnologías que requieren conocimientos especializados. En la nube, estas tecnologías se convierten en servicios que su equipo puede consumir, lo que les permite centrarse en el desarrollo del producto en lugar del aprovisionamiento y administración de recursos.
- Incorporarse al mercado global en minutos: la implementación de la carga de trabajo en varias regiones de AWS en todo el mundo permite ofrecer baja latencia y una mejor experiencia para sus clientes a un costo mínimo.
- Utilizar arquitecturas sin servidor: las arquitecturas sin servidor eliminan la necesidad de ejecutar y mantener servidores físicos para actividades informáticas tradicionales. Por ejemplo, los servicios de almacenamiento sin servidor pueden actuar como sitios web estáticos (eliminan la necesidad de servidores web) y los servicios para eventos pueden alojar un código. Esto elimina la carga operativa de administrar servidores físicos y puede reducir los costos transaccionales porque los servicios administrados operan a escala de la nube.
- Experimentar con más frecuencia: con los recursos automatizables y virtuales, puede llevar a cabo con rapidez pruebas comparativas con diferentes tipos de instancias, almacenamiento o configuraciones.
- Considerar la afinidad mecánica: comprenda de qué manera se consumen los servicios en la nube y siempre utilice el enfoque tecnológico que se adapte mejor a los objetivos de la carga de trabajo. Por ejemplo, tenga en cuenta los patrones de acceso de datos cuando selecciona las bases de datos o los enfoques de almacenamiento.

Definición

Existen cuatro áreas de prácticas recomendadas para la eficiencia de rendimiento en la nube:

- Selección
- Revisión
- Monitoreo
- Compensaciones

Adopte un enfoque basado en datos para crear una arquitectura de alto rendimiento. Recopile datos sobre todos los aspectos de la arquitectura, desde el diseño de alto nivel hasta la selección y configuración de tipos de recursos.

La revisión de sus opciones de forma regular garantiza que se aproveche de la continua evolución de la nube de AWS. El monitoreo garantiza que esté al tanto de cualquier desviación del rendimiento esperado. Realice compensaciones en la arquitectura para mejorar el rendimiento, como el uso de compresión o almacenamiento en caché o la flexibilización de los requisitos de consistencia.

Prácticas recomendadas

Temas

- Selección (p. 27)
- Informática (p. 28)
- Almacenamiento (p. 28)
- Base de datos (p. 29)
- Red (p. 30)
- Revisión (p. 30)
- Monitoreo (p. 31)
- Compensaciones (p. 31)

Selección

La solución óptima para una carga de trabajo particular varía y las soluciones suelen combinar múltiples enfoques. Las cargas de trabajo de buena arquitectura utilizan múltiples soluciones y permiten diferentes características para mejorar el rendimiento.

Los recursos de AWS están disponibles en muchos tipos y configuraciones, lo que facilita encontrar un enfoque que se ajuste a las necesidades de la carga de trabajo. También puede encontrar opciones que no son fáciles de lograr con la infraestructura en las instalaciones. Por ejemplo, un servicio administrado como Amazon DynamoDB ofrece una base de datos NoSQL completamente administrada con latencia en milisegundos de un solo dígito en cualquier escala.

La siguiente pregunta se enfoca en estas consideraciones para la eficiencia de rendimiento. (Para ver la lista de las preguntas y las prácticas recomendadas de eficiencia de rendimiento, consulte el Apéndice (p. 69)).

RENDIMIENTO 1 ¿Cómo se selecciona la arquitectura con el mejor rendimiento?
A menudo, se requieren múltiples enfoques para obtener un rendimiento óptimo en una carga de trabajo. Los sistemas de buena arquitectura utilizan múltiples soluciones y permiten diferentes características para mejorar el rendimiento.

Utilice un enfoque basado en datos para seleccionar los patrones y la implementación para su arquitectura y logre una solución rentable. Los arquitectos de soluciones de AWS, los socios de las arquitecturas de referencia de AWS y la red de socios de AWS (APN) pueden ayudarlo a seleccionar una arquitectura en función del conocimiento del sector, pero los datos obtenidos a través de pruebas de puntos de referencia o de carga serán necesarios para optimizar su arquitectura.

Es probable que su arquitectura combine varios enfoques de arquitectura diferentes (por ejemplo, impulsados por eventos, ETL o canalización). La implementación de la arquitectura utilizará los servicios de AWS que son específicos para la optimización del rendimiento de la arquitectura. En las siguientes secciones discutiremos los cuatro tipos de recurso principales a considerar (informática, almacenamiento, base de datos y red).

Informática

La selección de recursos informáticos que cumplan con sus requisitos, necesidades de rendimiento y ofrezcan gran eficiencia de costo y esfuerzo le permitirán lograr más con la misma cantidad de recursos. Cuando evalúe las opciones informáticas, tenga en cuenta los requisitos para el rendimiento de la carga de trabajo y el costo y utilícelos para tomar decisiones fundamentadas.

En AWS, la informática está disponible de tres formas: instancias, contenedores y funciones:

- Instancias son servidores virtualizados, que permiten cambiar sus capacidades con un botón o una llamada a la API. Como las decisiones de recursos en la nube no son fijas, puede experimentar con diferentes tipos de servidores. En AWS, estas instancias de servidores virtuales vienen en diferentes familias y tamaños y ofrecen una amplia variedad de capacidades, incluidas unidades de estado sólido (SSD) y unidades de procesamiento de gráficos (GPU).
- Contenedores son un método de virtualización de sistema operativo que permiten ejecutar una aplicación y sus dependencias en procesos aislados de los recursos. AWS Fargate es informática sin servidor para contenedores. También, puede utilizar Amazon EC2 si necesita tener el control sobre la instalación, la configuración y la administración del entorno informático. También puede elegir entre plataformas organizadoras de contenedores múltiples: Amazon Elastic Container Service (ECS) o Amazon Elastic Kubernetes Service (EKS).
- Funciones abstraen el entorno de ejecución desde el código que desea ejecutar. Por ejemplo, AWS Lambda permite ejecutar el código sin ejecutar una instancia.

La siguiente pregunta se enfoca en estas consideraciones para la eficiencia de rendimiento.

RENDIMIENTO 2 ¿Cómo se selecciona una solución de informática?
La solución de informática óptima para una carga de trabajo específica puede variar en función del diseño de la aplicación, los patrones de uso y los ajustes de configuración. Las arquitecturas pueden utilizar diferentes soluciones de informática para varios componentes y habilitar distintas características para mejorar el rendimiento. Si se elige la solución de informática incorrecta para una arquitectura, esto puede reducir la eficiencia del rendimiento.

Cuando diseñe su uso de la informática, aproveche los mecanismos de elasticidad disponibles para garantizar que posee suficiente capacidad para sostener el rendimiento a medida que cambia la demanda.

Almacenamiento

El almacenamiento en la nube es un componente esencial de la informática en la nube, contiene la información utilizada por su carga de trabajo. El almacenamiento en la nube es normalmente más fiable, escalable y seguro que los sistemas tradicionales de almacenamiento en las instalaciones. Seleccione entre los servicios de almacenamiento de archivos, bloques y objetos, así como las opciones de migración de datos en la nube para la carga de trabajo.

En AWS, el almacenamiento está disponible de tres formas: objeto, bloque y archivo:

- El almacenamiento de objetos ofrece una plataforma duradera y escalable para que los datos sean accesibles desde cualquier ubicación de Internet para el contenido generado por el usuario, el archivo activo, la informática sin servidor, el almacenamiento de Big Data o copia de seguridad y recuperación. Amazon Simple Storage Service (Amazon S3) es un servicio de almacenamiento de objetos que ofrece escalabilidad, disponibilidad de los datos, seguridad y rendimiento líderes en el sector. Amazon S3 está diseñado para el 99,999999999 % (11 nueves) de durabilidad, y almacena datos para millones de aplicaciones para compañías en todo el mundo.
- El almacenamiento de bloques proporciona un almacenamiento de bloques altamente disponible, consistente y de baja latencia para cada alojamiento virtual y es análogo al almacenamiento de conexión directa (DAS) o a la red de área de almacenamiento (SAN). Amazon Elastic Block Store (Amazon EBS) está diseñado para cargas de trabajo que demandan almacenamiento persistente accesible por instancias de EC2 que ayuda a ajustar las aplicaciones con el rendimiento, el costo y la capacidad de almacenamiento adecuados.
- El almacenamiento de archivos ofrece acceso a un sistema de archivos compartidos en múltiples sistemas. Las soluciones de almacenamiento de archivos como Amazon Elastic File System (EFS) son ideales para casos de uso, como repositorios de contenido grandes, entornos de desarrollo, almacenamientos de contenido multimedia o directorios de inicio del usuario. Amazon FSx hace que sea fácil y rentable lanzar y ejecutar sistemas de archivos conocidos, para así aprovechar los conjuntos con abundantes características y el rápido rendimiento de los sistemas de archivos de código abierto y con licencias comerciales usados ampliamente.

La siguiente pregunta se enfoca en estas consideraciones para la eficiencia de rendimiento.

RENDIMIENTO 3 ¿Cómo se selecciona una solución de almacenamiento?
La solución de almacenamiento óptimo para un sistema varía según el tipo de método de acceso (bloque, archivo u objeto), patrones de acceso (aleatorio o secuencial), rendimiento requerido, frecuencia de acceso (en línea, sin conexión, de archivo) frecuencia de actualización (WORM, dinámico) y restricciones de durabilidad y disponibilidad. Los sistemas de buena arquitectura utilizan múltiples soluciones de almacenamiento y permiten que diferentes características mejoren el rendimiento y usen los recursos de manera eficiente.

Cuando selecciona una solución de almacenamiento, garantizar que coincida con sus patrones de acceso será esencial para lograr el rendimiento que desea.

Base de datos

La nube ofrece servicios de base de datos creadas con fines específicos que abordan diferentes problemas que presenta su carga de trabajo. Puede elegir entre muchos motores de bases de datos creadas específicamente, incluidas las bases de datos relacionales, valor clave, documento, en la memoria, gráfico, serie temporal y libro mayor. Cuando elige la mejor base de datos para resolver un problema específico (o un grupo de problemas), puede separarse de las bases de datos monolíticas, universales y restrictivas y centrarse en la creación de aplicaciones que satisfagan las necesidades de rendimiento de los clientes.

En AWS, puede elegir entre múltiples motores de bases de datos creadas específicamente, incluidas las bases de datos relacionales, valor clave, documento, en la memoria, gráfico, serie temporal y libro mayor. Con las bases de datos de AWS, no necesita preocuparse por las tareas de administración de la base de datos, como el aprovisionamiento del servidor, la aplicación de parches, los ajustes, la configuración, las copias de seguridad o la recuperación. AWS monitorea de forma continua los clústeres para mantener la carga de trabajo activa y en funcionamiento con almacenamiento de autorrecuperación y escalado automatizado. De esta forma, usted puede centrarse en el desarrollo de aplicaciones de mayor valor.

La siguiente pregunta se enfoca en estas consideraciones para la eficiencia de rendimiento.

RENDIMIENTO 4: ¿Cómo se selecciona una solución de base de datos?
La solución de base de datos óptima para un sistema varía según los requerimientos de disponibilidad, consistencia, tolerancia en las particiones, latencia, durabilidad, escalabilidad y capacidad de consulta. Muchos sistemas utilizan soluciones de bases de datos diferentes para varios subsistemas y permiten que distintas características mejoren el rendimiento. La selección de las características y soluciones de base de datos incorrectas puede resultar en una menor eficiencia de rendimiento.

El enfoque de la base de datos de la carga de trabajo tiene un impacto significativo en la eficiencia del rendimiento. Con frecuencia, se trata de un área que se elige según las predeterminaciones organizativas y no mediante un enfoque basado en los datos. Como con el almacenamiento, es esencial considerar los patrones de acceso de la carga de trabajo y, además, debe tener en cuenta si otras soluciones que no sean de base de datos pueden resolver el problema de manera más eficiente (como el uso de gráficos, series temporales o bases de datos de almacenamiento en la memoria).

Red

Como la red se encuentra entre todos los componentes de la carga de trabajo, puede tener grandes impactos positivos o negativos en el rendimiento y comportamiento de la carga de trabajo. También existen cargas de trabajo que dependen fuertemente del rendimiento de la red, como la informática de alto rendimiento (HPC) donde la comprensión profunda de la red es importante para aumentar el rendimiento del clúster. Debe determinar los requisitos de la carga de trabajo para el ancho de banda, la latencia, la fluctuación y el rendimiento.

En AWS, la red se virtualiza y está disponible en varios tipos y configuraciones diferentes. Esto facilita la coincidencia entre los métodos de red con las necesidades. AWS ofrece características de productos (por ejemplo, redes mejoradas, instancias optimizadas de Amazon EBS, Amazon S3 Transfer Acceleration y Amazon CloudFront dinámico) para optimizar el tráfico de red. AWS también ofrece características de red (por ejemplo, direccionamiento de latencia de Amazon Route 53, puntos de enlace de Amazon VPC, AWS Direct Connect y AWS Global Accelerator) para reducir la distancia o fluctuación de la red.

La siguiente pregunta se enfoca en estas consideraciones para la eficiencia de rendimiento.

RENDIMIENTO 5: ¿Cómo se configura la solución de red?
La solución de red óptima para una carga de trabajo varía según la latencia, los requisitos de rendimiento, la fluctuación y el ancho de banda. Las restricciones físicas, como el usuario o los recursos en las instalaciones, determinan las opciones de ubicación. Estas restricciones se pueden compensar con ubicaciones de borde o ubicación de recurso.

Se debe considerar la ubicación al lanzar la red. Puede elegir colocar los recursos cerca de donde se utilizarán para reducir la distancia. Utilice las métricas de la red para hacer cambios en la configuración de la red a medida que evoluciona la carga de trabajo. Si aprovecha las regiones, los grupos de ubicación y los servicios de borde, puede mejorar el rendimiento significativamente. Las redes basadas en la nube se pueden recrear o modificar rápidamente, por lo tanto es necesario que la arquitectura en la red evolucione con el tiempo para mantener la eficiencia del rendimiento.

Revisión

Las tecnologías en la nube evolucionan rápidamente y usted debe garantizar que los componentes de las cargas de trabajo utilicen los enfoques y las tecnologías más recientes para mejorar el rendimiento de manera continua. Debe evaluar de manera continua y tener en cuenta los cambios para los componentes de la carga de trabajo a fin de garantizar que cumple con los objetivos de rendimiento y costo. Las

nuevas tecnologías, como el aprendizaje automático y la inteligencia artificial (AI), permiten replantear las experiencias de los clientes e innovar en todas las cargas de trabajo empresariales.

Aproveche las innovaciones continuas en AWS impulsadas por la necesidad del cliente. Lanzamos nuevas regiones, ubicaciones de borde, servicios y características regularmente. Cualquiera de estos lanzamientos puede mejorar de manera positiva la eficiencia del rendimiento de la arquitectura.

La siguiente pregunta se enfoca en estas consideraciones para la eficiencia de rendimiento.

RENDIMIENTO 6: ¿Cómo se desarrolla la carga de trabajo para aprovechar los nuevos lanzamientos?
Cuando diseña las cargas de trabajo, hay una cantidad limitada de opciones entre las que puede elegir. Sin embargo, con el tiempo, las nuevas tecnologías y enfoques estarán disponibles para que pueda mejorar el rendimiento de la carga de trabajo.

Las arquitecturas con rendimiento deficiente son generalmente el resultado de un proceso de revisión del rendimiento inexistente o dañado. Si su arquitectura presenta un rendimiento deficiente, implementar un proceso de revisión de rendimiento le permitirá aplicar el ciclo de planificación, ejecución, verificación y reacción (PDCA) de Deming para impulsar una mejora iterativa.

Monitoreo

Después de implementar la carga de trabajo, debe monitorear su rendimiento, de esta manera puede solucionar cualquier problema antes de que afecte a los clientes. El monitoreo de las métricas se debe utilizar para activar alarmas cuando se alcanzan los límites.

Amazon CloudWatch es un servicio de monitoreo y observación que proporciona datos e información práctica para monitorear la carga de trabajo, responder a cambios de rendimiento en todo el sistema, optimizar la utilización de recursos y obtener una vista unificada del estado de las operaciones. CloudWatch recopila datos del monitoreo y las operaciones en forma de registros, métricas y eventos de cargas de trabajo que se ejecutan en AWS y en los servidores en las instalaciones. AWS X-Ray ayuda a los desarrolladores a analizar y depurar la producción y las aplicaciones distribuidas. Con AWS X-Ray, puede deducir información sobre el rendimiento de la aplicación y descubrir las causas raíz e identificar los cuellos de botella en el rendimiento. Puede utilizar esta información para reaccionar rápidamente y mantener la carga de trabajo en funcionamiento sin inconvenientes.

La siguiente pregunta se enfoca en estas consideraciones para la eficiencia de rendimiento.

RENDIMIENTO 7: ¿Cómo se monitorean los recursos para garantizar que el rendimiento es óptimo?
El rendimiento del sistema se puede degradar con el tiempo. Monitoree el rendimiento del sistema para identificar la degradación y solucionar los factores internos y externos, como el sistema operativo o la carga de la aplicación.

Para una solución de monitoreo efectiva es clave asegurarse de no ver falsos positivos. Los desencadenadores automatizados evitan el error humano y pueden reducir el tiempo que toma solucionar los problemas. Planifique los días de prueba, en los que se realizan simulaciones en el entorno de producción, para probar las soluciones de alarma y garantizar que reconozca los problemas de manera correcta.

Compensaciones

Al diseñar las soluciones de arquitectura, piense en las compensaciones para garantizar un enfoque óptimo. En función de su situación, puede intercambiar la consistencia, la durabilidad y el espacio por tiempo o latencia, para entregar un rendimiento mayor.

Con AWS, puede incorporarse al mercado global rápidamente e implementar recursos en múltiples ubicaciones en el mundo para acercarse a sus usuarios finales. También puede agregar de manera dinámica réplicas de solo lectura a los almacenes de información (como sistemas de bases de datos) para reducir la carga en la base de datos primaria.

La siguiente pregunta se enfoca en estas consideraciones para la eficiencia de rendimiento.

RENDIMIENTO 8: ¿Cómo se utilizan las compensaciones para mejorar el rendimiento?
Cuando diseñe soluciones, determinar las compensaciones le permite seleccionar un enfoque óptimo. A menudo, puede mejorar el rendimiento con el intercambio de la consistencia, la durabilidad y el espacio por tiempo y latencia.

A medida que implementa cambios en la carga de trabajo, recopile y evalúe las métricas para determinar el impacto de esos cambios. Mida el impacto en los sistemas y en el usuario final para comprender de qué manera las compensaciones repercuten en la carga de trabajo. Utilice un enfoque sistemático, como la prueba de carga, para explorar si las compensaciones mejoran el rendimiento.

Recursos

Consulte los siguientes recursos para obtener más información sobre nuestras prácticas recomendadas para la eficiencia del rendimiento.

Documentación

- Optimización de rendimiento de Amazon S3
- Rendimiento por volumen de Amazon EBS

Documento técnico

- Pilar de eficiencia de rendimiento

Video

- AWS re:Invent 2019: conceptos básicos de Amazon EC2 (CMP211-R2)
- AWS re:Invent 2019: sesión de liderazgo: estado de madurez del almacenamiento (STG201-L)
- AWS re:Invent 2019: sesión de liderazgo: bases de datos construidas para un fin específico de AWS (DAT209-L)
- AWS re:Invent 2019: conectividad a AWS y arquitecturas de red de AWS híbridas (NET317-R1)
- AWS re:Invent 2019: impulsando Amazon EC2 de próxima generación: análisis profundo del sistema Nitro (CMP303-R2)
- AWS re:Invent 2019: escalar hasta los primeros 10 millones de usuarios (ARC211-R)

Optimización de costos

El pilar de optimización de costos incluye la capacidad de ejecutar sistemas para entregar valor empresarial al menor precio.

El pilar de optimización de costos proporciona información general sobre los principios de diseño, las prácticas recomendadas y las preguntas. Puede encontrar orientación normativa acerca de la implementación en el Documento técnico sobre el pilar de optimización de costos.

Temas

- Principios de diseño (p. 33)
- Definición (p. 33)
- Prácticas recomendadas (p. 34)
- Recursos (p. 38)

Principios de diseño

Existen cinco principios de diseño para la optimización de costos en la nube:

- Implementar la administración financiera en la nube: para lograr el éxito financiero y acelerar la materialización del valor de negocio en la nube, debe invertir en la administración financiera en la nube o la optimización de costos. Su organización debe dedicar tiempo y recursos al desarrollo de capacidades en este nuevo ámbito de la tecnología y de la administración del uso. De manera similar a la capacidad de seguridad o de excelencia operativa, debe desarrollar capacidades a través de conocimientos, programas, recursos y procesos que lo ayuden a convertirse en una organización rentable.
- Adoptar un modelo de consumo: pague solo por los recursos informáticos que necesite y aumente o disminuya el nivel de uso en función de los requisitos empresariales y no mediante estimaciones elaboradas. Por ejemplo, los entornos de desarrollo y prueba suelen utilizarse solo ocho horas al día durante la semana laboral. Puede detener estos recursos cuando no se estén utilizando para obtener un posible ahorro de costos del 75 % (40 horas frente a 168 horas).
- Medir la eficiencia general: mida el resultado empresarial de la carga de trabajo y los costos asociados con la entrega. Utilice esta medición para conocer las ganancias que obtiene de aumentar los resultados y reducir los costos.
- Dejar de gastar dinero en tareas complicadas no diferenciadas: AWS se encarga de las tareas complicadas que corresponden a las operaciones del centro de datos, como montar servidores en bastidores, apilarlos y proporcionarles electricidad. También elimina la carga operativa de administrar los sistemas operativos y las aplicaciones con servicios administrados. Esto le permite centrarse en los clientes y los proyectos empresariales, en lugar de en la infraestructura de TI.
- Analizar y asignar los gastos: la nube facilita la identificación precisa del uso y de los costos de los sistemas, lo que luego permite atribuir de forma transparente los costos de TI a los propietarios de las cargas de trabajo individuales. Esto ayuda a medir el retorno de la inversión (ROI) y ofrece a los propietarios de las cargas de trabajo la oportunidad de optimizar sus recursos y reducir los costos.

Definición

Existen cinco áreas de prácticas recomendadas para la optimización de los costos en la nube:

- Práctica de la administración financiera en la nube
- Concientización sobre los gastos y el uso
- Recursos rentables
- Administración de los recursos de oferta y demanda
- Optimización con el paso del tiempo

Al igual que con los demás pilares del Well-Architected Framework, existen compensaciones que se deben tener en cuenta, como, por ejemplo, si se debe optimizar la velocidad de la comercialización o los costos. En algunos casos, lo mejor es optimizar la velocidad (introducirse en el mercado con rapidez, lanzar características nuevas o, simplemente, cumplir un plazo) en lugar de invertir en la optimización de costos iniciales. A veces, las decisiones de diseño se rigen por el apuro en lugar de los datos, y siempre existe la tentación de sobrecompensar “solo por si acaso”, en lugar de dedicar más tiempo a los puntos de referencia para que la implementación sea más rentable. Esto puede conducir a implementaciones con

demasiado aprovisionamiento y poca optimización. Sin embargo, es una opción razonable para cuando debe migrar mediante “lift and shift” los recursos de su entorno en las instalaciones hacia la nube y, luego, optimizarlos. Invertir la cantidad de esfuerzo correcta en una estrategia de optimización de costos por anticipado le permite obtener los beneficios económicos de la nube con mayor facilidad, garantizar una adherencia consistente a las prácticas recomendadas y evitar el exceso de aprovisionamiento innecesario. En la siguientes secciones, se ofrecen técnicas y prácticas recomendadas para la implementación inicial y continua de la administración financiera en la nube y la optimización de costos para sus cargas de trabajo.

Prácticas recomendadas

Temas

- Práctica de la administración financiera en la nube (p. 34)
- Concientización sobre los gastos y el uso (p. 35)
- Recursos rentables (p. 36)
- Administración de los recursos de oferta y demanda (p. 37)
- Optimización con el paso del tiempo (p. 37)

Práctica de la administración financiera en la nube

Con la adopción de la nube, los equipos de tecnología innovan más rápido debido a que se acortan los ciclos de aprobación, adquisición e implementación de la infraestructura. Se necesita un nuevo enfoque hacia la administración financiera en la nube a fin de materializar el valor del negocio y el éxito financiero. Este enfoque se centra en la administración financiera en la nube y desarrolla capacidades en su organización mediante la implementación de un amplio desarrollo de conocimientos, programas, recursos y procesos organizacionales.

Muchas organizaciones están compuestas por numerosas unidades con prioridades diferentes. La capacidad de alinear su organización hacia un conjunto acordado de objetivos financieros y de proporcionar a su organización los mecanismos necesarios para lograrlos creará una organización más eficiente. Una organización competente innovará y creará con mayor rapidez, será más ágil y se adaptará a cualquier factor interno o externo.

En AWS, puede utilizar Cost Explorer y, de forma opcional, Amazon Athena y Amazon QuickSight con el informe de uso y costo (CUR) para generar conciencia del uso y de los costos en toda la organización. AWS Budgets proporciona notificaciones proactivas con respecto al uso y a los costos. Los blogs de AWS ofrecen información sobre los nuevos servicios y las funciones a fin de garantizar que se mantenga actualizado respecto del lanzamiento de nuevos servicios.

La siguiente pregunta se enfoca en estas consideraciones para la optimización de costos. (Para ver una lista con las preguntas y las prácticas recomendadas sobre la optimización de costos, consulte el Apéndice (p. 75)).

COSTOS 1: ¿Cómo implementar la administración financiera en la nube?
La implementación de la administración financiera en la nube permite a las organizaciones comprender el valor de negocio y éxito financiero a medida que optimizan los costos, el uso y el escalado en AWS.

Cuando desarrolle una función de optimización de costos, trabaje con los miembros del equipo, pero también incluya expertos en administración financiera en la nube y optimización de costos para complementarlo. Aquellas personas que ya forman parte del equipo comprenderán cómo funciona la organización actualmente y aprenderán a implementar las mejoras con rapidez. Además, considere incluir personas que cuenten con habilidades especializadas o complementarias, como las habilidades analíticas y de administración de proyectos.

Cuando implemente la concientización sobre los costos en su organización, mejore o desarrolle programas o procesos ya existentes. Es mucho más rápido agregar características a procesos y programas ya existentes que desarrollar nuevos. Los resultados se lograrán con mayor rapidez.

Concientización sobre los gastos y el uso

El aumento de flexibilidad y agilidad que posibilita la nube incentiva la innovación, el desarrollo y la implementación acelerados. La nube elimina los procesos manuales y reduce el tiempo que toma el aprovisionamiento de la infraestructura en las instalaciones, incluidas la identificación de las especificaciones del hardware, la negociación de las cotizaciones de precios, la administración de las órdenes de compra, la programación de los envíos y la implementación de los recursos. Sin embargo, la facilidad de uso y la capacidad bajo demanda prácticamente ilimitada requiere de una nueva forma de pensamiento sobre los gastos.

Muchas empresas constan de varios sistemas ejecutados por varios equipos. La capacidad de asignar los costos de los recursos a la organización individual o a los propietarios de los productos impulsa un comportamiento de uso eficiente y ayuda a reducir los gastos innecesarios. La asignación precisa de los costos le permite saber qué productos son realmente rentables y tomar decisiones bien fundamentadas sobre el destino del presupuesto.

En AWS, puede crear una estructura de cuenta con AWS Organizations o AWS Control Tower, que ayuda a separar y asignar los costos y el uso. Además, puede etiquetar los recursos para implementar la información de la empresa y de la organización en los costos y en el uso. Utilice AWS Cost Explorer para visualizar los costos y el uso o cree paneles y análisis personalizados y análisis con Amazon Athena y Amazon QuickSight. El control de los costos y el uso se efectúa mediante notificaciones a través de AWS Budgets. También se pueden realizar controles con AWS Identity and Access Management (IAM) y Service Quotas.

Las siguientes preguntas se enfocan en estas consideraciones para la optimización de costos.

COSTOS 2 ¿Cómo se controla el uso?
Establezca políticas y mecanismos a fin de asegurar que se incurra en los costos adecuados a la vez que se logran los objetivos. Mediante la aplicación del enfoque de distribución de la autoridad y la responsabilidad, puede implementar innovaciones sin gastar demasiado.

COSTOS 3 ¿Cómo se monitorean el uso y los costos?
Establezca políticas y procedimientos para monitorear y asignar de forma adecuada los costos. Esto le permite medir y mejorar los niveles de rentabilidad correspondientes a esta carga de trabajo.

COSTOS 4 ¿Cómo se retiran los recursos?
Implemente el control de cambios y la administración de recursos desde el inicio de los proyectos hasta el final de su vida útil. Esto garantizará que pueda desactivar o terminar los recursos que no utilice a fin de reducir el desperdicio.

Puede utilizar las etiquetas de asignación de costos para clasificar los costos y el uso de AWS y realizar un seguimiento de ellos. Cuando etiqueta sus recursos de AWS (como las instancias EC2 o los buckets de S3), AWS genera un informe de uso y costo con sus etiquetas y su uso. Puede aplicar las etiquetas que representen las categorías de la organización (como los centros de costos, los nombres de cargas de trabajo o los propietarios) a fin de organizar sus costos en varios servicios.

Asegúrese de utilizar el nivel adecuado de detalle y especificación en el monitoreo y los informes de uso y costo. Para obtener información y tendencias de alto nivel, aproveche la granularidad diaria con AWS Cost Explorer. Para ejecutar un análisis y una inspección más detallados, aproveche la granularidad por hora de AWS Cost Explorer o de Amazon Athena y Amazon QuickSight con el informe de uso y costo (CUR) con granularidad por hora.

La combinación de los recursos etiquetados con el seguimiento del ciclo de vida de las entidades (trabajadores, proyectos) posibilita la identificación de los recursos huérfanos o los proyectos que ya no generan valor para la organización y que deberían retirarse. Puede configurar alertas de facturación que lo notifiquen sobre los gastos excesivos previstos.

Recursos rentables

El uso de las instancias y los recursos adecuados para su carga de trabajo es fundamental a la hora de ahorrar en los costos. Por ejemplo, un proceso de elaboración de informes puede tardar cinco horas en ejecutarse en un servidor más pequeño, pero puede tardar una hora en un servidor más grande, que es el doble de costoso. Ambos servidores le brindan el mismo resultado, pero el más pequeño incurre en más costos con el paso del tiempo.

Una carga de trabajo de buena arquitectura utiliza los recursos más rentables y que pueden generar un impacto económico positivo y significativo. Además, tiene la posibilidad de utilizar servicios administrados para reducir sus costos. Por ejemplo, en lugar de mantener servidores para enviar correos electrónicos, puede utilizar un servicio que cobre por mensaje.

AWS ofrece una amplia variedad de opciones de precios flexibles y rentables para adquirir instancias de Amazon EC2 y otros servicios de la forma que mejor se ajuste a sus necesidades. Las Instancias a petición permiten pagar la capacidad de cómputo por hora y no requieren compromisos mínimos. Los Savings Plans y las instancias reservadas (IR) ofrecen ahorros de hasta un 75 % de descuento con respecto al precio bajo demanda. Con las instancias de spot, puede aprovechar la capacidad de Amazon EC2 no utilizada, además de obtener ahorros de hasta el 90 % menos sobre los precios bajo demanda. Instancias de spot son adecuadas cuando el sistema puede tolerar el uso de una flota de servidores en la que los servidores individuales pueden intercambiarse de forma dinámica, como los servidores web sin estado, el procesamiento por lotes o cuando se utiliza la informática de alto rendimiento (HPC) o big data.

La selección adecuada del servicio también puede reducir el uso y los costos, como CloudFront para minimizar la transferencia de datos, o eliminar por completo los costos, como utilizar Amazon Aurora on RDS para eliminar los costos altos de las licencias de bases de datos.

Las siguientes preguntas se enfocan en estas consideraciones para la optimización de costos.

COSTOS 5 ¿Cómo se evalúan los costos al momento de elegir los servicios?
Amazon EC2, Amazon EBS y Amazon S3 son servicios de componentes básicos de AWS. Los servicios administrados, como Amazon RDS y Amazon DynamoDB, son servicios de AWS de mayor nivel o de nivel de aplicaciones. Si selecciona los bloques de creación y los servicios administrados adecuados, puede optimizar los costos de la carga de trabajo. Por ejemplo, si usa servicios administrados, puede reducir o eliminar una gran parte de los gastos generales administrativos y operativos, lo que le brindará la libertad para trabajar en las aplicaciones y las actividades relacionadas con el negocio.

COSTOS 6: ¿Cómo se cumplen los objetivos de costos al seleccionar el tipo, el tamaño y el número de recursos?
Asegúrese de elegir el tamaño de recurso y el número de recursos adecuados para la tarea en cuestión. El gasto se minimiza seleccionando el tipo, el tamaño y el número de recursos más rentables.

COSTOS 7: ¿Cómo se utilizan los modelos de precios para reducir los costos?
Use el modelo de precios más adecuado para sus recursos con el fin de minimizar los gastos.

COSTOS 8: ¿Cómo se planean los cargos por transferencia de datos?
Asegúrese de planear y monitorear los cargos por transferencia de datos para poder tomar decisiones sobre arquitectura con el fin de minimizar los costos. Un pequeño pero efectivo cambio en la arquitectura puede reducir radicalmente sus costos operativos con el paso del tiempo.

Si se consideran los costos durante la selección del servicio y se utilizan herramientas, como Cost Explorer y AWS Trusted Advisor, para revisar regularmente el uso de AWS, puede monitorearlo de forma activa y ajustar las implementaciones como corresponda.

Administración de los recursos de oferta y demanda

Una vez que migre hacia la nube, solo pagará lo que necesite. Puede suministrar recursos para adaptarse a la demanda de la carga de trabajo en el momento en que se necesitan, lo que elimina la necesidad de un sobreaprovisionamiento costoso y que desperdicia recursos. También puede modificar la demanda a través de la limitación controlada, un búfer o una cola para reducir la demanda y satisfacerla con menos recursos, lo que resulta en menores costos, o puede procesarla más tarde con un servicio por lotes.

En AWS, puede aprovisionar recursos de forma automática para satisfacer la demanda de la carga de trabajo. El escalado automático según el enfoque basado en la demanda o en el tiempo le permite agregar y eliminar recursos según sea necesario. Si puede anticipar los cambios en la demanda, puede ahorrar más dinero y asegurarse de que sus recursos satisfagan las necesidades de la carga de trabajo. Puede utilizar Amazon API Gatewaypara implementar limitaciones controladas, o Amazon SQS para implementar una cola en la carga de trabajo. Ambos servicios le permiten modificar la demanda de los componentes de su carga de trabajo.

La siguiente pregunta se enfoca en estas consideraciones para la optimización de costos.

COSTOS 9: ¿Cómo se administran los recursos de la oferta y demanda?
Para una carga de trabajo que tiene gastos y rendimiento equilibrados, asegúrese de que se use todo lo que pague y evite significativamente las instancias subutilizadas. Una métrica de utilización manipulada en cualquier dirección tiene un impacto adverso en su organización, ya sea en los costos operativos (rendimiento degradado debido a la sobreutilización) o los gastos de AWS desperdiciados (debido al sobreaprovisionamiento).

Cuando diseñe la modificación de los recursos de la oferta y la demanda, piense de forma activa sobre los patrones de uso, el tiempo que toma aprovisionar nuevos recursos y la predictibilidad de los patrones de la demanda. Al administrar la demanda, asegúrese de tener una cola o un búfer del tamaño correcto y de responder a la demanda de la carga de trabajo en el periodo requerido.

Optimización con el paso del tiempo

A medida que AWS lanza nuevos servicios y características, una práctica recomendada es revisar las decisiones sobre la arquitectura existente para garantizar que siguen siendo la opción más rentable. Cuando los requisitos cambian, debe ser enérgico a la hora de retirar recursos, servicios y sistemas completos que ya no necesite.

La implementación de nuevos tipos de características o recursos puede optimizar su carga de trabajo progresivamente, mientras minimiza el esfuerzo necesario para implementar el cambio. Esto ofrece mejoras continuas en la eficiencia a lo largo del tiempo y garantiza que sigue utilizando la tecnología más actualizada para reducir los costos operativos. Además, puede reemplazar los componentes de la carga de trabajo o agregarle nuevos, así como también nuevos servicios. Esto puede proporcionar aumentos significativos en la eficiencia, por lo que es fundamental revisar regularmente la carga de trabajo e implementar nuevos servicios y características.

La siguiente pregunta se enfoca en estas consideraciones para la optimización de costos.

COSTOS 10: ¿Cómo se evalúan los nuevos servicios?
A medida que AWS lanza nuevos servicios y características, una práctica recomendada es revisar las decisiones sobre la arquitectura existente para garantizar que siguen siendo la opción más rentable.

Cuando revise sus implementaciones con regularidad, evalúe cómo los servicios más nuevos pueden ayudarlo a ahorrar dinero. Por ejemplo, Amazon Aurora on RDS puede reducir los costos de las bases de datos relacionales. El uso de servicios sin servidor, como Lambda, puede eliminar la necesidad de operar y administrar las instancias para ejecutar el código.

Recursos

Consulte los siguientes recursos para obtener más información sobre nuestras prácticas recomendadas para la optimización de costos.

Documentación

- Documentación de AWS

Documento técnico

- Pilar de optimización de costos

Proceso de revisión

La revisión de la arquitectura se debe realizar de manera consistente y adoptar un enfoque libre de culpas que fomente la reflexión profunda. Debe ser un proceso ligero (horas, no días) que sea una conversación y no una auditoría. El objetivo de revisar una arquitectura consiste en identificar todos los problemas graves que puedan necesitar solucionarse o las áreas que se puedan mejorar. El resultado de la revisión es un conjunto de acciones que debe mejorar la experiencia de un cliente que utiliza la carga de trabajo.

Como se analiza en la sección “Sobre la arquitectura”, el objetivo es que cada miembro del equipo se responsabilice por la calidad de su arquitectura. Recomendamos que los miembros del equipo que construyen una arquitectura utilicen el Marco de Buena Arquitectura para revisarla continuamente, en lugar de llevar a cabo una reunión formal de revisión. Un enfoque continuo permite que los miembros del equipo actualicen las respuestas a medida que la arquitectura evoluciona y la mejoren a medida que se entregan las características.

AWS Well-Architected Framework se adapta a la forma en que AWS revisa los sistemas y servicios internamente. Se basa en un conjunto de principios de diseño que influye en el enfoque arquitectónico y en preguntas que aseguran que las personas no descuiden las áreas que a menudo aparecen en el análisis de causa raíz (RCA). Siempre que haya un problema grave con un sistema interno, el servicio de AWS o el cliente, analizamos el RCA para ver si podemos mejorar los procesos de revisión que utilizamos.

Las revisiones se deben aplicar a los hitos clave en el ciclo de vida del producto al principio de la etapa de diseño para evitar caminos sin retorno que son difíciles de cambiar antes de la fecha de la puesta en marcha. Luego de entrar en producción, su carga de trabajo seguirá evolucionando a medida que agregue características y cambie las implementaciones tecnológicas. La arquitectura de una carga de trabajo cambia con el tiempo. Tendrá que cumplir con las prácticas recomendadas de higiene para evitar que sus características arquitectónicas se degraden a medida que evolucionan. A medida que hace cambios importantes en la arquitectura, debe cumplir con una serie de procesos de higiene que incluyen una revisión de Well-Architected.

Si desea utilizar la revisión como una instantánea única o como una medición independiente, querrá asegurarse de tener a todas las personas adecuadas en la conversación. A menudo descubrimos que las revisiones son la primera vez que un equipo entiende realmente lo que ha implementado. Un enfoque que funciona bien al revisar la carga de trabajo de otro equipo es tener una serie de conversaciones informales sobre su arquitectura en la que se pueden obtener las respuestas a la mayoría de las preguntas. Luego puede seguir con una o dos reuniones en las que puede ganar claridad o profundizar en áreas de ambigüedad o riesgo percibido.

Estos son algunos puntos sugeridos para facilitar sus reuniones:

- Una sala de reunión con pizarras
- La impresión de diagramas o notas de diseño
- Lista de acción de preguntas que necesitan investigación fuera de banda para responderlas (p. ej., “¿activamos el cifrado o no?”)

Después de haber hecho la revisión, debe tener una lista de problemas a los que puede dar prioridad en función de su contexto empresarial. También querrá tener en cuenta el impacto de esos problemas en el trabajo diario de su equipo. Si aborda estos problemas a tiempo, podría liberar tiempo para trabajar en la creación de valor empresarial en lugar de resolver problemas recurrentes. A medida que aborda los problemas, puede actualizar su revisión para ver cómo mejora la arquitectura.

Si bien el valor de una revisión es evidente después de haber realizado una, puede que un nuevo equipo se resista al principio. Estas son algunas de las objeciones que se pueden manejar a través de la capacitación del equipo sobre los beneficios de una revisión:

- “Estamos muy ocupados” (Suele decirse cuando el equipo se prepara para un gran lanzamiento).
  - Si se está preparando para un gran lanzamiento, no querrá que haya problemas. La revisión le ayudará a comprender todos los problemas que pudo pasar por alto.
  - Le recomendamos que lleve a cabo revisiones al principio del ciclo de vida del producto para descubrir los riesgos y desarrollar un plan de mitigación que cumpla con la hoja de ruta de entrega de características.
- “No disponemos del tiempo para hacer algo con los resultados” (Se dice con frecuencia cuando hay un evento impostergable que abordan, como el Super Bowl).
  - Estos eventos son inamovibles. ¿De verdad quiere ir sin saber los riesgos de su arquitectura? Incluso si no aborda estos problemas, puede tener manuales de estrategia para abordarlos si suceden
- “No queremos que otros conozcan los secretos de nuestra implementación de la solución”
  - Si señala al equipo las preguntas del Marco de Buena Arquitectura, verán que ninguna revela información de propiedad comercial o técnica.

A medida que lleve a cabo varias revisiones con los equipos en su organización, podrá identificar problemas temáticos. Por ejemplo, puede ver que un grupo de equipos tiene conjuntos de problemas en un pilar o tema en particular. Querrá analizar todas sus revisiones de manera holística e identificar los mecanismos, las capacitaciones o las charlas de ingeniería principal que pueden ayudar a abordar esos problemas temáticos.

Conclusión

AWS Well-Architected Framework ofrece las prácticas recomendadas de arquitectura a través de los cinco pilares para diseñar y operar sistemas en la nube confiables, seguros, eficientes y rentables. El Marco ofrece una serie de preguntas que permiten revisar una arquitectura actual o propuesta. También ofrece una serie de prácticas recomendadas de AWS para cada pilar. Utilice el Marco en su arquitectura para producir sistemas estables y eficientes, que le permitan enfocarse en sus requisitos funcionales.

Colaboradores

Las siguientes personas y organizaciones colaboraron a la hora de crear este documento:

- Rodney Lester, director sénior de Well-Architected, Amazon Web Services
- Brian Carlson, líder de operaciones de Well Architected, Amazon Web Services
- Ben Potter, líder de seguridad de Well-Architected, Amazon Web Services
- Eric Pullen, líder de rendimiento de Well-Architected, Amazon Web Services
- Seth Eliot, líder de fiabilidad de Well-Architected, Amazon Web Services
- Nathan Besh, líder de costos de Well-Architected, Amazon Web Services
- Jon Steele, técnico de cuentas sénior, Amazon Web Services
- Ryan King, director técnico de programas, Amazon Web Services
- Erin Rifkin, directora sénior de productos, Amazon Web Services
- Max Ramsay, arquitecto de soluciones principales de seguridad, Amazon Web Services
- Scott Paddock, arquitecto de soluciones de seguridad, Amazon Web Services
- Callum Hughes, arquitecto de soluciones, Amazon Web Services

Documentación adicional

- Conformidad de la nube de AWS
- Programa para socios de AWS Well-Architected
- AWS Well-Architected Tool
- Página de inicio de AWS Well-Architected
- Documento técnico sobre el pilar de optimización de costos
- Documento técnico sobre el pilar de excelencia operativa
- Documento técnico sobre el pilar de eficiencia de rendimiento
- Documento técnico sobre el pilar de fiabilidad
- Documento técnico sobre el pilar de seguridad
- Amazon Builders' Library

Revisiones del documento

Para recibir notificaciones sobre actualizaciones a este documento técnico, suscríbase a la fuente RSS.

update-history-change	update-history-description	update-history-date
Actualización menor (p. 44)	Cambios editoriales menores a lo largo del documento.	July 15, 2020
Actualizaciones para el nuevo marco (p. 44)	Revisión y reescritura de la mayoría de las preguntas y respuestas.	July 8, 2020
Documento técnico actualizado (p. 44)	Se agregaron AWS Well-Architected Tool, enlaces a los laboratorios de AWS Well-Architected y correcciones menores para permitir una versión del marco en varios idiomas.	July 1, 2019
Documento técnico actualizado (p. 44)	Se revisaron y reescribieron la mayoría de las preguntas y respuestas para asegurar que las primeras se enfoquen en un tema a la vez. Esto hizo que algunas de las preguntas anteriores se dividieran en varias preguntas. Se agregaron otros términos comunes a las definiciones (carga de trabajo, componente, etc.). Se cambió la presentación de la pregunta del cuerpo principal para incluir texto descriptivo.	November 1, 2018
Documento técnico actualizado (p. 44)	Se actualizó para simplificar el texto de las preguntas, estandarizar las respuestas y mejorar la legibilidad.	June 1, 2018
Documento técnico actualizado (p. 44)	Se trasladó la excelencia operativa al frente de los pilares y se reescribió para enmarcar los demás pilares. Se actualizaron los demás pilares para reflejar la evolución de AWS.	November 1, 2017
Documento técnico actualizado (p. 44)	Se actualizó el Marco para incluir el pilar de excelencia operativa, y se revisaron y actualizaron los demás pilares para reducir la duplicación e incorporar aprendizajes de las revisiones que se llevaron a cabo con miles de clientes.	November 1, 2016

Actualizaciones menores (p. 44)	Se actualizó el apéndice con la información vigente de Amazon CloudWatch Logs.	November 1, 2015
Publicación inicial (p. 44)	AWS Well-Architected Framework publicado.	October 1, 2015

Apéndice: Preguntas y prácticas recomendadas

Temas

- Excelencia operativa (p. 46)
- Seguridad (p. 54)
- Fiabilidad (p. 61)
- Eficiencia de rendimiento (p. 69)
- Optimización de costos (p. 75)

Excelencia operativa

Temas

- Organización (p. 46)
- Preparación (p. 49)
- Operación (p. 52)
- Evolución (p. 54)

Organización

OPS 1 ¿Cómo determina cuáles son sus prioridades?
Todos deben entender su rol en el proceso que permite alcanzar el éxito empresarial. Cuente con objetivos compartidos a fin de establecer prioridades para los recursos. Esto maximizará los beneficios de sus esfuerzos.

Prácticas recomendadas:

- Evalúe las necesidades de los clientes externos: involucre a las partes interesadas clave, incluidos los equipos de negocios, desarrollo y operaciones, para determinar dónde concentrar los esfuerzos orientados a las necesidades de los clientes externos. Esto garantizará que comprenda por completo el respaldo operativo que se necesita para lograr los resultados empresariales deseados.
- Evalúe las necesidades de los clientes internos: involucre a las partes interesadas clave, incluidos los equipos de negocios, desarrollo y operaciones, al momento de determinar dónde concentrar los esfuerzos orientados a las necesidades de los clientes internos. Esto garantizará que comprenda por completo el respaldo operativo que se necesita para lograr los resultados empresariales.
- Evalúe los requisitos de gobernanza: asegúrese de conocer las directrices o las obligaciones que estableció su organización y que pueden exigir o resaltar un enfoque específico. Evalúe los factores internos, como la política, los estándares y los requisitos de la organización. Compruebe que cuenta con los mecanismos necesarios para identificar cambios en la gobernanza. Si no se identifican requisitos de gobernanza, asegúrese de haber aplicado la debida diligencia a esta tarea.

- Evalúe los requisitos de conformidad: evalúe los factores externos, como los requisitos de conformidad normativa y los estándares del sector, para asegurarse de conocer las directrices o las obligaciones que pueden exigir o resaltar un enfoque específico. Si no se identifican requisitos de conformidad, asegúrese de aplicar la debida diligencia a esta tarea.
- Evalúe el panorama de amenazas: evalúe las amenazas a su negocio (por ejemplo, la competencia, los riesgos y las cargas empresariales, los riesgos operativos y las amenazas a la seguridad de la información) y mantenga la información actual en un registro de riesgos. Incluya el impacto de los riesgos a la hora de determinar dónde concentrar los esfuerzos.
- Evalúe las compensaciones: evalúe el efecto de las compensaciones entre intereses contrapuestos o enfoques alternativos para poder tomar decisiones con fundamentos al momento de determinar en dónde concentrar esfuerzos o a la hora de establecer un curso de acción. Por ejemplo, se puede priorizar la aceleración de la comercialización de características nuevas por encima de la optimización de costos. También puede elegir una base de datos relacional para datos no relacionales con el fin de simplificar el esfuerzo de migración de un sistema, en lugar de migrar a una base de datos optimizada para su tipo de datos y actualizar la aplicación.
- Administre los beneficios y los riesgos: administre los beneficios y los riesgos para tomar decisiones con fundamentos al momento de determinar dónde concentrar los esfuerzos. Por ejemplo, puede resultar beneficioso implementar una carga de trabajo que tenga problemas sin resolver de manera que nuevas características importantes puedan estar disponibles para los clientes. Tal vez sea posible mitigar los riesgos asociados o quizás se vuelva inaceptable permitir que un riesgo permanezca, en cuyo caso tendrá que tomar medidas para abordarlo.

OPS 2 ¿Cómo estructura su organización de manera que respalde los resultados empresariales?

Los equipos deben comprender el rol que juegan en el logro de los resultados empresariales. Los equipos deben comprender el rol que tienen en el éxito de otros equipos, conocer el rol de los demás equipos en su propio éxito y tener objetivos en común. Comprender la responsabilidad, la propiedad, la manera en que se toman las decisiones y quién tiene la autoridad para hacerlo ayudará a concentrar los esfuerzos y a maximizar los beneficios de sus equipos.

Prácticas recomendadas:

- Los recursos tienen propietarios identificados: se debe comprender quién es propietario de cada aplicación, carga de trabajo, plataforma y componente de infraestructura; qué valor de negocio proporciona ese componente; y por qué existe esa propiedad. Comprender el valor de negocio de estos componentes individuales y la manera en que respaldan los resultados empresariales determina los procesos y los procedimientos que se les aplican.
- Los procesos y los procedimientos tienen propietarios identificados: se debe comprender quién es propietario de la definición de los procesos y los procedimientos individuales, por qué se usan esos procesos y procedimientos específicos, y por qué existe esa propiedad. Comprender las razones por las que se usan procesos y procedimientos específicos permite identificar oportunidades de mejora.
- Las actividades operativas tienen propietarios identificados que son responsables de su rendimiento: se debe comprender quién tiene la responsabilidad de llevar a cabo actividades específicas en cargas de trabajo definidas y por qué existe esa responsabilidad. Comprender quién tiene la responsabilidad de llevar a cabo actividades determina quién realizará la actividad, validará el resultado y proporcionará retroalimentación al propietario de la actividad.
- Los miembros del equipo saben de qué son responsables: comprender las responsabilidades de su rol y de qué manera contribuye a los resultados empresariales determina la priorización de las tareas y por qué su rol es importante. Esto permite a los miembros del equipo reconocer las necesidades y responder de forma adecuada.
- Existen mecanismos para identificar la responsabilidad y la propiedad: cuando no se identifica ni a una persona ni a un equipo, existen vías de escalamiento definidas, las cuales llevan a alguien con la autoridad suficiente como para asignar propiedad o planear para que se aborde esa necesidad.

- Existen mecanismos para solicitar incorporaciones, cambios y excepciones: usted puede realizar solicitudes a los propietarios de procesos, procedimientos y recursos. Tome decisiones con fundamento para aprobar solicitudes siempre que sean posibles y se determine que son adecuadas después de una evaluación de los beneficios y los riesgos.
- Las responsabilidades entre equipos se negocian o definen de manera anticipada: existen acuerdos definidos o negociados entre los equipos que describen cómo trabajan entre sí y se respaldan mutuamente (por ejemplo, tiempos de respuesta, objetivos de nivel de servicio o acuerdos de nivel de servicio). Comprender el efecto del trabajo de los equipos sobre los resultados empresariales y los resultados de otros equipos y organizaciones determina la priorización de sus tareas y les permite responder de manera adecuada.

OPS 3 ¿Cómo respalda su cultura organizativa los resultados empresariales?

Brinde soporte a los miembros de su equipo para que puedan ser más eficaces a la hora de tomar medidas y de respaldar los resultados empresariales.

Prácticas recomendadas:

- Patrocinio ejecutivo: los líderes principales establecen expectativas claras para la organización y evalúan el éxito. Son patrocinadores, defensores e impulsores de la adopción de las prácticas recomendadas y de la evolución de la organización
- Permitir a los miembros del equipo tomar medidas cuando los resultados estén en riesgo: el propietario de la carga de trabajo definió la orientación y el alcance, lo que permite a los miembros del equipo responder cuando los resultados estén en riesgo. Los mecanismos de escalamiento se utilizan para recibir indicaciones cuando los eventos están fuera del alcance definido.
- Se alienta el escalamiento: ya que cuentan con los mecanismos necesarios para hacerlo, se alienta a los miembros del equipo a que remitan sus inquietudes a los responsables de la toma de decisiones y las partes interesadas si creen que los resultados están en peligro. El escalamiento debe realizarse a tiempo y con frecuencia, de manera que se puedan identificar los riesgos y se pueda evitar que causen incidentes.
- Las comunicaciones deben ser oportunas, claras y factibles: existen mecanismos que se utilizan para notificar oportunamente a los miembros del equipo sobre los riesgos conocidos y los eventos planificados. Se brinda el contexto, los detalles y el tiempo (cuando es posible) necesarios para ayudar a determinar si se requiere alguna acción, y de qué acción se trata, y también para actuar a tiempo. Por ejemplo, notificar sobre las vulnerabilidades del software para que se pueda acelerar la implementación de parches o notificar sobre las promociones planificadas de ventas para que se pueda implementar un congelamiento de cambios a fin de evitar el riesgo de interrupción del servicio.
- Se alienta a la experimentación: la experimentación acelera el aprendizaje y mantiene a los miembros del equipo interesados e involucrados. Un resultado no deseado es un experimento exitoso que identificó un camino que no conduce al éxito. No se penaliza a los miembros del equipo por experimentos exitosos con resultados no deseados. Es necesario experimentar para dar lugar a la innovación y para que las ideas se transformen en resultados.
- Se permite y se alienta a que los miembros del equipo mantengan y desarrollen sus habilidades: los miembros deben desarrollar sus conjuntos de habilidades para adoptar nuevas tecnologías y admitir cambios en la demanda y las responsabilidades a favor de las cargas de trabajo. Con frecuencia, el desarrollo de las habilidades en tecnologías nuevas es una fuente de satisfacción para los miembros del equipo y respalda a la innovación. Apoye a los miembros de su equipo en la búsqueda y el mantenimiento de certificaciones del sector que validen y reconozcan sus habilidades en desarrollo. Proporcione formación interdisciplinaria para promover el intercambio de conocimientos y reducir el riesgo de un impacto significativo si se pierden miembros del equipo capacitados y experimentados con conocimiento institucional. Ofrezca tiempo definido y específico para el aprendizaje.
- Brindar recursos a los equipos de manera adecuada: mantenga la capacidad de los miembros del equipo y ofrezca herramientas y recursos para respaldar las necesidades de la carga de trabajo.

Saturar de cargas a los miembros del equipo aumenta el riesgo de incidentes que surgen de errores humanos. Invertir en herramientas y recursos (por ejemplo, automatizar las actividades frecuentes) puede aumentar la efectividad de su equipo, lo que les permite admitir otras actividades.

- Se alientan y se buscan las opiniones diversas en cada equipo y entre ellos: aproveche la diversidad entre las organizaciones para buscar varias perspectivas únicas. Utilice esta perspectiva para aumentar el nivel de innovación, desafiar sus suposiciones y reducir el riesgo de sesgo de confirmación. Aumente los niveles de inclusión, diversidad y accesibilidad dentro de sus equipos para obtener perspectivas beneficiosas.

Preparación

OPS 4 ¿Cómo diseña la carga de trabajo de manera que pueda comprender su estado?

Diseñe su carga de trabajo de manera que brinde la información necesaria de todos los componentes (por ejemplo, métricas, registros y rastreos) y pueda comprender su estado interno. Esto le permite ofrecer respuestas efectivas cuando sea necesario.

Prácticas recomendadas:

- Implementar telemetría de la aplicación: provea al código de la aplicación herramientas que permitan emitir información acerca del estado interno, el estado y la obtención de resultados empresariales. Por ejemplo, profundidad de la cola, mensajes de error y tiempos de respuesta. Utilice esta información para determinar cuándo se necesita una respuesta.
- Implementar y configurar la telemetría de la carga de trabajo: diseñe y configure la carga de trabajo para que emita información acerca del estado interno y el estado actual. Por ejemplo, considere el volumen de llamadas a la API, los códigos de estado HTTP y los eventos de escalado. Utilice esta información para poder determinar cuándo se necesita una respuesta.
- Implementar telemetría de la actividad del usuario: provea al código de la aplicación herramientas que permitan emitir información acerca de la actividad del usuario, como, por ejemplo, secuencias de clics o transacciones que se han iniciado, abandonado o completado. Utilice esta información para comprender cómo se utiliza la aplicación, identificar patrones de uso y determinar cuándo se necesita una respuesta.
- Implementar telemetría de la dependencia: diseñe y configure la carga de trabajo de manera que emita información acerca del estado (por ejemplo, accesibilidad o tiempo de respuesta) de los recursos de los que depende. Algunos ejemplos de dependencias externas son las bases de datos externas, los DNS y la conectividad a la red. Utilice esta información para determinar cuándo se necesita una respuesta.
- Implementar la trazabilidad de las transacciones: implemente el código de la aplicación y configure los componentes de la carga de trabajo de manera que emitan información sobre el flujo de transacciones en toda la carga de trabajo. Utilice esta información para determinar cuándo se necesita una respuesta y para ayudarlo a identificar los factores que contribuyen a un problema.

OPS 5 ¿Cómo reduce los defectos, facilita la corrección y mejora el flujo a la producción?

Adopte enfoques que mejoren el flujo de los cambios en la producción y que permitan la refactorización, la retroalimentación rápida sobre la calidad y la corrección de errores. Estos enfoques aceleran los cambios beneficiosos que se aplican a la fase de producción, limitan los problemas implementados y permiten una rápida identificación y solución de los problemas que acarrearon las actividades de implementación.

Prácticas recomendadas:

- Utilizar el control de versiones: utilice el control de versiones para habilitar el seguimiento de los cambios y las versiones.
- Evaluar y validar los cambios: pruebe y valide los cambios para ayudar a limitar y detectar errores. Automatice las pruebas a fin de reducir los errores causados por procesos manuales y, también, reducir el nivel de esfuerzo necesario para realizar las pruebas.
- Utilizar sistemas de administración de la configuración: utilice sistemas de administración de la configuración para realizar cambios en la configuración y rastrearlos. Estos sistemas reducen los errores causados por los procesos manuales y reducen el nivel de esfuerzo necesario para implementar cambios.
- Utilizar sistemas de administración de implementaciones y creaciones: utilice sistemas de administración de implementaciones y creaciones. Estos sistemas reducen los errores causados por los procesos manuales y reducen el nivel de esfuerzo necesario para implementar cambios.
- Llevar a cabo la administración de parches: lleve a cabo la administración de parches para obtener características, abordar problemas y mantener la conformidad con la gobernanza. Automatice la administración de parches a fin de reducir los errores causados por procesos manuales y, también, reducir el nivel de esfuerzo necesario para aplicar parches.
- Compartir estándares de diseño: comparta las prácticas recomendadas con los equipos para incrementar el conocimiento y maximizar los beneficios de los esfuerzos de desarrollo.
- Implementar prácticas para mejorar la calidad del código: implemente prácticas para mejorar la calidad del código y minimizar los defectos. Por ejemplo, el desarrollo basado en pruebas, las revisiones de códigos y la adopción de estándares.
- Utilizar varios entornos: utilice varios entornos para experimentar, desarrollar y evaluar la carga de trabajo. Utilice niveles de control en crecimiento a medida que los entornos se acercan a la producción con el fin de adquirir confianza en que las cargas de trabajo funcionarán como se previó al momento de la implementación.
- Realizar cambios pequeños, reversibles y frecuentes: los cambios frecuentes, pequeños y reversibles reducen el alcance y el impacto de un cambio. Esto facilita la resolución de problemas, permite correcciones más rápidas y proporciona la opción de restaurar los cambios.
- Automatizar por completo la integración y la implementación: automatice la creación, la implementación y la realización de pruebas de la carga de trabajo. Esto reduce los errores causados por los procesos manuales y reduce el esfuerzo necesario para implementar los cambios.

OPS 6 ¿Cómo mitiga los riesgos de implementación?

Adopte enfoques que ofrezcan una rápida valoración acerca de la calidad y permitan una rápida recuperación de aquellos cambios que no tengan los resultados deseados. La aplicación de estas prácticas mitiga el impacto de los problemas que surgen como consecuencia de la implementación de cambios.

Prácticas recomendadas:

- Planifique los cambios incorrectos: haga planes para volver a un estado correcto conocido o para corregir el entorno de producción en el caso de que un cambio no produzca el resultado deseado. Esta preparación reduce el tiempo de recuperación a través de respuestas más rápidas.
- Evaluar y validar los cambios: evalúe los cambios y valide los resultados en todas las etapas del ciclo de vida a fin de confirmar las nuevas características y minimizar el riesgo y el impacto de las implementaciones con errores.
- Utilice los sistemas de administración de implementaciones: utilice los sistemas de administración de implementaciones para hacer un seguimiento de los cambios e implementarlos. Esto reduce los errores causados por los procesos manuales y reduce los esfuerzos para implementar cambios.
- Evalúe con implementaciones limitadas: realice pruebas con implementaciones limitadas junto con sistemas existentes para confirmar los resultados deseados antes de implementarlos a una escala

completa. Por ejemplo, utilice pruebas de valor controlado de implementaciones o implementaciones únicas.

- Implementación con entornos paralelos: implemente cambios en entornos paralelos y, luego, haga la transición al nuevo entorno. Mantenga el entorno anterior hasta obtener una confirmación de que la implementación fue correcta. De este modo, se minimizan los tiempos de recuperación, ya que se permite la restauración del entorno anterior.
- Implementar cambios reversibles, pequeños y frecuentes: utilice cambios reversibles, pequeños y frecuentes para reducir su alcance. Esto permite que la resolución de problemas sea más sencilla y que las correcciones sean más rápidas, además de la posibilidad de revertir el cambio.
- Automatizar por completo la integración y la implementación: automatice la creación, la implementación y la realización de pruebas de la carga de trabajo. Esto reduce los errores causados por los procesos manuales y reduce los esfuerzos para implementar cambios.
- Automatice las pruebas y la restauración: automatice la prueba de los entornos implementados para confirmar los resultados deseados. Automatice la restauración al anterior estado correcto conocido cuando no se logren los resultados esperados, con el fin de minimizar los tiempos de recuperación y reducir los errores causados por los procesos manuales.

OPS 7 ¿Cómo sabe que está listo para dar respaldo a una carga de trabajo?

Evalúe la disposición operativa de sus cargas de trabajo, procesos y procedimientos y personal con el fin de comprender los riesgos operativos relacionados con su carga de trabajo.

Prácticas recomendadas:

- Garantice la capacidad del personal: cuente con un mecanismo para confirmar que dispone de la cantidad apropiada de personal capacitado para ofrecer respaldo a las necesidades operativas. Entrene a su personal y ajuste su capacidad según sea necesario a fin de mantener un respaldo eficaz.
- Garantice la revisión constante de la disposición operativa: garantice que se realice una revisión constante del nivel de preparación para operar una carga de trabajo. Las revisiones deben incluir, como mínimo, la disposición operativa de los equipos y la carga de trabajo y los requisitos de seguridad. Implemente actividades de revisión como código y active revisiones automáticas en respuesta a los eventos, cuando sea apropiado, a fin de garantizar la consistencia, la velocidad de ejecución y reducir los errores causados por los procesos manuales.
- Utilice manuales de procedimiento para su ejecución: los manuales de procedimientos consisten en procedimientos documentados para lograr resultados específicos. Permita respuestas rápidas y constantes para eventos que se comprendan bien a través de la documentación de los procedimientos en los manuales. Implemente manuales de procedimientos como código y active su ejecución en respuesta a los eventos, cuando sea apropiado, a fin de asegurar la consistencia, la velocidad de las respuestas y reducir los errores causados por los procesos manuales.
- Utilice los manuales de estrategias para investigar los problemas: habilite respuestas constantes y rápidas para los problemas que no se comprendan correctamente. Para ello, documente el proceso de investigación en los manuales de estrategias. Los manuales de estrategias contienen los pasos predefinidos que se realizan para identificar los factores que contribuyen a una situación de error. Los resultados de cualquier paso en el proceso se utilizan para determinar los próximos pasos a seguir hasta que se identifique o escale el problema.
- Tome decisiones fundamentadas para implementar sistemas y cambios: evalúe las capacidades del equipo para respaldar la carga de trabajo y la conformidad de la carga de trabajo con la gobernabilidad. Lleve a cabo esta evaluación en función de los beneficios de su implementación cuando determine si se debe realizar la transición de un sistema o de un cambio a la fase de producción. Comprenda los beneficios y los riesgos para tomar decisiones fundamentadas.

Operación

OPS 8 ¿Cómo comprende el estado de la carga de trabajo?

Defina, registre y analice las métricas de las cargas de trabajo para obtener visibilidad en los eventos de carga de trabajo y poder tomar las medidas adecuadas.

Prácticas recomendadas:

- Identifique los indicadores clave de rendimiento: identifique los indicadores clave de rendimiento (KPI) en función de los resultados empresariales deseados (por ejemplo, la tasa de pedidos, la tasa de retención de clientes y las ganancias frente a los gastos operativos) y los resultados de los clientes (por ejemplo, la satisfacción del cliente). Evalúe los KPI para determinar el éxito de la carga de trabajo.
- Defina las métricas de la carga de trabajo: defina las métricas de la carga de trabajo para medir el logro de los KPI (por ejemplo, los carros de compras abandonados, los pedidos realizados, el costo, el precio y los gastos de la carga de trabajo asignada). Defina las métricas de la carga de trabajo para medir el estado de dicha carga de trabajo (por ejemplo, el tiempo de respuesta de la interfaz, la tasa de error, las solicitudes realizadas, las solicitudes completadas y la utilización). Evalúe las métricas para determinar si la carga de trabajo logra los resultados deseados y para comprender el estado de la carga de trabajo.
- Recopile y analice las métricas de la carga de trabajo: lleve a cabo revisiones proactivas y regulares de las métricas para identificar las tendencias y determinar dónde se necesitan las respuestas adecuadas.
- Establezca puntos de referencia de las métricas de la carga de trabajo: establezca puntos de referencia para las métricas con el fin de ofrecer valores esperados como base para la comparación e identificación de los componentes de rendimiento bajo y alto. Identifique los límites para mejoras, investigaciones e intervenciones.
- Conozca los patrones esperados de actividad para la carga de trabajo: establezca los patrones de actividad de la carga de trabajo para identificar anomalías en su comportamiento y responder adecuadamente si es necesario.
- Genere una alerta cuando los resultados de una carga de trabajo estén en riesgo: genere una alerta cuando los resultados de la carga de trabajo estén en riesgo para que pueda responder adecuadamente si es necesario.
- Genere una alerta cuando se detecten anomalías en la carga de trabajo: genere una alerta cuando se detecten anomalías en la carga de trabajo para que pueda responder adecuadamente si es necesario.
- Valide el logro de los resultados y la efectividad de los KPI y las métricas : cree una vista de nivel empresarial de las operaciones de la carga de trabajo para ayudar a determinar si las necesidades se satisfacen y para identificar las áreas que necesitan mejoras con el fin de alcanzar los objetivos comerciales. Valide la efectividad de los KPI y de las métricas y revíselos si es necesario.

OPS 9 ¿Cómo comprende el estado de las operaciones?

Defina, registre y analice las métricas de las operaciones para obtener visibilidad en los eventos operativos y poder tomar las medidas adecuadas.

Prácticas recomendadas:

- Identifique los indicadores clave de rendimiento: identifique los indicadores clave de rendimiento (KPI) en función de la actividad comercial deseada (por ejemplo, entrega de nuevas características) y los resultados del cliente (por ejemplo, casos de servicio de atención al cliente). Evalúe los KPI para determinar el éxito de las operaciones.
- Defina las métricas de las operaciones: defina las métricas de las operaciones para medir el logro de los KPI (por ejemplo, implementaciones correctas e implementaciones con errores). Defina las métricas de

las operaciones para medir el estado de las actividades de dichas operaciones (por ejemplo, el tiempo promedio para la detección de un incidente [MTTD] y el tiempo promedio para la recuperación [MTTR] de un incidente). Evalúe las métricas para determinar si las operaciones logran los resultados deseados y para comprender el estado de sus actividades operativas.

- Recopile y analice las métricas de las operaciones: lleve a cabo revisiones proactivas y regulares de las métricas para identificar las tendencias y determinar dónde se necesitan las respuestas adecuadas.
- Establezca puntos de referencia de las métricas de las operaciones: establezca puntos de referencia para las métricas con el fin de ofrecer valores esperados como base para la comparación e identificación de actividades operativas de rendimiento alto y bajo.
- Conozca los patrones esperados de actividad para las operaciones: establezca los patrones de actividades operativas para identificar actividades anómalas, y así tener la capacidad responder adecuadamente si es necesario.
- Genere una alerta cuando los resultados de las operaciones estén en riesgo: genere una alerta cuando los resultados de las operaciones estén en riesgo para que pueda responder adecuadamente si es necesario.
- Genere una alerta cuando se detecten anomalías en las operaciones: genere una alerta cuando se detecten anomalías en las operaciones para que pueda responder adecuadamente si es necesario.
- Valide el logro de los resultados y la efectividad de los KPI y las métricas : cree una vista de nivel empresarial de las actividades operativas para ayudar a determinar si las necesidades se satisfacen y para identificar las áreas que necesitan mejoras con el fin de alcanzar los objetivos comerciales. Valide la efectividad de los KPI y de las métricas y revíselos si es necesario.

OPS 10 ¿Cómo administra los eventos de carga de trabajo y operaciones?

Prepare y valide procedimientos para responder a los eventos con el fin de minimizar la interrupción de su carga de trabajo.

Prácticas recomendadas:

- Utilizar procesos para la administración de eventos, incidentes y problemas: disponga de procesos para abordar eventos observados, eventos que necesitan intervención (incidentes) y eventos que necesitan intervención y que pueden repetirse o no se pueden resolver actualmente (problemas). Además, utilice estos procesos para mitigar el impacto que dichos eventos pueden causar en la empresa y sus clientes a través de respuestas adecuadas y oportunas.
- Disponer de un proceso por alerta: disponga de una respuesta clara (manual de procedimientos o de estrategias), que cuente con un propietario específicamente identificado, ante cualquier evento en el que se genere una alerta. De esta forma, garantiza respuestas rápidas y efectivas ante eventos operativos y evita que las notificaciones menos importantes oculten a los eventos que se pueden corregir.
- Priorizar los eventos operativos según el impacto empresarial: cuando varios eventos necesiten intervención, garantice que se traten en primer lugar los eventos más importantes para la empresa. Por ejemplo, los impactos pueden incluir la pérdida de la vida o lesiones, pérdidas financieras o daños a la reputación o la confianza.
- Definir rutas de escalamiento: defina rutas de escalamiento en los manuales de procedimientos y de estrategias. Incluya aquello que impulsa el escalamiento y los procedimientos que se necesitan para ello. Identifique propietarios específicos para cada acción y, de esta forma, garantice respuestas efectivas y rápidas para los eventos operativos.
- Habilitar las notificaciones push: comuníquese directamente con los usuarios (por ejemplo, a través de un correo electrónico o un SMS) cuando los servicios que utilizan se vean afectados y, nuevamente, cuando los servicios regresen a las condiciones operativas habituales. De esta forma, permitirá que tomen las medidas adecuadas.

- Comunicar su estado a través de paneles: proporcione paneles que se ajusten a sus audiencias de destino (por ejemplo, equipos técnicos internos, líderes y clientes) para comunicar el estado operativo actual de la empresa y proporcionar métricas de interés.
- Automatizar las respuestas a eventos: automatice las respuestas a eventos para reducir los errores ocasionados por los procesos manuales, así como para garantizar las respuestas rápidas y coherentes.

Evolución

OPS 11 ¿Cómo impulsa el progreso de las operaciones?

Dedique tiempo y recursos a la mejora gradual y continua a fin de desarrollar la efectividad y la eficiencia de sus operaciones.

Prácticas recomendadas:

- Cuente con un proceso para la mejora continua: evalúe y priorice regularmente las oportunidades de mejora para centrar sus esfuerzos donde estas oportunidades puedan brindar mayores beneficios.
- Ejecute análisis posteriores a los incidentes: revise los eventos que afectan a los clientes e identifique tanto los factores contribuyentes, como las acciones preventivas. Utilice esta información para desarrollar estrategias de mitigación a fin de limitar o evitar la recurrencia. Desarrolle procedimientos para ofrecer respuestas efectivas y rápidas. Comunique las acciones correctivas y los factores contribuyentes según corresponda, adaptados a las audiencias de destino.
- Implemente bucles de retroalimentación: incluya bucles de retroalimentación en los procedimientos y las cargas de trabajo que ayuden a identificar los problemas y las áreas que necesitan mejoras.
- Administre los conocimientos: existen mecanismos para que los miembros del equipo encuentren a tiempo la información que buscan, accedan a ella e identifiquen si se trata de información completa y vigente. Estos mecanismos están presentes para identificar contenido necesario, contenido que debe actualizarse y contenido que debe archivarse para que no se vuelva a utilizar como referencia.
- Defina los factores de motivación para la mejora: identifique los factores que impulsan la mejora para que lo ayuden a evaluar y priorizar las oportunidades.
- Valide los conocimientos: revise los resultados y las respuestas del análisis con equipos interdisciplinarios y propietarios de empresas. Utilice estas revisiones para fijar bases en común, identificar efectos adicionales y determinar procedimientos. Ajuste las respuestas según corresponda.
- Realice revisiones de las métricas operativas: lleve a cabo con regularidad análisis retrospectivos de las métricas operativas con participantes de distintos equipos y diferentes áreas de la empresa. Utilice estas revisiones para identificar oportunidades de mejora y posibles procedimientos, además de compartir las lecciones aprendidas.
- Documente y comparta las lecciones aprendidas: documente y comparta las lecciones aprendidas a partir de la ejecución de actividades operativas para poder usarlas de forma interna y entre todos los equipos.
- Dedique tiempo a implementar mejoras: dedique tiempo y recursos en los procesos para posibilitar mejoras continuas y graduales.

Seguridad

Temas

- Bases de seguridad (p. 55)
- Administración de identidades y accesos (p. 56)
- Detección (p. 57)

- Protección de la infraestructura (p. 58)
- Protección de los datos (p. 59)
- Respuesta ante incidentes (p. 61)

Bases de seguridad

SEC 1 ¿Cómo operar la carga de trabajo de manera segura?

A fin de operar la carga de trabajo de forma segura, debe aplicar prácticas recomendadas generales en todas las áreas de la seguridad. Tome los requisitos y los procesos que ha definido en la excelencia operativa a nivel de la organización y carga de trabajo y aplíquelos en todas las áreas. Mantenerse al día con las recomendaciones del sector y AWS y la inteligencia de amenazas facilita la evolución del modelo de amenazas y los objetivos de control. La automatización de los procesos de seguridad, las pruebas y la validación permiten escalar las operaciones de seguridad.

Prácticas recomendadas:

- Separe las cargas de trabajo mediante el uso de cuentas:organice las cargas de trabajo en cuentas individuales y en cuentas de grupos según la función o en un conjunto común de controles en lugar de imitar la estructura de generación de informes de la empresa. Comience teniendo en cuenta la seguridad y la infraestructura para permitirle a su organización establecer medidas de seguridad a medida que crezcan las cargas de trabajo.
- Proteja la cuenta de AWS: proteja el acceso a las cuentas mediante, por ejemplo, la habilitación de la MFA y el uso restringido del usuario raíz. Además, configure los contactos de la cuenta.
- Identifique y valide los objetivos de control: obtenga y valide los objetivos de control y los controles que necesita aplicar a la carga de trabajo en función de los requisitos de conformidad y los riesgos identificados en el modelo de amenazas. La validación constante de los controles y los objetivos de control facilitan la medición de la efectividad de la mitigación de riesgos.
- Manténgase al día con las amenazas de seguridad: reconozca los vectores de ataque. Para ello, manténgase al día con las amenazas de seguridad más recientes para facilitar la definición e implementación de los controles apropiados.
- Manténgase al día con las recomendaciones de seguridad: manténgase al día con las recomendaciones seguridad del sector y de AWS a fin de desarrollar la posición de seguridad de la carga de trabajo.
- Automatice las pruebas y validación de los controles de seguridad en canalizaciones: establezca plantillas y puntos de referencia seguros para los mecanismos de seguridad que sean probados y validados como parte de la creación, las canalizaciones y los procesos. Utilice herramientas y la automatización para probar y validar todos los controles de seguridad de forma continua. Por ejemplo, analice elementos, como las imágenes de máquinas y la infraestructura, como plantillas de código para detectar vulnerabilidades de seguridad, irregularidades y desviaciones con respecto al punto de referencia establecido en cada etapa.
- Identifique y priorice riesgos mediante un modelo de amenazas: utilice un modelo de amenazas para identificar y mantener un registro actualizado de posibles amenazas. Priorice las amenazas y adapte los controles de seguridad para prevenirlas, detectarlas y responder ante ellas. Revise y mantenga esto en el contexto del panorama de seguridad en evolución.
- Evalúe e implemente características y servicios de seguridad regularmente: los socios de AWS y APN lanzan nuevas características y servicios de manera constante que permiten desarrollar la postura de seguridad de la carga de trabajo.

Administración de identidades y accesos

SEC 2 ¿Cómo se administra la autenticación para las personas y las máquinas?

Hay dos tipos de identidades que necesitará administrar cuando aborde las cargas de trabajo operativas de AWS. Conocer el tipo de identidad que debe administrar y a la cual debe conceder acceso lo ayuda asegurarse de que las identidades correctas tengan acceso a los recursos correctos bajo las condiciones correctas.

Identidades humanas: los administradores, los desarrolladores, los operadores y los usuarios finales requieren una identidad para obtener acceso a los entornos y a las aplicaciones de AWS. Estos son miembros de su organización o usuarios externos con los que colabora, que interactúan con sus recursos de AWS mediante un navegador web, una aplicación cliente o herramientas interactivas de línea de comandos.

Identidades de máquinas: las aplicaciones de servicios, las herramientas operativas y las cargas de trabajo requieren una identidad para realizar solicitudes a los servicios de AWS, como, por ejemplo, para leer datos. Estas identidades incluyen máquinas que se ejecutan en su entorno de AWS, como las instancias de Amazon EC2 o las funciones de AWS Lambda. También puede administrar las identidades de máquinas para los usuarios externos que necesiten acceso. Además, también puede tener máquinas fuera de AWS que necesiten acceso a su entorno de AWS.

Prácticas recomendadas:

- Uso de mecanismos de inicio de sesión seguros: aplique contraseñas de longitud mínima e instruya a los usuarios para que eviten elegir contraseñas comunes o que ya utilizaron. Aplique la autenticación multifactor (MFA) con mecanismos de software o hardware para ofrecer una capa adicional.
- Uso de credenciales temporales: solicite a las identidades que adquieran credenciales temporales de manera dinámica. Para las identidades del personal, utilice AWS Single Sign-On o la federación con roles de IAM para acceder a las cuentas de AWS. Para las identidades de máquinas, solicite la utilización de roles de IAM en lugar de claves de acceso a largo plazo.
- Almacenamiento y uso seguro de los secretos: para las identidades del personal y de máquinas que requieran secretos, tales como contraseñas para aplicaciones de terceros, almacene estos secretos con rotación automática con los estándares más modernos del sector en un servicio especializado.
- Uso de un proveedor centralizado de identidad: para las identidades del personal, utilice un proveedor de identidad que permita administrar las identidades en un lugar centralizado. Esto le permite crear, administrar y revocar el acceso desde una sola ubicación, facilitando la administración del acceso. Esto elimina el requisito de necesitar credenciales múltiples y brinda la oportunidad de integrar los procesos de RR. HH.
- Auditoría y rotación periódica de las credenciales: cuando no pueda utilizar las credenciales temporales y requiera de credenciales a largo plazo, audite las credenciales para garantizar que los controles definidos (por ejemplo, MFA) se apliquen, roten regularmente y tengan un nivel de acceso adecuado.
- Uso de los grupos y los atributos de usuarios: ubique a los usuarios con requisitos de seguridad comunes en grupos definidos por el proveedor de identidad e implemente mecanismos para garantizar que los atributos de los usuarios que puedan ser utilizados para controlar el acceso (por ejemplo, departamento o ubicación) sean correctos y estén actualizados. Utilice estos grupos y atributos, en lugar de los usuarios individuales, para controlar el acceso. Esto le permite administrar el acceso de manera centralizada al cambiar la pertenencia a un grupo de usuarios o los atributos solo una vez, en lugar de actualizar varias políticas individuales cuando necesita cambiar el acceso de un usuario.

SEC 3 ¿Cómo administra los permisos para las personas y las máquinas?

Administre los permisos para controlar el acceso a las identidades de las personas y de las máquinas que requieran acceso a AWS y a su carga de trabajo. Los permisos controlan a qué se tiene acceso, quién puede acceder y bajo qué condiciones lo hace.

Prácticas recomendadas:

- Defina los requisitos de acceso: cada componente o recurso de la carga de trabajo debe ser accedido por los administradores, los usuarios finales u otros componentes. Se debe tener una definición clara de quién o qué debe obtener acceso a cada componente. A continuación, se debe elegir el tipo de identidad y el método de autenticación y autorización adecuados.
- Autorización de acceso con privilegios mínimos: conceda solo el acceso requerido por las identidades al permitir el acceso a determinadas acciones en ciertos recursos de AWS bajo condiciones específicas. Utilice los grupos y los atributos de identidad para establecer permisos a escala de manera dinámica, en lugar de definir los permisos para usuarios individuales. Por ejemplo, puede permitir el acceso a un grupo de desarrolladores para que solo administren los recursos de su proyecto. De esta manera, cuando se elimina del grupo a un desarrollador, se revoca su acceso a los lugares en los que el grupo tiene control de acceso, sin que se necesite algún cambio en las políticas de acceso.
- Establecimiento de un proceso de acceso de emergencia: un proceso que permita el acceso de emergencia a la carga de trabajo en el caso poco probable de que se produzca un problema de canalización o un proceso automatizado. Esto lo ayudará a utilizar privilegios mínimos para el acceso, pero asegúrese de que los usuarios puedan obtener el nivel correcto de acceso cuando lo requieran. Por ejemplo, establezca un proceso para que los administradores verifiquen y aprueben su solicitud.
- Reducción de la cantidad permisos de manera continua: a medida que los equipos y las cargas de trabajo determinen qué acceso necesitan, elimine los permisos que ya no se utilicen y establezca procesos de revisión para obtener permisos de privilegios mínimos. Monitoree y reduzca de manera continua las identidades y los permisos que no se utilicen.
- Definición de las medidas de seguridad de los permisos para su organización: establezca controles comunes que limiten el acceso a todas las identidades de la organización. Por ejemplo, puede limitar el acceso a regiones específicas de AWS o evitar que los operadores borren recursos comunes, tales como los roles de IAM utilizados por su equipo de seguridad central.
- Administre el acceso en función del ciclo de vida: integre los controles de acceso con el ciclo de vida de la aplicación y el operador, así como con el proveedor de la federación centralizada. Por ejemplo, retire el acceso de un usuario cuando abandone la organización o cambie de rol.
- Análisis del acceso público y el acceso entre cuentas: monitoree de manera continua los hallazgos que destaquen el acceso público y el acceso entre cuentas. Limite el acceso público y el acceso entre cuentas solo los recursos que lo requieran.
- Uso compartido seguro de los recursos: gobierne el consumo de los recursos compartidos en las cuentas o dentro de su organización de AWS. Monitoree los recursos compartidos y revise el acceso a ellos.

Detección

SEC 4 ¿Cómo se detectan e investigan los eventos de seguridad?

Capture y analice los eventos a partir de registros y métricas para obtener visibilidad. Tome medidas con respecto a los eventos de seguridad y las amenazas potenciales a fin de ayudar a asegurar su carga de trabajo.

Prácticas recomendadas:

- Configure el registro de servicios y aplicaciones: configure el registro en toda la carga de trabajo, incluidos los registros de aplicaciones, recursos y servicios de AWS. Por ejemplo, asegúrese de que AWS CloudTrail, Amazon CloudWatch Logs, Amazon GuardDuty y AWS Security Hub estén habilitados para todas las cuentas dentro de la organización.
- Analice los registros, hallazgos y métricas de forma centralizada: todos los registros, métricas y telemetrías se deben recopilar de forma centralizada y se deben analizar automáticamente para detectar anomalías e indicadores de actividad no autorizada. Un panel de gestión puede proporcionarle una visión de fácil acceso del estado en tiempo real. Por ejemplo, asegúrese de que los registros de Amazon GuardDuty y Security Hub se envíen a una ubicación central para alertar y analizar.
- Automatice respuestas a eventos: el uso de la automatización para investigar y remediar los eventos reduce el esfuerzo y el error humano y permite escalar las capacidades de investigación. Las revisiones periódicas lo ayudarán a ajustar las herramientas de automatización y a iterar de forma continua. Por ejemplo, automatice las respuestas a los eventos de Amazon GuardDuty mediante la automatización del primer paso de investigación y luego itere para eliminar gradualmente el esfuerzo humano.
- Implemente eventos de seguridad que se puedan accionar: cree alertas que su equipo pueda recibir y accionar. Asegúrese de que las alertas incluyan información relevante que le sirva al equipo para tomar medidas. Por ejemplo, asegúrese de que las alertas de Amazon GuardDuty y AWS Security Hub se envíen al equipo para que actúe o se envíen a las herramientas de automatización de respuesta, con el equipo aún informado por medio de mensajes del marco de automatización.

Protección de la infraestructura

SEC 5 ¿Cómo se protegen los recursos de red?
Cualquier carga de trabajo que tenga alguna forma de conectividad de red, ya sea de Internet o una red privada, requiere varios niveles de defensa para ayudar a protegerse de las amenazas externas e internas relacionadas con la red.

Prácticas recomendadas:

- Cree niveles de red: agrupe en niveles a los componentes que comparten los requisitos de accesibilidad. Por ejemplo, un clúster de bases de datos en una VPC sin necesidad de acceso a Internet debería ser colocado en subredes sin ruta hacia o desde Internet. En una carga de trabajo sin servidor que funcione sin un VPC, una segmentación y nivelación similar con microservicios puede cumplir el mismo objetivo.
- Controle el tráfico en todos los niveles: aplique los controles con un enfoque de defensa profundo, tanto para el tráfico de entrada como el de salida. Por ejemplo, para Amazon Virtual Private Cloud (VPC), esto incluye grupos de seguridad, listas de control de acceso de red y subredes. Para AWS Lambda, considere la posibilidad de ejecutar en la VPC privada con controles basados en VPC.
- Automatice la protección de la red: automatice los mecanismos de protección para proporcionar una red capaz de defenderse a sí misma, basada en la inteligencia contra amenazas y la detección de anomalías. Por ejemplo, herramientas de detección y prevención de intrusiones que se pueden adaptar de manera proactiva a las amenazas actuales y reducir su impacto.
- Implemente inspección y protección: inspeccione y filtre el tráfico en cada nivel. Por ejemplo, utilice un firewall de aplicaciones web para ayudar a protegerse contra el acceso inadvertido en el nivel de red de la aplicación. Para las funciones de Lambda, las herramientas de terceros pueden agregar un firewall de capa de aplicaciones al entorno de tiempo de ejecución.

SEC 6 ¿Cómo se protegen los recursos informáticos?

Los recursos informáticos de la carga de trabajo requieren varios niveles de defensa para facilitar la protección contra las amenazas internas y externas. Los recursos informáticos incluyen instancias de EC2, contenedores, funciones de AWS Lambda, servicios de base de datos, dispositivos de IoT y más.

Prácticas recomendadas:

- Administre las vulnerabilidades: analice y aplique parches con frecuencia para detectar las vulnerabilidades del código, las dependencias y la infraestructura a fin de facilitar la protección contra nuevas amenazas.
- Reduzca la superficie expuesta a ataques: reduzca la superficie expuesta a ataques mediante el refuerzo de los sistemas operativos, al minimizar los componentes, las bibliotecas y los servicios consumibles externos en uso.
- Implemente servicios administrados: implemente servicios que administren los recursos, como Amazon RDS, AWS Lambda y Amazon ECS, a fin de reducir las tareas de mantenimiento de la seguridad en el marco del modelo de responsabilidad compartida.
- Automatice la protección informática: automatice los mecanismos informáticos de protección, incluida la administración de vulnerabilidades, la reducción de la superficie expuesta a ataques y la administración de recursos.
- Permita que las personas realicen acciones a distancia: eliminar la capacidad de acceso interactivo reduce el riesgo de error humano y las posibilidades de configuración o administración manual. Por ejemplo, utilice un flujo de trabajo de administración de cambios para implementar las instancias de EC2 mediante el uso de infraestructura como código. Luego administre las instancias de EC2 mediante herramientas en lugar de permitir el acceso directo o un alojamiento bastión.
- Valide la integridad del software: implemente mecanismos (por ejemplo, la firma de código) para validar que el software, el código y las bibliotecas que se utilizan en la carga de trabajo provienen de fuentes confiables y no han sido manipulados.

Protección de los datos

SEC 7 ¿Cómo se clasifican los datos?

La clasificación de datos proporciona una forma de categorizar los datos en función de la criticidad y la confidencialidad, a fin de determinar los controles de protección y retención adecuados.

Prácticas recomendadas:

- Identifique los datos en el interior de la carga de trabajo: esto incluye el tipo y la clasificación de los datos, los procesos empresariales asociados, el propietario de los datos, los requisitos legales y de conformidad aplicables, el lugar de almacenamiento y los controles resultantes que se deben aplicar. Esto puede incluir clasificaciones que indiquen si los datos son de acceso público o exclusivamente de uso interno, como la información de identificación personal (PII) del cliente, así como si los datos son de acceso más restringido, como la propiedad intelectual, la información legalmente privilegiada o marcada como confidencial, entre otras categorías.
- Defina los controles de protección de datos: proteja los datos según su nivel de clasificación. Por ejemplo, asegure los datos clasificados como públicos mediante las recomendaciones pertinentes, mientras protege los datos confidenciales con controles adicionales.
- Automatice la identificación y la clasificación: automatice la identificación y clasificación de los datos para reducir el riesgo de errores humanos en las interacciones manuales.

- Defina la administración del ciclo de vida de los datos: la estrategia de ciclo de vida definida se debe basar en el nivel de confidencialidad, así como en los requisitos legales y de la organización. Se deben tener en cuenta aspectos como la duración de la retención de datos, los procesos de destrucción de los datos, la administración del acceso a los datos, la transformación y el intercambio de datos.

SEC 8 ¿Cómo se protegen los datos en reposo?

Proteja sus datos en reposo mediante la implementación de varios controles a fin de reducir el riesgo de acceso no autorizado o de manipulación indebida.

Prácticas recomendadas:

- Implemente una gestión segura de las claves: las claves de cifrado se deben almacenar de forma segura, con un estricto control de acceso. Por ejemplo, mediante el uso de un servicio de administración de claves como AWS KMS. A fin de alinear los niveles de clasificación de datos y los requisitos de segregación, considere la posibilidad de utilizar claves diferentes y el control de acceso a las claves combinado con AWS IAM y las políticas de recursos.
- Aplique el cifrado en reposo: aplique los requisitos de cifrado en función de los más recientes estándares y recomendaciones para ayudar a proteger los datos en reposo.
- Automatice la protección de datos en reposo: utilice herramientas automatizadas para validar y aplicar continuamente la protección de datos en reposo. Por ejemplo, verifique que solo haya recursos de almacenamiento cifrados.
- Aplique el control de acceso: aplique el control de acceso con privilegios mínimos y mecanismos, incluidos el aislamiento, el control de versiones y las copias de seguridad, para ayudar a proteger los datos en reposo. Considere cuáles de sus datos son de acceso público.
- Utilice mecanismos para alejar a las personas de los datos: mantenga a todos los usuarios alejados del acceso directo a los datos y los sistemas confidenciales en circunstancias operacionales normales. Por ejemplo, proporcione un panel en lugar de acceso directo a un almacén de datos para realizar consultas. En los casos en que no se utilicen canalizaciones de CI/CD, determine qué controles y procesos se requieren para proporcionar adecuadamente un mecanismo de acceso de emergencia “break-glass” normalmente desactivado.

SEC 9 ¿Cómo se protegen los datos en tránsito?

Proteja sus datos en tránsito mediante la implementación de varios controles a fin de reducir el riesgo de acceso no autorizado o pérdida.

Prácticas recomendadas:

- Implemente la gestión segura de claves y certificados: almacene los certificados y las claves de cifrado de forma segura y rótelos a intervalos de tiempo apropiados al aplicar un estricto control de acceso, p. ej., mediante el uso de un servicio de administración de certificados, como AWS Certificate Manager (ACM).
- Aplique el cifrado en tránsito: aplique los requisitos de cifrado definidos en función de los estándares y las recomendaciones pertinentes para facilitar el cumplimiento de los requisitos organizativos, legales y de conformidad.
- Automatice la detección del acceso no deseado a los datos: utilice herramientas como GuardDuty para detectar automáticamente los intentos de trasladar datos fuera de los límites definidos en función del nivel de clasificación de los datos, p. ej., para detectar un troyano que copia datos a una red desconocida o no fiable mediante el protocolo DNS.

- Autentique conexiones de red: verifique la identidad de las comunicaciones mediante el uso de protocolos que admitan la autenticación, como Transport Layer Security (TLS) o IPsec.

Respuesta ante incidentes

SEC 10 ¿Cómo se anticipa, responde y recupera de los incidentes?
La preparación es esencial para la investigación, respuesta y recuperación oportuna y efectiva de incidentes de seguridad a fin de ayudar a minimizar la interrupción en su organización.

Prácticas recomendadas:

- Identifique el personal clave y los recursos externos: identifique los recursos, el personal y las obligaciones jurídicas a nivel externo y a nivel interno que ayudarían a la organización a responder ante un incidente.
- Desarrolle planes de administración de incidentes: cree planes que lo ayuden a responder, comunicarse y recuperarse ante un incidente. Por ejemplo, puede comenzar a planificar la respuesta ante incidentes a partir de los escenarios más probables en función de la carga de trabajo y la organización. Incluya la forma en que se comunicaría durante el incidente y cómo escalaría tanto interna como externamente.
- Prepare las capacidades forenses: identifique y prepare las capacidades de investigación forense que sean adecuadas, incluidos los especialistas externos, las herramientas y la automatización.
- Automatice la capacidad de contención: automatice la contención y la recuperación ante un incidente a fin de reducir los tiempos de respuesta y el impacto en la organización.
- Aprovisione con antelación el acceso: asegúrese de que quienes responden a los incidentes tengan el acceso correcto aprovisionado con antelación en AWS a fin de reducir el tiempo que transcurre desde la investigación hasta la recuperación.
- Implemente herramientas con antelación: asegúrese de que el personal de seguridad tenga las herramientas adecuadas previamente implementadas en AWS para reducir el tiempo que transcurre desde la investigación hasta la recuperación.
- Organice los días de prueba: realice días de prueba en respuesta a incidentes (simulaciones) regularmente, incorpore las lecciones aprendidas en los planes de administración de incidentes y mejore de manera continua.

Fiabilidad

Temas

- Bases (p. 62)
- Arquitectura de las cargas de trabajo (p. 63)
- Administración de los cambios (p. 65)
- Administración de los errores (p. 67)

Bases

REL 1 ¿Cómo se administran las cuotas y las restricciones de servicio?

Para las arquitecturas de cargas de trabajo basadas en la nube, existen las cuotas de servicio (que también se denominan límites de servicio). Estas cuotas existen para evitar el aprovisionamiento accidental de más recursos de los que necesita y para limitar la tasa de solicitudes en las operaciones de la API a fin de proteger los servicios de un uso inadecuado. Además, existen restricciones de recursos, por ejemplo, la tasa con la que puede enviar bits por un cable de fibra óptica o la cantidad de almacenamiento en un disco físico.

Prácticas recomendadas:

- Conocer las cuotas y restricciones de servicio: conoce las cuotas predeterminadas y las solicitudes de aumento de cuota para la arquitectura de la carga de trabajo. También sabe qué restricciones de recursos, como el disco o la red, podrían tener un impacto.
- Administre las cuotas de servicio en todas las cuentas y regiones: si utiliza varias cuentas o regiones de AWS, asegúrese de solicitar las cuotas adecuadas en todos los entornos en los que se ejecutan las cargas de trabajo de producción.
- Adapte las cuotas y las restricciones de servicio fijas en la arquitectura: tenga en cuenta las cuotas de servicio y los recursos físicos que no se pueden cambiar y diseñe la arquitectura para evitar que afecten la fiabilidad.
- Monitoree y administre las cuotas: evalúe el uso potencial y aumente las cuotas de forma adecuada, ya que esto permitirá un crecimiento planificado en el uso.
- Automatice la administración de las cuotas: implemente herramientas que lo alerten cuando se acerque a los límites. Si utiliza las API de Service Quotas de AWS, puede automatizar las solicitudes de aumento de cuota.
- Asegúrese de que exista una brecha entre las cuotas actuales y el uso máximo que sea suficiente para adaptarse a la conmutación por error: cuando un recurso falla, todavía se puede contar de acuerdo con las cuotas hasta que se termine con éxito. Asegúrese de que sus cuotas cubran la superposición de todos los recursos con errores con reemplazos, antes de que se terminen dichos recursos. Cuando calcule esta brecha, debe considerar un error en la zona de disponibilidad.

REL 2 ¿Cómo se planifica la topología de red?

A menudo, las cargas de trabajo se encuentran en varios entornos. Entre ellos se incluyen varios entornos en la nube (de acceso público y privado) y, posiblemente, su infraestructura de centros de datos existente. Los planes deben incluir las consideraciones sobre la red, como la conectividad dentro del sistema y entre sistemas, la administración de direcciones IP públicas y privadas y la resolución de nombres de dominio.

Prácticas recomendadas:

- Utilice la conectividad de red de alta disponibilidad para sus puntos de enlace públicos de carga de trabajo: estos puntos de enlace y el direccionamiento hacia ellos deben ser de alta disponibilidad. Para lograr esto, utilice el DNS de alta disponibilidad, las redes de entrega de contenidos (CDN), API Gateway, el balanceo de cargas o los proxies inversos.
- Aprovisione conectividad redundante entre las redes privadas en la nube y los entornos en las instalaciones: utilice varias conexiones de AWS Direct Connect (DX) o túneles VPN entre redes privadas implementadas por separado. Utilice varias ubicaciones de DX para obtener alta disponibilidad. Si utiliza

varias regiones de AWS, asegúrese de tener redundancia en al menos dos de ellas. Es posible que quiera evaluar los dispositivos de AWS Marketplace que terminan las VPN. Si utiliza los dispositivos de AWS Marketplace, implemente instancias redundantes para obtener alta disponibilidad en diferentes zonas de disponibilidad.

- Garantice las cuentas de asignación de subredes IP para expansión y disponibilidad: los intervalos de direcciones IP de Amazon VPC deben ser lo suficientemente amplios como para adaptarse a los requisitos de las carga de trabajo, lo que incluye tener en cuenta futuras expansiones y asignaciones de direcciones IP a subredes en las zonas de disponibilidad. Esto incluye balanceadores de carga, instancias EC2 y aplicaciones basadas en contenedores.
- Opte por las topologías radiales, en lugar de las topologías de mallas de varios a varios: si existen más de dos espacios de direcciones de red (por ejemplo, VPC y redes en las instalaciones) conectados a través de la interconexión de VPC, AWS Direct Connect o VPN, utilice un modelo radial, como los que ofrece AWS Transit Gateway.
- Implemente intervalos de direcciones IP privadas que no se superpongan en todos los espacios de direcciones privadas, en los cuales estén conectadas: los intervalos de direcciones IP de cada VPC no deben superponerse cuando se conectan a través de una VPN. Del mismo modo, debe evitar los conflictos de direcciones IP entre la VPC y los entornos en las instalaciones o con otros proveedores en la nube que utilice. Además, debe disponer de una forma para asignar los intervalos de direcciones IP privadas cuando sea necesario.

Arquitectura de las cargas de trabajo

REL 3 ¿Cómo se diseña la arquitectura de servicios para la carga de trabajo?

Cree cargas de trabajo sumamente escalables y confiables a través de una arquitectura orientada a servicios (SOA) o una arquitectura de microservicios. La arquitectura orientada a servicios (SOA) es la práctica de crear componentes de software reutilizables a través de las interfaces de servicios. La arquitectura de microservicios ha avanzado en la creación de los componentes proporcionando componentes más pequeños y simples.

Prácticas recomendadas:

- Elija cómo segmentar su carga de trabajo: se debe evitar la arquitectura monolítica. En su lugar, debe elegir entre la SOA y los microservicios. Cuando tome cada decisión, equilibre los beneficios con las complejidades, lo que se considera correcto en el caso de un nuevo producto que compite para ser lanzado en primer lugar es diferente de lo que necesita una carga de trabajo creada para escalar desde el comienzo. Los beneficios que resultan de utilizar segmentos más pequeños incluyen escalado, mayor agilidad y flexibilidad organizacional. Por otro lado, las complejidades incluyen un posible aumento de la latencia, una depuración más compleja y una mayor carga operativa
- Desarrolle servicios centrados en funcionalidades y dominios empresariales específicos: la arquitectura orientada a servicios crea servicios con funciones bien definidas por las necesidades del negocio. Los microservicios utilizan modelos de dominio y de contexto delimitado para limitar las funciones aún más, de modo que cada servicio se encarga de satisfacer solo una necesidad. Enfocarse en la funcionalidad específica le permite diferenciar los requisitos de fiabilidad de los diferentes servicios, además de dirigir las inversiones más específicamente. Un problema empresarial conciso y contar con un equipo pequeño asociado con cada servicio también le permitirán llevar a cabo un escalado organizativo más sencillo.
- Proporcione contratos de servicios por API: los contratos de servicios son acuerdos documentados entre los equipos sobre la integración de servicios e incluyen una definición de la API legible por máquina, los límites de las tasas y las expectativas de rendimiento. Contar con una estrategia de control de versiones permite a los clientes seguir utilizando la API existente y migrar sus aplicaciones a la API más nueva cuando estén listas. La implementación se puede llevar a cabo en cualquier momento, siempre y cuando no se incumpla el contrato. El equipo del proveedor de servicios puede utilizar el componente

tecnológico que desee para cumplir con el contrato de API. Del mismo modo, el consumidor del servicio puede utilizar su propia tecnología.

REL 4 ¿Cómo se diseñan interacciones en un sistema distribuido para evitar errores?

Los sistemas distribuidos dependen de las redes de comunicación para interconectar los componentes, como servidores o servicios. A pesar de la pérdida de datos o la latencia en estas redes, su carga de trabajo debe operar de manera confiable. Los componentes del sistema distribuido deben funcionar de manera que no afecten negativamente a otros componentes o a la carga de trabajo. Las prácticas recomendadas evitan errores y mejoran el tiempo promedio entre los errores (MTBF).

Prácticas recomendadas:

- Identifique qué tipo de sistema distribuido se requiere: los sistemas distribuidos de tiempo real rígidos requieren que las respuestas se brinden de manera sincronizada y rápida, mientras que los sistemas de tiempo real flexibles disponen de una franja de tiempo en minutos mucho más amplia para proporcionar respuestas. Los sistemas sin conexión gestionan las respuestas a través del procesamiento asíncrono o por lotes. Los sistemas distribuidos de tiempo real estrictos presentan los requisitos de fiabilidad más rigurosos.
- Implemente dependencias con acoplamiento bajo: las dependencias, como los sistemas de cola, los sistemas de streaming, los flujos de trabajo y los balanceadores de carga, están acopladas en un nivel bajo. El bajo acoplamiento ayuda a aislar el comportamiento de un componente de los demás componentes que dependen de él, lo que aumenta la resistencia y la agilidad
- Proporcione respuestas idempotentes: un servicio idempotente garantiza que cada solicitud se complete exactamente una vez, de manera que hacer múltiples solicitudes idénticas tiene el mismo efecto que hacer una solicitud única. Un servicio idempotente facilita a los clientes la implementación de reintentos sin temor a que una solicitud se procese erróneamente varias veces. Para implementar los reintentos, los clientes pueden emitir solicitudes de la API con un token de idempotencia; el mismo token se utiliza cuando se repite la solicitud. Una API de servicio idempotente usa el token para generar una respuesta idéntica a la respuesta que se generó la primera vez que se completó la solicitud.
- Realice un trabajo constante: los sistemas pueden producir errores cuando hay cambios grandes y rápidos en la carga. Por ejemplo, un sistema de comprobación de estado que monitorea el estado de miles de servidores debe enviar cada vez una carga del mismo tamaño (una instantánea completa del estado actual). Aunque ningún servidor fallara o todos lo hicieran, el sistema de comprobación de estado realiza un trabajo constante sin cambios grandes ni rápidos.

REL 5 ¿Cómo se diseñan interacciones en un sistema distribuido para mitigar o tolerar errores?

Los sistemas distribuidos dependen de las redes de comunicación para interconectar los componentes (como servidores o servicios). A pesar de la pérdida de datos o la latencia sobre estas redes, su carga de trabajo debe funcionar de manera confiable. Los componentes del sistema distribuido deben funcionar de manera que no afecten negativamente a otros componentes o a la carga de trabajo. Las prácticas recomendadas permiten que las cargas de trabajo toleren errores o presiones, se recuperen más rápido de estos y mitiguen el impacto de dichas dificultades. El resultado es un mejor tiempo promedio de recuperación (MTTR).

Prácticas recomendadas:

- Implemente una degradación ordenada para transformar las dependencias estrictas aplicables en dependencias flexibles: cuando las dependencias de un componente no están en buen estado, el componente en sí puede funcionar, aunque de manera degradada. Por ejemplo, cuando una llamada de dependencia falla, se conmuta por error a una respuesta estática predeterminada.

- Limite las solicitudes: se trata de un patrón de mitigación para responder a un aumento inesperado en la demanda. Algunas solicitudes se cumplen, pero aquellas solicitudes que superan un límite definido son rechazadas y devuelven un mensaje que indica que fueron limitadas. Se espera que los clientes se retiren y abandonen la solicitud o lo intenten nuevamente a una velocidad mucho menor.
- Controle y limite las llamadas de reintento: utilice un retardo exponencial para volver a intentar después de intervalos progresivamente más largos. Introduzca la fluctuación para aleatorizar esos intervalos de reintentos y limite la cantidad máxima de reintentos.
- Implemente las notificaciones rápidas de errores y limite las colas: si la carga de trabajo no puede responder de forma correcta a una solicitud, entonces se presenta un error rápidamente. Esto permite la liberación de recursos asociados con una solicitud. Además, si se están agotando los recursos, permite al servicio recuperarse. Si la carga de trabajo puede responder correctamente, pero la tasa de solicitudes es demasiado alta, en su lugar, utilice una cola para almacenar en búfer las solicitudes. Sin embargo, no permita que se formen colas largas que lo lleven a tratar solicitudes obsoletas que el cliente ya ha desestimado.
- Establezca tiempos de espera para los clientes: establezca tiempos de espera adecuadamente, verifíquelos de manera sistemática y no confíe en los valores predeterminados, ya que, por lo general, están establecidos demasiado altos
- Cree servicios sin estado siempre que sea posible: los servicios no deberían requerir un estado o deberían descargar el estado de manera tal que, entre las solicitudes de clientes diferentes, no haya dependencia en datos almacenados localmente en un disco o una memoria. Esto permite que los servidores se reemplacen voluntariamente sin afectar la disponibilidad. Amazon ElastiCache o Amazon DynamoDB son buenos destinos para el estado descargado.
- Implemente palancas de emergencia: se trata de procesos rápidos que pueden mitigar el impacto en la disponibilidad de la carga de trabajo. Se pueden ejecutar en caso de ausencia de una causa raíz. La palanca de emergencia ideal reduce la carga cognitiva de los encargados de solucionar los problemas a cero a través de criterios totalmente deterministas de activación y desactivación. Algunos ejemplos de palancas incluyen bloquear todo el tráfico robotizado o brindar una respuesta estática. Por lo general, las palancas son manuales, pero también pueden ser automatizadas.

Administración de los cambios

REL 6 ¿Cómo se monitorean los recursos de las cargas de trabajo?

Los registros y las métricas son herramientas poderosas para obtener información sobre el estado de su carga de trabajo. Puede configurar su carga de trabajo para monitorear los registros y las métricas y enviar notificaciones cuando se superen los límites o se produzcan eventos significativos. El monitoreo permite que su carga de trabajo reconozca cuándo se superan los límites de bajo rendimiento o cuándo se producen errores, de manera que se pueda recuperar automáticamente como respuesta.

Prácticas recomendadas:

- Monitoree todos los componentes de la carga de trabajo (Generación): monitoree los componentes de la carga de trabajo con Amazon CloudWatch o herramientas de terceros. Monitoree los servicios de AWS con Personal Health Dashboard
- Defina y calcule las métricas (Agregación): almacene datos de registros y aplique filtros donde sea necesario a fin de calcular métricas, como los recuentos de un evento de registro específico o la latencia calculada a partir de las marcas de tiempo de los eventos de registros
- Envíe notificaciones (Procesamiento y activación de alarmas en tiempo real): las organizaciones que necesitan esta información reciben notificaciones cuando se producen eventos significativos
- Automatice las respuestas (Procesamiento y activación de alarmas en tiempo real): utilice la automatización para tomar las medidas necesarias cuando se detecte un evento, por ejemplo, para reemplazar los componentes que presenten errores

- Almacenamiento y análisis: recopile archivos de registro y e historiales de métricas y analícelos para encontrar tendencias más amplias e información sobre la carga de trabajo
- Realice revisiones de forma regular: revise con frecuencia el modo en que implementa el monitoreo de la carga de trabajo y actualícelo en función de los eventos y cambios significativos
- Monitorear el rastreo total de solicitudes a través de su sistema: utilice AWS X-Ray o herramientas de terceros para que los desarrolladores puedan analizar y depurar los sistemas distribuidos de manera más fácil. De esta forma, comprenderán cómo funcionan las aplicaciones y los servicios subyacentes

REL 7 ¿Cómo se diseña la carga de trabajo para que se adapte a los cambios en la demanda?

Una carga de trabajo escalable proporciona elasticidad para agregar o eliminar recursos de forma automática, de manera que coincidan estrechamente con la demanda actual en cualquier momento específico.

Prácticas recomendadas:

- Utilice la automatización cuando adquiera o escale recursos: cuando reemplace los recursos dañados o escale la carga de trabajo, automatice el proceso mediante los servicios administrados de AWS, como Amazon S3 y AWS Auto Scaling. Además, puede utilizar herramientas de terceros y los SDK de AWS para automatizar el escalado.
- Obtenga recursos cuando detecte errores en una carga de trabajo: si la disponibilidad se ve afectada, escale los recursos en forma reactiva cuando sea necesario a fin de restaurar la disponibilidad de la carga de trabajo.
- Adquiera recursos cuando detecte que una carga de trabajo necesita más recursos: escale los recursos de manera proactiva a fin de satisfacer la demanda y evitar que la disponibilidad se vea afectada.
- Realice pruebas de carga a su carga de trabajo: adopte una metodología de prueba de carga para medir si la actividad de escalado cumplirá con los requisitos de la carga de trabajo.

REL 8 ¿Cómo se implementan los cambios?

Los cambios controlados son necesarios para implementar nuevas funcionalidades y para asegurarse de que el entorno operativo, así como también las cargas de trabajo, ejecutan un software conocido, que se puede reemplazar de una manera predecible o que contiene los parches adecuados. Si no se controlan estos cambios, es más difícil predecir los efectos de estos cambios o abordar los problemas que surjan como consecuencia de ellos.

Prácticas recomendadas:

- Use manuales de procedimientos para actividades estándar como la implementación: los manuales de procedimientos son los pasos predefinidos que se utilizan para lograr resultados específicos. Utilice manuales de procedimientos para llevar a cabo actividades estándar, ya sea de forma manual o automática. Algunos ejemplos incluyen la implementación de una carga de trabajo, la implementación de parches en dicha carga o las modificaciones de DNS.
- Integre las pruebas funcionales como parte de su implementación: las pruebas funcionales se ejecutan como parte de la implementación automatizada. Si no se cumplen los criterios para el éxito, la canalización se detiene o se restaura.
- Integre las pruebas de resistencia como parte de su implementación: las pruebas de resistencia (las cuales forman parte de la ingeniería del caos) se ejecutan como parte de la canalización de implementación automatizada en un entorno de preproducción.
- Efectúe implementaciones con infraestructuras inmutables: se trata de un modelo que no exige se realicen actualizaciones, aplicaciones de parches de seguridad o cambios de configuración en el lugar

en las cargas de trabajo de producción. Cuando se necesita un cambio, la arquitectura se crea en una nueva infraestructura y se implementa en la producción.

- Implemente cambios con automatización: las implementaciones y la aplicación de parches se automatizan para eliminar el impacto negativo.

Administración de los errores

REL 9 ¿Cómo se realizan copias de seguridad de los datos?

Realice copias de seguridad de los datos, las aplicaciones y las configuraciones a fin de cumplir con los requisitos de los objetivos de tiempo de recuperación (RTO) y los objetivos de puntos de recuperación (RPO).

Prácticas recomendadas:

- Identifique todos los datos y haga una copia de seguridad de aquellos que la necesitan o reproduzca los datos desde sus orígenes: Amazon S3 se puede utilizar como destino de copia de seguridad para múltiples orígenes de datos. Los servicios de AWS como Amazon EBS, Amazon RDS y Amazon DynamoDB cuentan con capacidades integradas para crear copias de seguridad. También se puede utilizar software de copia de seguridad de terceros. Otra alternativa es que, si los datos se pueden reproducir desde otros orígenes para cumplir con los RPO, tal vez no sea necesario hacer una copia de seguridad de ellos
- Proteja y cifre las copias de seguridad: detecte el acceso mediante la autenticación y autorización, como IAM de AWS y detecte el riesgo de la integridad de los datos mediante el cifrado.
- Realice copias de seguridad de los datos de manera automática: configure las copias de seguridad, de modo que se realicen de manera automática en función de un programa periódico o debido a los cambios en el conjunto de datos. Las instancias de RDS, los volúmenes de EBS, las tablas de DynamoDB y los objetos de S3 se pueden configurar para copias de seguridad automáticas. Las soluciones de AWS Marketplace o las soluciones de terceros también se pueden utilizar.
- Realice la recuperación periódica de los datos para verificar la integridad y los procesos de la copia de seguridad: mediante una prueba de recuperación, corrobore que la implementación del proceso de copia de seguridad cumpla con los objetivos de tiempo de recuperación (RTO) y objetivos de punto de recuperación (RPO).

REL 10 ¿Cómo se utiliza el aislamiento de errores para proteger la carga de trabajo?

Los límites del aislamiento de errores restringen los efectos de un error dentro de la carga de trabajo a una cantidad limitada de componentes. Los componentes que se encuentren por fuera de los límites no se ven afectados por el error. La implementación de varios límites de aislamiento de errores le permite restringir el impacto de los errores en su carga de trabajo.

Prácticas recomendadas:

- Implemente la carga de trabajo en varias ubicaciones: distribuya los datos y los recursos de la carga de trabajo en varias zonas de disponibilidad o, cuando sea necesario, en distintas regiones de AWS. Estas ubicaciones pueden ser tan variadas como se necesite.
- Automatice la recuperación de componentes restringidos a una sola ubicación: si los componentes de una carga de trabajo solo se pueden ejecutar en una única zona de disponibilidad o en el centro de datos en las instalaciones, debe implementar la capacidad de efectuar una reconstrucción completa de la carga de trabajo dentro de los objetivos de recuperación definidos.

- Utilice arquitecturas de mamparo: al igual que los mamparos de un barco, este patrón garantiza que un error sea contenido dentro de un pequeño subconjunto de solicitudes o usuarios para que el número de solicitudes dañadas sea limitado y la mayoría de ellas pueda continuar sin errores. Los mamparos para datos generalmente se denominan particiones, mientras que los mamparos para servicios se conocen como células.

REL 11 ¿Cómo se diseña la carga de trabajo para tolerar errores de componentes?

Las cargas de trabajo que presenten requisitos de alta disponibilidad y tiempo medio de recuperación (MTTR) bajo se deben diseñar de forma que sean resistentes.

Prácticas recomendadas:

- Monitoree todos los componentes de la carga de trabajo para detectar errores: monitoree continuamente el estado de la carga de trabajo para que usted y sus sistemas automatizados estén informados de la degradación o del error total tan pronto como ocurran. Monitoree los indicadores de rendimiento clave (KPI) en función del valor de negocio.
- Conmutación por error a recursos en buen estado: asegúrese de que si se produce un error en un recurso, los recursos en buen estado puedan atender las solicitudes. En caso de errores de ubicación (como la zona de disponibilidad o la región de AWS), asegúrese de que dispone de sistemas establecidos para realizar una conmutación por error a recursos en buen estado en ubicaciones no dañadas.
- Automatización de la recuperación en todas las capas: tras la detección de un error, utilice las capacidades automatizadas para realizar acciones para corregirlo.
- Utilización de la estabilidad estática para prevenir el comportamiento bimodal: el comportamiento bimodal se produce cuando la carga de trabajo exhibe una conducta diferente en los modos normal y de error, por ejemplo, depender de lanzar nuevas instancias si se presenta un error en una zona de disponibilidad. En su lugar, debe crear cargas de trabajo que sean estáticamente estables y que funcionen en un solo modo. En este caso, aprovisione suficientes instancias en cada zona de disponibilidad para manejar la carga de la carga de trabajo si se eliminase una zona de disponibilidad y luego use las comprobaciones de estado de Elastic Load Balancing o de Amazon Route 53 para mover la carga de las instancias dañadas.
- Envío de notificaciones cuando los eventos afectan la disponibilidad: las notificaciones se envían cuando se detectan eventos importantes, incluso si el problema causado por el evento se resolvió automáticamente.

REL 12 ¿Cómo se prueba la fiabilidad?

Después de haber diseñado su carga de trabajo para que sea resistente a las presiones de la producción, las pruebas son la única forma de garantizar que funcionará como se diseñó y proporcionará la resistencia que espera.

Prácticas recomendadas:

- Utilice manuales de estrategias para investigar los errores: a través de la documentación del proceso de investigación de los manuales de estrategias, habilite respuestas consistentes y rápidas para las situaciones de errores que no se comprendan correctamente. Los manuales de estrategias contienen los pasos predefinidos que se realizan para identificar los factores que contribuyen a una situación de error. Los resultados de cualquier paso en el proceso se utilizan para determinar los próximos pasos a seguir hasta que se identifique o escale el problema.
- Ejecute análisis posteriores a los incidentes: revise los eventos que afectan a los clientes e identifique tanto los factores contribuyentes, como los elementos de acción preventiva. Utilice esta información para

desarrollar estrategias de mitigación a fin de limitar o evitar la recurrencia. Desarrolle procedimientos para ofrecer respuestas efectivas y rápidas. Comunique las acciones correctivas y los factores contribuyentes según corresponda, adaptados a las audiencias de destino. Tenga un método para comunicar estas causas a los demás según sea necesario.

- Pruebe los requisitos funcionales: estos incluyen pruebas de unidades y pruebas de integración que validan la funcionalidad requerida.
- Pruebe los requisitos de escalado y de rendimiento: esto incluye las pruebas de carga para validar que la carga de trabajo cumple con los requisitos de escalado y de rendimiento.
- Pruebe la resistencia a través de la ingeniería del caos: ejecute pruebas que inyecten errores de forma regular en los entornos de preproducción y producción. Elabore una hipótesis sobre cómo reaccionará su carga de trabajo frente al error. A continuación, compare su hipótesis con los resultados de la prueba y repita el proceso si los resultados no coinciden. Asegúrese de que las pruebas de producción no afecten a los usuarios.
- Lleve a cabo días de prueba de forma regular: utilice los días de prueba para practicar de forma regular los procedimientos de errores lo más cerca posible de la producción (incluso en los entornos de producción) con las personas que estarán involucradas en los escenarios de errores reales. Los días de prueba aplican medidas para garantizar que las pruebas de producción no tengan impacto en los usuarios.

REL 13 ¿Cómo se planifica la recuperación ante desastres (DR)?

Tener copias de seguridad y componentes de carga de trabajo redundantes en las instalaciones es el primer paso de su estrategia de recuperación de desastres (DR). Los objetivos de tiempo y punto de recuperación son los objetivos que debe cumplir para lograr la restauración de la disponibilidad. Debe establecer estos objetivos en función de las necesidades de la empresa. Implemente una estrategia para cumplir estos objetivos, teniendo en cuenta la ubicación y la función de los recursos y los datos de la carga de trabajo.

Prácticas recomendadas:

- Defina los objetivos de recuperación para el tiempo de inactividad y la pérdida de datos: la carga de trabajo tiene un objetivo de tiempo de recuperación (RTO) y un objetivo de punto de recuperación (RPO).
- Utilice estrategias de recuperación definidas para cumplir los objetivos de recuperación: se ha definido una estrategia de recuperación ante desastres (DR) para cumplir los objetivos.
- Pruebe la implementación de recuperación de desastres para validar la implementación: pruebe regularmente la conmutación por error a DR para asegurarse de que se cumplan los RTO y RPO.
- Administre la desviación de configuración en el sitio o región DR: asegúrese de que la infraestructura, los datos y la configuración se encuentren en su sitio o región DR según sea necesario. Por ejemplo, verifique que las cuotas de servicio y de AMI estén actualizadas.
- Recuperación automática: utilice AWS o herramientas de terceros para automatizar la recuperación del sistema y dirigir el tráfico al sitio o región DR.

Eficiencia de rendimiento

Temas

- Selección (p. 70)
- Revisión (p. 73)
- Monitoreo (p. 74)
- Compensaciones (p. 74)

Selección

PERF 1 ¿Cómo se selecciona la mejor arquitectura de rendimiento?

A menudo, se requieren múltiples enfoques para obtener un rendimiento óptimo en una carga de trabajo. Los sistemas de buena arquitectura utilizan múltiples soluciones y permiten diferentes características para mejorar el rendimiento.

Prácticas recomendadas:

- Comprenda los recursos y servicios disponibles: conozca y comprenda la amplia gama de servicios y recursos disponibles en la nube. Identifique los servicios relevantes y opciones de configuración para la carga de trabajo y comprenda de qué manera puede lograr un rendimiento óptimo.
- Defina un proceso para opciones de arquitectura: utilice el conocimiento y la experiencia interna de la nube o los recursos externos, como los casos de uso publicados, la documentación relevante o los documentos técnicos para definir un proceso para elegir recursos y servicios. Debe definir un proceso que promueva la experimentación y los puntos de referencia con los servicios que se pueden utilizar en la carga de trabajo.
- Gestione los requisitos de costo en las decisiones : las cargas de trabajo suelen tener requisitos de costo para las operaciones. Utilice los controles de costos internos para seleccionar los tipos y tamaños de recursos según la necesidad de recursos prevista.
- Utilice políticas o arquitecturas de referencia: maximice el rendimiento y la eficiencia mediante la evaluación de políticas internas y arquitecturas de referencia existentes y utilice su análisis para seleccionar los servicios y las configuraciones para la carga de trabajo.
- Utilice la guía del proveedor de la nube o un socio adecuado: utilice los recursos en la nube de la empresa, como arquitectos de soluciones, servicios profesionales o un socio adecuado para guiar las decisiones. Estos recursos pueden ayudar a revisar y mejorar su arquitectura para un rendimiento óptimo.
- Compare las cargas de trabajo existentes: compare el rendimiento de una carga de trabajo existente para comprender de qué manera se desempeña en la nube. Utilice los datos recopilados de los puntos de referencia para impulsar decisiones sobre arquitectura.
- Realice pruebas de carga a su carga de trabajo: implemente su última arquitectura de carga de trabajo en la nube con diferentes tipos y tamaños de recursos. Monitoree la implementación para capturar las métricas de rendimiento que identifican los cuellos de botella o los excesos de capacidad. Utilice esta información de rendimiento para diseñar o mejorar su selección de recursos y arquitectura.

PERF 2 ¿Cómo seleccionar una solución de informática?

La solución de informática óptima para una carga de trabajo específica puede variar en función del diseño de la aplicación, los patrones de uso y los ajustes de configuración. Las arquitecturas pueden utilizar diferentes soluciones de informática para varios componentes y habilitar distintas características para mejorar el rendimiento. Si se elige la solución de informática incorrecta para una arquitectura, esto puede reducir la eficiencia del rendimiento.

Prácticas recomendadas:

- Evalúe las opciones de informática disponibles: comprenda las características de rendimiento de las opciones relacionadas con la informática disponibles para usted. Conozca cómo funcionan las instancias, los contenedores y las funciones y qué ventajas o desventajas incorporan a su carga de trabajo.

- Comprenda las opciones de configuración informática disponibles: comprenda de qué manera distintas opciones complementan la carga de trabajo y qué opciones de configuración son mejores para el sistema. Los ejemplos de estas opciones incluyen familia de instancias, tamaños, características (GPU, E/S), tamaños de funciones, instancias de contenedor y tenencia única contra múltiple.
- Recopile métricas relacionadas con la informática: una de las mejores formas de comprender cómo rinden los sistemas informáticos es registrar y realizar un seguimiento del verdadero uso de diversos recursos. Estos datos se pueden utilizar para realizar determinaciones más precisas sobre los requisitos de los recursos.
- Determine la configuración necesaria mediante el dimensionamiento: analice las diversas características de rendimiento de la carga de trabajo y de qué manera se relacionan con el uso de la CPU, la red y la memoria. Utilice estos datos para elegir los recursos que mejor se adapten al perfil de su carga de trabajo. Por ejemplo, una carga de trabajo de memoria intensiva, como una base de datos, puede ser el mejor modo de alcanzar la familia de instancias r. Sin embargo, una carga de trabajo ampliada puede obtener mayores beneficios de un sistema de contenedor elástico.
- Utilice la elasticidad disponible de los recursos: la nube ofrece la flexibilidad de expandir o reducir los recursos de manera dinámica mediante una variedad de mecanismos para satisfacer los cambios en la demanda. En combinación con métricas relacionadas con la informática, una carga de trabajo puede responder a cambios de manera automática y utilizar el conjunto de recursos óptimos para lograr este objetivo.
- Reevalúe las necesidades informáticas en función de las métricas: utilice métricas a nivel del sistema para identificar las conductas y solicitudes de la carga de trabajo a lo largo del tiempo. Evalúe las necesidades de su carga de trabajo mediante la comparación de los recursos disponibles con estas solicitudes y realice cambios en su entorno informático para que coincidan mejor con el perfil de su carga de trabajo. Por ejemplo, con el tiempo se podría observar que un sistema consume más memoria de lo que se pensaba inicialmente, por lo que pasar a un tamaño o familia de instancias diferente podría mejorar tanto el rendimiento como la eficiencia.

PERF 3 ¿Cómo se selecciona una solución de almacenamiento?

La solución de almacenamiento óptimo para un sistema varía según el tipo de método de acceso (bloque, archivo u objeto), patrones de acceso (aleatorio o secuencial), rendimiento requerido, frecuencia de acceso (en línea, sin conexión, de archivo) frecuencia de actualización (WORM, dinámico) y restricciones de durabilidad y disponibilidad. Los sistemas de buena arquitectura utilizan múltiples soluciones de almacenamiento y permiten que diferentes características mejoren el rendimiento y usen los recursos de manera eficiente.

Prácticas recomendadas:

- Comprenda las características y requisitos de almacenamiento: comprenda las diferentes características (por ejemplo, compartible, tamaño del archivo, tamaño del caché, patrones de acceso, latencia, rendimiento y persistencia de los datos) que se requieren para seleccionar los servicios que mejor se adapten a la carga de trabajo, como almacenamiento de objetos, de bloques, de archivos o de instancias.
- Evalúe las opciones de configuración disponibles: evalúe las diferentes características y opciones de configuración y de qué manera se relacionan con el almacenamiento. Comprenda dónde y cómo usar IOPS provisionadas, SSD, almacenamiento magnético, almacenamiento de objetos, almacenamiento de archivos o almacenamiento efímero para optimizar el espacio de almacenamiento y el rendimiento para su carga de trabajo.
- Tome decisiones en función de métricas y patrones de acceso: elija sistemas de almacenamiento en función de los patrones de acceso de la carga de trabajo y configúrelos al establecer de qué manera la carga de trabajo accede a los datos. Elija el almacenamiento de objetos en lugar del almacenamiento en bloque para aumentar la eficiencia del almacenamiento. Configure las opciones de almacenamiento que elija para que coincidan con sus patrones de acceso a datos.

PERF 4 ¿Cómo se selecciona una solución de base de datos?

La solución de base de datos óptima para un sistema varía según los requerimientos de disponibilidad, consistencia, tolerancia en las particiones, latencia, durabilidad, escalabilidad y capacidad de consulta. Muchos sistemas utilizan soluciones de bases de datos diferentes para varios subsistemas y permiten que distintas características mejoren el rendimiento. La selección de las características y soluciones de base de datos incorrectas puede resultar en una menor eficiencia de rendimiento.

Prácticas recomendadas:

- Comprenda las características de los datos: comprenda las diferentes características de los datos en la carga de trabajo. Determine si la carga de trabajo necesita transacciones, cómo interactúa con los datos y cuáles son las demandas de rendimiento. Utilice estos datos para seleccionar el enfoque de base de datos de mejor rendimiento para su carga de trabajo (por ejemplo, bases de datos relacionales, de valor clave de NoSQL, documento, columna ancha, gráfico, serie temporal o almacenamiento en la memoria).
- Evalúe las opciones disponibles: evalúe los servicios y las opciones de almacenamiento que están disponibles como parte del proceso de selección para los mecanismos de almacenamiento de la carga de trabajo. Comprenda de qué manera y cuándo utilizar un servicio o sistema de almacenamiento de datos determinado. Aprenda sobre las opciones de configuración disponibles que pueden optimizar el rendimiento o la eficiencia de la base de datos, como las IOPS provisionadas, los recursos de memoria e informática y el almacenamiento de caché.
- Recopile y registre métricas de rendimiento de la base de datos: utilice herramientas, bibliotecas y sistemas que registren mediciones de rendimiento relacionadas con el rendimiento de la base de datos. Por ejemplo, mida las transacciones por segundo, las consultas lentas o los sistemas de latencia introducidos cuando accede a la base de datos. Utilice estos datos para comprender el rendimiento de los sistemas de su base de datos.
- Elija el almacenamiento de datos en función de los patrones de acceso: utilice los patrones de acceso de la carga de trabajo para decidir qué servicios y tecnologías utilizar. Por ejemplo, use una base de datos relacional para las cargas de trabajo que requieren transacciones o un almacén de valor clave que ofrece un rendimiento mayor, pero que finalmente sea constante donde se aplique.
- Optimice el almacenamiento de datos en función de los patrones de acceso y las métricas: utilice las características de rendimiento y los patrones de acceso que optimicen la forma en que los datos se almacenan o se consultan para lograr el mejor rendimiento posible. Mida de qué manera las optimizaciones, como el indexado, la distribución clave, el diseño de almacén de datos o las estrategias de caché, impactan en el rendimiento del sistema o la eficiencia general.

PERF 5 ¿Cómo se configura la solución de red?

La solución de red óptima para una carga de trabajo varía según la latencia, los requisitos de rendimiento, la fluctuación y el ancho de banda. Las restricciones físicas, como el usuario o los recursos en las instalaciones, determinan las opciones de ubicación. Estas restricciones se pueden compensar con ubicaciones de borde o ubicación de recurso.

Prácticas recomendadas:

- Comprenda de qué manera la red impacta en el rendimiento: analice y comprenda de qué manera las decisiones relacionadas con la red impactan en el rendimiento de la carga de trabajo. Por ejemplo, la latencia de la red suele impactar en la experiencia del usuario y, con los protocolos incorrectos, puede privar la capacidad de la red mediante gastos generales excesivos.
- Evalúe las características de red disponibles: evalúe las características de red en la nube que pueden aumentar el rendimiento. Mida el impacto de estas características mediante pruebas, métricas y análisis.

Por ejemplo, aproveche las características de nivel de red que están disponibles para reducir la latencia, la distancia de red o la fluctuación.

- Seleccione una conectividad específica de tamaño adecuado o una VPN para las cargas de trabajo híbridas: cuando exista un requisito para la comunicación en las instalaciones, asegúrese de contar con el ancho de banda adecuado para el rendimiento de la carga de trabajo. Según los requisitos de ancho de banda, una sola conexión dedicada o una única VPN puede que no sea suficiente y deba habilitar el equilibrio de carga de tráfico en varias conexiones.
- Aproveche el equilibrio de carga y la descarga cifrada: distribuya el tráfico a través de múltiples recursos o servicios para permitir que la carga de trabajo aproveche la elasticidad que ofrece la nube. También puede utilizar el equilibrio de carga para descargar la terminación de cifrado a fin de mejorar el rendimiento y administrar y direccionar el tráfico de manera efectiva.
- Seleccione protocolos de red para mejorar el rendimiento: tome decisiones sobre los protocolos para la comunicación entre sistemas y redes en función del impacto en el rendimiento de la carga de trabajo.
- Elija la ubicación de la carga de trabajo en función de los requisitos de red: utilice las opciones de ubicación en la nube disponibles para reducir la latencia de la red o mejorar el rendimiento. Utilice las regiones de AWS, las zonas de disponibilidad, los grupos de ubicación y las ubicaciones de borde, como Outposts, Local Regions (regiones locales) y Wavelength, para reducir la latencia de la red o mejorar el rendimiento.
- Optimice la configuración de la red en función de las métricas: utilice los datos recopilados y analizados para tomar decisiones fundamentadas sobre la optimización de la configuración de la red. Mida el impacto de esos cambios y utilice esas mediciones para tomar decisiones futuras.

Revisión

PERF 6 ¿Cómo se desarrolla la carga de trabajo para aprovechar los nuevos lanzamientos?

Cuando diseña las cargas de trabajo, hay una cantidad limitada de opciones entre las que puede elegir. Sin embargo, con el tiempo, las nuevas tecnologías y enfoques estarán disponibles para que pueda mejorar el rendimiento de la carga de trabajo.

Prácticas recomendadas:

- Manténgase actualizado sobre los nuevos recursos y servicios: evalúe las formas de mejorar el rendimiento a medida que estén disponibles los nuevos servicios, patrones de diseño y ofertas de productos. Determine cuál de ellos puede mejorar el rendimiento o aumentar la eficiencia de la carga de trabajo mediante la evaluación ad hoc, el debate interno o los análisis externos.
- Defina un proceso para mejorar el rendimiento de la carga de trabajo: defina un proceso para evaluar los servicios nuevos, los patrones de diseño, los tipos de recursos y las configuraciones a medida que estén disponibles. Por ejemplo, ejecute pruebas de rendimiento existentes en ofertas de instancias nuevas para determinar el potencial de mejorar la carga de trabajo.
- Permita que el rendimiento de la carga de trabajo evolucione con el paso del tiempo: como organización, utilice la información que se recopila mediante el proceso de evaluación para impulsar activamente la adopción de nuevos servicios o recursos a medida que estén disponibles.

Monitoreo

PERF 7 ¿Cómo se monitorean los recursos para garantizar que el rendimiento es óptimo?

El rendimiento del sistema se puede degradar con el tiempo. Monitoree el rendimiento del sistema para identificar la degradación y solucionar los factores internos y externos, como el sistema operativo o la carga de la aplicación.

Prácticas recomendadas:

- Registre las métricas relacionadas con el rendimiento: utilice un servicio de monitoreo y observabilidad para registrar las métricas relacionadas con el rendimiento. Por ejemplo, el registro de las transacciones de bases de datos, consultas lentas, latencia de E/S, rendimiento de solicitud HTTP, latencia de servicio u otro dato clave.
- Analice las métricas cuando ocurren eventos o incidentes: en respuesta a (o durante) un evento o incidente, utilice los paneles o informes de monitoreo para comprender y diagnosticar el impacto. Estas visualizaciones ofrecen información sobre qué partes de la carga de trabajo no funcionan como se esperaba.
- Establezca indicadores clave de rendimiento (KPI) para medir el rendimiento de la carga de trabajo: identifique los KPI que indican si la carga de trabajo tiene un rendimiento óptimo según lo previsto. Por ejemplo, una carga de trabajo basada en las API puede utilizar latencia de respuesta general como una indicación del rendimiento general y un sitio de comercio electrónico podría elegir usar el número de compras como su KPI.
- Utilice el monitoreo para generar notificaciones basadas en las alarmas: con los indicadores clave de rendimiento (KPI) relacionados con el rendimiento que ha definido, utilice un sistema de monitoreo que genere alarmas automáticamente cuando estas medidas están fuera de los límites esperados.
- Revise las métricas en intervalos regulares: como rutina de mantenimiento o en respuesta a eventos o incidentes, revise que métricas se recopilan. Utilice estas revisiones para identificar que métricas eran claves en abordar los problemas y qué métricas adicionales, si se estuviera realizando un seguimiento, ayudarían a identificar, abordar o prevenir problemas.
- Monitoree y active las alarmas de manera proactiva: utilice los indicadores clave de rendimiento (KPI), combinados con los sistemas de monitoreo y alerta, para abordar de manera proactiva los problemas relacionados con el rendimiento. Utilice alarmas para desencadenar acciones automatizadas a fin de solucionar los problemas donde sea posible. Escale la alarma a aquellos que puedan responder si no es posible una respuesta automatizada. Por ejemplo, puede tener un sistema que puede predecir los valores esperados de los indicadores clave de rendimiento (KPI) y la alarma cuando alcanzan ciertos límites o una herramienta que automáticamente puede detener o revertir las implementaciones si los KPI están fuera de los valores esperados.

Compensaciones

PERF 8 ¿Cómo se utilizan las compensaciones para mejorar el rendimiento?

Cuando diseñe soluciones, determinar las compensaciones le permite seleccionar un enfoque óptimo. A menudo, puede mejorar el rendimiento con el intercambio de la consistencia, la durabilidad y el espacio por tiempo y latencia.

Prácticas recomendadas:

- Comprenda las áreas donde el rendimiento es más crítico: comprenda e identifique las áreas donde el aumento del rendimiento de la carga de trabajo tendrá un impacto positivo en la eficiencia o la

experiencia del cliente. Por ejemplo, un sitio web que tiene una gran interacción con los clientes puede beneficiarse de utilizar servicios de borde para acercar la entrega de contenidos a los clientes.

- Aprenda sobre los servicios y los patrones de diseño: investigue y comprenda los diferentes servicios y patrones de diseño que ayuden a mejorar el rendimiento de la carga de trabajo. Como parte del análisis, identifique lo que podría intercambiar para lograr un mejor rendimiento. Por ejemplo, con un servicio de caché puede ayudar a reducir la carga en los sistemas de la base de datos; sin embargo, implementar un almacenamiento de caché seguro o una posible introducción de consistencia final en algunas áreas requiere de algo de ingeniería.
- Identifique cómo las compensaciones impactan en los clientes y en la eficiencia: cuando evalúe las mejoras relacionadas con el rendimiento, determine qué opciones impactarán en sus clientes y en la eficiencia de la carga de trabajo. Por ejemplo, si el uso de un almacén de datos de valor clave aumenta el rendimiento del sistema, es importante evaluar de qué manera la naturaleza finalmente constante de esto impactará en los clientes.
- Mida el impacto de las mejoras de rendimiento: a medida que los cambios se llevan a cabo para mejorar el rendimiento, evalúe las métricas y los datos recopilados. Utilice esta información para determinar el impacto que la mejora del rendimiento tuvo en la carga de trabajo, en los componentes de la carga de trabajo y en los clientes. Estas medidas ayudan a comprender las mejoras que resultan de las compensaciones y lo ayudan a determinar si se introdujo algún efecto secundario negativo.
- Utilice diversas estrategias relacionadas con el rendimiento: según corresponda, utilice múltiples estrategias para mejorar el rendimiento. Por ejemplo, el uso de estrategias, como el caché de datos para evitar demasiadas llamadas a la red o a la base de datos, el uso de réplicas de lectura para motores de bases de datos a fin de mejorar los índices de lectura, la partición o compresión de datos cuando sea posible para reducir volúmenes de datos y el almacenamiento en búfer y streaming de los resultados a medida que estén disponibles para evitar el bloqueo.

Optimización de costos

Temas

- Práctica de la administración financiera en la nube (p. 75)
- Concientización sobre los gastos y el uso (p. 76)
- Recursos rentables (p. 78)
- Administración de los recursos de oferta y demanda (p. 80)
- Optimización con el paso del tiempo (p. 80)

Práctica de la administración financiera en la nube

COSTOS 1 ¿Cómo se implementa la administración financiera en la nube?
La implementación de la administración financiera en la nube permite a las organizaciones comprender el valor de negocio y éxito financiero a medida que optimizan los costos, el uso y el escalado en AWS.

Prácticas recomendadas:

- Establecer una función de optimización de costos: cree un equipo que se encargue de establecer y mantener la concientización sobre los costos en toda la organización. El equipo debe contar con personas que ocupen roles financieros, tecnológicos y comerciales en la organización.
- Establezca una sociedad entre las finanzas y la tecnología: involucre a los equipos de finanzas y de tecnología en los análisis de costos y uso en todas las etapas del traspaso a la nube. Los equipos se

reúnen regularmente y analizan temas, como los objetivos y las metas de la organización, el estado actual de los costos y el uso, y las prácticas contables y financieras.

- Establezca presupuestos y predicciones de la nube: ajuste los procesos organizacionales de elaboración de presupuestos y predicciones para que sean compatibles con la naturaleza altamente variable de los costos y el uso de la nube. Los procesos deben ser dinámicos y utilizar algoritmos basados en tendencias, en impulsores empresariales o bien, la combinación de ambos.
- Implemente la concientización de costos en sus procesos organizacionales: implemente la concientización de costos en los procesos nuevos y existentes que afecten el uso. Además, aproveche los procesos existentes para este fin. Implemente la concientización de costos en la formación técnica de los trabajadores.
- Informe y notifique la optimización de costos: configure AWS Budgets de forma que proporcione notificaciones sobre los costos y el uso en relación con los objetivos. Organice reuniones con regularidad para analizar la eficiencia de los costos de esta carga de trabajo y para fomentar la cultura de concientización de costos.
- Monitoree los costos de forma proactiva: implemente herramientas y paneles para monitorear los costos de la carga de trabajo de forma proactiva. Cuando reciba las notificaciones, no solo debe prestar atención a los costos y las categorías. Esto lo ayudará a identificar las tendencias positivas y a fomentarlas en toda su organización.
- Manténgase actualizado con los lanzamientos de nuevos servicios: consulte regularmente con los expertos o los socios de APN para considerar cuáles servicios y características cuestan menos. Revise los blogs de AWS y otras fuentes de información.

Concientización sobre los gastos y el uso

COSTOS 2 ¿Cómo se controla el uso?

Establezca políticas y mecanismos a fin de asegurar que se incurra en los costos adecuados a la vez que se logran los objetivos. Mediante la aplicación del enfoque de distribución de la autoridad y la responsabilidad, puede implementar innovaciones sin gastar demasiado.

Prácticas recomendadas:

- Desarrolle políticas basadas en los requisitos de su organización: desarrolle políticas que definan cómo se administran los recursos en la organización. Las políticas deberían abordar los aspectos de los recursos y las cargas de trabajo que se relacionen con el costo, incluidos la creación, la modificación y el retiro durante la vida útil del recurso.
- Implemente los objetivos y las metas: implemente objetivos tanto de costos como de uso para la carga de trabajo. Los objetivos orientan a la organización en cuanto al uso y los costos, y las metas proporcionan resultados mensurables para sus cargas de trabajo.
- Implemente una estructura de cuentas: implemente una estructura para las cuentas que se asignen a la organización. Esto ayuda con la asignación y la administración de los costos en toda su organización.
- Implemente los grupos y los roles: implemente grupos y roles coherentes con las políticas. También controle quiénes pueden crear, modificar o retirar instancias y recursos en cada grupo. Por ejemplo, implemente grupos de desarrollo, prueba y producción. Esto se aplica tanto a los servicios de AWS como a las soluciones de terceros.
- Implemente los controles de costos: implemente controles basados en las políticas organizativas y en los roles y los grupos definidos. Estos garantizan que se incurra en los costos de acuerdo con lo definido por los requisitos de la organización, por ejemplo, el control de acceso a las regiones o los tipos de recursos con las políticas de IAM.
- Haga un seguimiento del ciclo de vida del proyecto: monitoree, mida y audite el ciclo de vida de los proyectos, los equipos y los entornos para evitar usar y pagar recursos innecesarios.

COSTOS 3 ¿Cómo se monitorean el uso y los costos?

Establezca políticas y procedimientos para monitorear y asignar de forma adecuada los costos. Esto le permite medir y mejorar los niveles de rentabilidad correspondientes a esta carga de trabajo.

Prácticas recomendadas:

- Configure fuentes de información detallada: configure el Informe de uso y costo de AWS, además de Cost Explorer con granularidad por hora, para que brinden información detallada acerca de los costos y el uso. Configure su carga de trabajo a fin de que documente entradas de registro para cada resultado empresarial entregado.
- Identifique las categorías de atribución de costos: identifique las categorías de la organización que se podrían utilizar para asignar los costos internamente.
- Establezca las métricas de la organización: establezca las métricas de la organización necesarias para esta carga de trabajo. Algunos ejemplos de las métricas de una carga de trabajo son los informes de clientes o las páginas web destinadas a los clientes.
- Configure las herramientas de facturación y administración de costos: configure AWS Cost Explorer y AWS Budgets de acuerdo con las políticas de la organización.
- Agregue información de la organización al uso y los costos: defina un esquema de etiquetado que se base en la organización, los atributos de la carga de trabajo y las categorías de asignación de costos. Implemente el etiquetado en todos los recursos. Utilice las categorías de costos para agrupar los costos y el uso de acuerdo con los atributos de la organización.
- Asigne los costos en función de las métricas de la carga de trabajo: asigne los costos de la carga de trabajo en función de las métricas o los resultados empresariales para medir la rentabilidad de la carga de trabajo. Implemente un proceso para analizar el Informe de uso y costo de AWS con Amazon Athena, lo que puede proporcionar información y capacidad de reembolso.

COST 4 ¿Cómo se retiran los recursos?

Implemente el control de cambios y la administración de recursos desde el inicio de los proyectos hasta el final de su vida útil. Esto garantizará que pueda desactivar o terminar los recursos que no utilice a fin de reducir el desperdicio.

Prácticas recomendadas:

- Realice un seguimiento de los recursos a lo largo de su vida útil: defina e implemente un método para realizar un seguimiento de los recursos y sus asociaciones a los sistemas durante su vida útil. Puede emplear el etiquetado a fin de identificar la carga de trabajo o la función del recurso.
- Implemente un proceso de retiro: implemente un proceso que identifique y retire los recursos huérfanos.
- Retire recursos: retire los recursos que se activan por eventos, como las auditorías periódicas o los cambios en el uso. Por lo general, el retiro se lleva a cabo de forma periódica y se efectúa de manera manual o automatizada.
- Retire recursos de forma automática: diseñe la carga de trabajo para que gestione con facilidad la terminación de los recursos a medida que identifica y retira los recursos que no son fundamentales, los que no son necesarios o los que tienen un bajo nivel de uso.

Recursos rentables

COSTOS 5 ¿Cómo se evalúan los costos al momento de elegir los servicios?

Amazon EC2, Amazon EBS y Amazon S3 son servicios de componentes básicos de AWS. Los servicios administrados, como Amazon RDS y Amazon DynamoDB, son servicios de AWS de mayor nivel o de nivel de aplicaciones. Si selecciona los bloques de creación y los servicios administrados adecuados, puede optimizar los costos de la carga de trabajo. Por ejemplo, si usa servicios administrados, puede reducir o eliminar una gran parte de los gastos generales administrativos y operativos, lo que le brindará la libertad para trabajar en las aplicaciones y las actividades relacionadas con el negocio.

Prácticas recomendadas:

- Identifique los requisitos de la organización para los costos: trabaje con los miembros del equipo para determinar cuándo se alcanza el equilibrio entre la optimización de costos y los demás pilares, como los de rendimiento y fiabilidad, para esta carga de trabajo.
- Analice todos los componentes de esta carga de trabajo: asegúrese de analizar cada componente de la carga de trabajo, independientemente del tamaño o los costos actuales. El esfuerzo de revisión debe reflejar el beneficio potencial, como los costos actuales y proyectados.
- Lleve a cabo un análisis exhaustivo de cada componente: observe los costos generales de cada componente para la organización. Analice los costos totales de propiedad teniendo en cuenta los costos de operaciones y administración, en especial cuando utilice servicios administrados. El esfuerzo de revisión debe reflejar el beneficio potencial; por ejemplo, el tiempo dedicado al análisis es proporcional al costo del componente.
- Seleccione software con licencias rentables: el software de código abierto eliminará los costos de licencias de software, que pueden generar costos significativos para las cargas de trabajo. Cuando sea necesario el software con licencia, evite las licencias vinculadas a atributos arbitrarios, como las CPU, y busque licencias vinculadas a los resultados o las salidas. El costo de estas licencias escala de manera más similar a los beneficios que proporcionan.
- Seleccione los componentes de esta carga de trabajo a fin de optimizar los costos en línea con las prioridades de la organización: tenga en cuenta los costos a la hora de seleccionar todos los componentes. Esto incluye el uso de servicios administrados y de nivel de aplicaciones, como Amazon RDS, Amazon DynamoDB, Amazon SNS y Amazon SES para reducir los costos generales de la organización. Utilice servicios sin servidor y contenedores para el cómputo, como AWS Lambda, Amazon S3 para los sitios web estáticos, y Amazon ECS. Minimice los costos de licencias mediante software de código abierto o software que no implique tarifas por licencias, como Amazon Linux para las cargas de trabajo de cómputo, o migre las bases de datos a Amazon Aurora.
- Lleve a cabo análisis de costos para los diferentes usos a través del tiempo: las cargas de trabajo pueden cambiar con el tiempo. Algunos servicios o características son más rentables en diferentes niveles de uso. Si efectúa análisis de cada componente a lo largo del tiempo y con el uso proyectado, se asegura de que la carga de trabajo mantenga la rentabilidad durante toda su vida útil.

COSTOS 6 ¿Cómo se cumple con los objetivos de costos al seleccionar un tipo, un tamaño y un número de recursos?

Asegúrese de elegir el tamaño de recurso y el número de recursos adecuados para la tarea en cuestión. El gasto se minimiza seleccionando el tipo, el tamaño y el número de recursos más rentables.

Prácticas recomendadas:

- Realizar el modelado de costos: identifique los requisitos de la organización y lleve a cabo el modelado de costos de la carga de trabajo y de cada uno de sus componentes. Realice actividades de

comparación para la carga de trabajo con diferentes cargas estimadas y compare los costos. El esfuerzo que implica el modelado debería reflejar el beneficio potencial; por ejemplo, que el tiempo dedicado sea proporcional al costo de los componentes.

- Seleccionar el tipo y el tamaño de recurso en función de los datos: seleccione el tamaño o el tipo de los recursos en función de los datos acerca de la carga de trabajo y de las características de los recursos, como, por ejemplo, el cómputo, la memoria, el rendimiento o el uso intensivo de la escritura. En general, esta selección se efectúa usando una versión previa de la carga de trabajo (como una versión en las instalaciones), documentación u otras fuentes de información acerca de la carga de trabajo.
- Seleccionar el tipo y el tamaño de recurso de forma automática en función de las métricas: utilice las métricas de la carga de trabajo que se ejecuta actualmente para seleccionar el tamaño y el tipo adecuados para optimizar los costos. Aprovisione de forma adecuada el rendimiento, el tamaño y el almacenamiento para servicios como Amazon EC2, Amazon DynamoDB, Amazon EBS (PIOPS), Amazon RDS, Amazon EMR y la red. Esto puede hacerse con un bucle de retroalimentación como un escalado automático o mediante código personalizado en la carga de trabajo.

COSTOS 7 ¿Cómo se utilizan los modelos de precios para reducir el costo?

Use el modelo de precios más adecuado para sus recursos con el fin de minimizar los gastos.

Prácticas recomendadas:

- Realizar análisis de modelos de precios: analice cada componente de la carga de trabajo. Determine si el componente y los recursos funcionarán durante periodos extendidos (para obtener descuentos por compromiso) o si funcionarán de manera dinámica durante lapsos cortos (para optar por modelos de spot o bajo demanda). Lleve a cabo un análisis de la carga de trabajo con la característica de recomendaciones de AWS Cost Explorer.
- Implementar regiones en función del costo: los precios de los recursos pueden ser diferentes en cada región. Considerar el costo de cada región garantiza que pague el precio total más bajo para esta carga de trabajo
- Seleccionar acuerdos con terceros con términos rentables: los términos y acuerdos rentables garantizan que el costo de estos servicios escale con los beneficios que proporcionan. Seleccione acuerdos y precios que escalen cuando le brinden beneficios adicionales a su organización.
- Implementar modelos de precios para todos los componentes de esta carga de trabajo: los recursos que permanentemente se encuentran en ejecución deben utilizar capacidad reservada como los Savings Plans o las instancias reservadas. La capacidad a corto plazo se configura para usar instancias de spot o una flota de spot. El modelo bajo demanda solo se usa para cargas de trabajo a corto plazo que no pueden interrumpirse y no funcionan durante un tiempo suficiente como para utilizar la capacidad reservada, entre un 25 % y un 75 % del periodo, según el tipo de recurso.
- Realizar análisis de modelos de precios al nivel de la cuenta maestra: utilice las recomendaciones de Cost Explorer Savings Plans e instancias reservadas para llevar a cabo análisis regulares al nivel de la cuenta maestra para obtener descuentos por compromisos.

COSTOS 8 ¿Cómo se planean los cargos por transferencia de datos?

Asegúrese de planear y monitorear los cargos por transferencia de datos para poder tomar decisiones sobre arquitectura con el fin de minimizar los costos. Un pequeño pero efectivo cambio en la arquitectura puede reducir radicalmente sus costos operativos con el paso del tiempo.

Prácticas recomendadas:

- Realizar un modelado de transferencia de datos: reúna los requisitos de la organización y lleve a cabo el modelado de la transferencia de datos de la carga de trabajo y cada uno de sus componentes. Esto identifica el punto de costo más bajo para los requisitos de transferencia de datos actuales.
- Seleccionar componentes para optimizar los costos de transferencia de datos: todos los componentes se seleccionan y la arquitectura se diseña para reducir los costos de transferencia de datos. Esto incluye usar componentes como la optimización de WAN y las configuraciones Multi-AZ.
- Implementar servicios para reducir costos de transferencia de datos: implemente servicios para reducir los costos de las transferencias de datos, como, por ejemplo, usar una red de entrega de contenido como Amazon CloudFront para entregar contenido a los usuarios finales, capas de almacenamiento en caché con Amazon ElastiCache, o usar AWS Direct Connect en lugar de la VPN para la conectividad con AWS.

Administración de los recursos de oferta y demanda

COSTOS 9 ¿Cómo se administran los recursos de la oferta y demanda?

Para una carga de trabajo que tiene gastos y rendimiento equilibrados, asegúrese de que se use todo lo que pague y evite significativamente las instancias subutilizadas. Una métrica de utilización manipulada en cualquier dirección tiene un impacto adverso en su organización, ya sea en los costos operativos (rendimiento degradado debido a la sobreutilización) o los gastos de AWS desperdiciados (debido al sobreaprovisionamiento).

Prácticas recomendadas:

- Realizar un análisis de la demanda de la carga de trabajo: analice la demanda de la carga de trabajo con el paso del tiempo. Asegúrese de que el análisis cubra las tendencias estacionales y represente de manera precisa las condiciones operativas durante toda la vida útil de la carga de trabajo. El esfuerzo de análisis debe reflejar el beneficio potencial; por ejemplo, el tiempo dedicado es proporcional al costo de los componentes.
- Implemente un búfer o una limitación controlada para administrar la demanda: el almacenamiento en búfer y la limitación controlada modifican la demanda de la carga de trabajo, lo que atenúa los picos. Implemente una limitación controlada cuando sus clientes lleven a cabo reintentos. Implemente el almacenamiento en búfer para almacenar la solicitud y postergar el procesamiento para más adelante. Asegúrese de que sus limitaciones y búferes estén diseñados de manera que los clientes reciban una respuesta en el tiempo requerido.
- Suministrar los recursos de manera dinámica: los recursos se aprovisionan de manera planeada. Esto se puede hacer en función de la demanda, como a través del escalado automático, o en función del tiempo, donde la demanda es predecible y los recursos se suministran en función del tiempo. Estos métodos generan la cantidad menor de aprovisionamiento excesivo e insuficiente.

Optimización con el paso del tiempo

COSTOS 10 ¿Cómo se evalúan los nuevos servicios?

A medida que AWS lanza nuevos servicios y características, una práctica recomendada es revisar las decisiones sobre la arquitectura existente para garantizar que siguen siendo la opción más rentable.

Prácticas recomendadas:

- Desarrolle un proceso de revisión de la carga de trabajo: desarrolle un proceso que defina los criterios y los pasos para revisar la carga de trabajo. Los esfuerzos de revisión deben reflejar el beneficio potencial, por ejemplo, las cargas de trabajo principales o las cargas de trabajo cuyo valor represente más del 10 % de la facturación se deben revisar cada tres meses, mientras que las cargas de trabajo que representan menos del 10 % se deben revisar una vez por año.
- Revise y analice la carga de trabajo regularmente: las cargas de trabajo existentes se revisan regularmente en función de procesos definidos.

Avisos

Los clientes son responsables de hacer su propia evaluación independiente de la información en este documento. Este documento: (a) solo tiene fines informativos, (b) representa las prácticas y las ofertas de productos de AWS actuales, las cuales están sujetas a cambios sin aviso previo, y (c) no crea compromisos ni promesas de parte de AWS y sus empresas afiliadas, proveedores o licenciantes. Los servicios o los productos de AWS se ofrecen “como son”, sin garantías, declaraciones ni condiciones de ningún tipo, ya sean expresas o implícitas. Las responsabilidades y obligaciones de AWS frente a sus clientes se rigen por los acuerdos celebrados con AWS, y este documento no forma parte de ningún acuerdo entre AWS y sus clientes, ni lo modifica.