AWS Well-Architected Framework

Pilar de optimización de costos

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Pilar de optimización de costos: AWS Well-Architected Framework

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

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

Table of Contents

Resumen e introducción 1
Resumen 1
Introducción 1
Optimización de costos 3
Principios de diseño 3
Definición 4
Práctica de administración financiera en la nube 5
COST01-BP01: establecimiento de la responsabilidad de la optimización de costos 8
Guía para la implementación 8
Recursos 10
COST01-BP02 Establecimiento de la colaboración entre los departamentos de finanzas y tecnología 11
Guía para la implementación 8
Recursos 10
COST01-BP03 Establecimiento de presupuestos y previsiones de la nube 16
Guía para la implementación 8
Recursos 10
COST01-BP04 Implementación de conciencia de costos en los procesos organizativos 20
Guía para la implementación 8
Recursos 10
COST01-BP05 Creación de informes y notificaciones sobre la optimización de costos 23
Guía para la implementación 8
Recursos 10
COST01-BP06 Supervisión proactiva de los costos 25
Guía para la implementación 8
Recursos 10
COST01-BP07 Seguimiento de la información sobre las nuevas versiones de los servicios 27
Guía para la implementación 8
Recursos 10
COST01-BP08 Creación de una cultura sensibilizada con los costos 29
Guía para la implementación 8
Recursos 10
COST01-BP09: cuantificación del valor empresarial de la optimización de costos 31
Guía para la implementación 8

Recursos 10
Conocimiento del gasto y del uso 34
Gobernanza 34
COST02-BP01 Desarrollo de políticas basadas en los requisitos de su organización 35
COST02-BP02 Implementación de objetivos y metas 39
COST02-BP03 Implementación de una estructura de cuentas 45
COST02-BP04 Implementación de grupos y roles 49
COST02-BP05 Implementación de controles de costos 51
COST02-BP06 Seguimiento del ciclo de vida de los proyectos 54
Supervisión del costo y el uso 56
COST03-BP01 Configuración de los orígenes de información detallados 57
COST03-BP02 Incorporación de información de la organización a los costos y el uso 59
COST03-BP03 Identificación de las categorías de atribución de costos 62
COST03-BP04 Establecimiento de métricas de la organización 66
COST03-BP05 Configuración de herramientas de administración de facturación y costos 67
COST03-BP06 Asignación de costos según las métricas de la carga de trabajo 71
Retirada de recursos 72
COST04-BP01 Seguimiento de los recursos a lo largo de su ciclo de vida 72
COST04-BP02 Implementación de un proceso de retirada 74
COST04-BP03 Retirada de recursos 77
COST04-BP04 Retirada automática de los recursos 78
COST04-BP05 Aplicación de políticas de retención de datos 79
Recursos rentables 81
Evaluación del costo al seleccionar los servicios 81
COST05-BP01 Identificación de los requisitos de la organización en relación con el costo 81
COST05-BP02 Análisis de todos los componentes de la carga de trabajo 83
COST05-BP03 Ejecución de un análisis exhaustivo de cada componente 86
COST05-BP04 Selección de software con licencias rentables 89
COST05-BP05 Selección de los componentes de la carga de trabajo para optimizar los costos de acuerdo con las prioridades de la organización 90
COST05-BP06 Análisis de costos para diferentes usos a lo largo del tiempo 93
Selección del tipo, tamaño y número correctos de recursos 94
COST06-BP01 Modelado de costos 95
COST06-BP02 Selección del tipo, tamaño y número de recursos en función de los datos 97
COST06-BP03 Selección automática del tipo, tamaño y número de recursos en función de las métricas 100

COST06-BP04 Consideración del uso de los recursos compartidos 103
Selección del mejor modelo de precios 106
COST07-BP01 Análisis de los modelos de precios 112
COST07-BP02 Elección de regiones según el costo 115
COST07-BP03 Selección de acuerdos de terceros con condiciones rentables 117
COST07-BP04 Implementación de modelos de precios para todos los componentes de la carga de trabajo 119
COST07-BP05 Análisis de modelos de precios en el nivel de la cuenta de administración ... 121
Planificación para la transferencia de datos 124
COST08-BP01 Modelado de transferencia de datos 124
COST08-BP02 Selección de componentes para optimizar el costo de la transferencia de datos 127
COST08-BP03 Implementación de servicios para reducir los costos de transferencia de datos 129
Administración de la demanda y suministro de recursos 132
COST09-BP01 Análisis de la demanda de la carga de trabajo 132
Guía para la implementación 8
Recursos 10
COST09-BP02 Implementación de un búfer o una limitación para administrar la demanda 135
Guía para la implementación 8
Recursos 10
COST09-BP03 Suministro dinámico de recursos 138
Guía para la implementación 8
Pasos para la implementación 9
Recursos 10
Optimización a lo largo del tiempo 146
Definición de un proceso de revisión y análisis periódico de la carga de trabajo 146
COST10-BP01 Desarrollo de un proceso de revisión de la carga de trabajo 146
COST10-BP02 Revisión y análisis regulares de esta carga de trabajo 149
Automatización de operaciones 151
COST11-BP01 Automatización de las operaciones 151
Conclusión 156
Colaboradores 157
Documentación adicional 158
Revisiones del documento 159
Avisos 162

Glosario de AWS 163

Pilar de optimización de costos: Marco de AWS Well-Architected

Fecha de publicación: 27 de junio de 2024 (Revisiones del documento)

Resumen

Este documento técnico se centra en el pilar de optimización de costos del Marco de Amazon Web Services (AWS) Well-Architected. Proporciona una guía para ayudar a los clientes a aplicar las prácticas recomendadas a la hora de diseñar, entregar y mantener los entornos de AWS.

Una carga de trabajo con costos optimizados utiliza plenamente todos los recursos, consigue un resultado al menor precio posible y satisface sus requisitos funcionales. En este documento técnico se ofrece una guía exhaustiva para generar capacidad en su organización, diseñar su carga de trabajo, seleccionar sus servicios, configurar y operar los servicios y aplicar técnicas de optimización de costos.

Introducción

El Marco de AWS Well-Architected le ayuda a comprender las decisiones que toma al crear cargas de trabajo en AWS. El marco proporciona las prácticas recomendadas de arquitectura para diseñar y operar cargas de trabajo fiables, seguras, eficientes, rentables y sostenibles en la nube. Demuestra una forma de medir sus arquitecturas de manera constante en función de las prácticas recomendadas y de identificar las áreas de mejora. Creemos que contar con cargas de trabajo de Well-Architected aumenta en gran medida la probabilidad de éxito empresarial.

El marco se basa en seis pilares:

- Excelencia operativa
- Seguridad
- Fiabilidad
- Eficacia del rendimiento
- Optimización de costes
- Sostenibilidad

Este documento se centra en el pilar de optimización de costos y en cómo diseñar cargas de trabajo con el uso más eficaz de los servicios y los recursos, para conseguir resultados empresariales al menor precio posible.

Aprenderá a aplicar las prácticas recomendadas del pilar de optimización de costos en su organización. La optimización de costos puede ser un desafío en las soluciones en las instalaciones tradicionales, porque hay que predecir la capacidad futura y las necesidades empresariales mientras se llevan a cabo procesos de adquisición complejos. La adopción de las prácticas de este documento ayudará a su organización a alcanzar los siguientes objetivos:

- Práctica de administración financiera en la nube
- Conocimiento del gasto y del uso
- Recursos rentables
- Administración de la demanda y suministro de recursos
- Optimización a lo largo del tiempo

Este documento está dirigido a quienes desempeñan roles tecnológicos y financieros, como directores de tecnología (CTO), directores financieros (CFO), arquitectos, desarrolladores, controladores financieros, planificadores financieros, analistas empresariales y miembros del equipo de operaciones. En este documento no se ofrecen detalles de implementación ni patrones arquitectónicos, pero sí se incluyen referencias a los recursos adecuados.

Optimización de costos

La optimización de costos es un proceso continuo de refinamiento y mejora durante el ciclo de vida de una carga de trabajo. Las prácticas de este documento lo ayudan a crear y operar cargas de trabajo sensibilizadas con los costos que logren resultados empresariales y, al mismo tiempo, minimicen los costos y permitan a su organización maximizar el retorno de la inversión.

Temas

- Principios de diseño
- Definición

Principios de diseño

Tenga en cuenta los siguientes principios de diseño de optimización de costos:

Implementación de la administración financiera en la nube: para alcanzar el éxito financiero y acelerar la materialización del valor empresarial en la nube, es necesario invertir en la administración financiera de la nube. Su organización debe dedicar el tiempo y los recursos necesarios para desarrollar capacidades en este nuevo ámbito de la tecnología y de administración del uso. De forma similar a su capacidad de seguridad u operaciones, debe desarrollar capacidades a través de la creación de conocimientos, programas, recursos y procesos que lo ayuden a convertirse en una organización rentable.

Adopte un modelo de consumo: pague solo por los recursos de computación que utilice y aumente o reduzca el uso según los requisitos del negocio. Por ejemplo, los entornos de desarrollo y pruebas se utilizan normalmente solo ocho horas al día durante la semana laboral. Puede detener estos recursos cuando no se utilicen para generar un ahorro potencial de costos del 75 % (40 horas frente a 168 horas).

Evalúe la eficacia global: mida el rendimiento empresarial de la carga de trabajo y los costos asociados a su entrega. Use estos datos para comprender las ganancias que obtiene al aumentar la producción, aumentar la funcionalidad y reducir los costos.

Fin del gasto de dinero en tareas pesadas poco diferenciadas: AWS se encarga del trabajo pesado de las operaciones del centro de datos, como el montaje de los servidores en bastidores, su apilamiento y su alimentación. También elimina la carga operativa de administrar sistemas operativos

y aplicaciones con servicios administrados. De este modo, podrá centrarse en sus clientes y proyectos empresariales en lugar de hacerlo en la infraestructura de TI.

Analice y atribuya los gastos: la nube facilita la identificación precisa del uso y el costo de las carga de trabajo, lo que permite atribuir de forma transparente los costos de TI a fuentes de ingresos y propietarios de cargas de trabajo individuales. Esto lo ayuda a medir el retorno de la inversión (ROI) y da a los propietarios de cargas de trabajo la oportunidad de optimizar sus recursos y reducir costos.

Definición

Existen cinco áreas de enfoque para la optimización de costos en la nube:

- Práctica de administración financiera en la nube
- Conocimiento del gasto y del uso
- Recursos rentables
- Administración de la demanda y suministro de recursos
- Optimización a lo largo del tiempo

Al igual que en los otros pilares del Marco de Well-Architected, hay que tener en cuenta las ventajas y desventajas de la optimización de costos. Por ejemplo, si optimizar la velocidad de comercialización o el costo. En algunos casos, es mejor optimizar la velocidad (salida rápida al mercado, envío de nuevas características o simplemente el cumplimiento de una fecha límite), en lugar de la inversión en la optimización de costos iniciales.

Las decisiones de diseño a veces se guían por la prisa en lugar de basarse en los datos y siempre existe la tentación de sobrecompensar en lugar de dedicar tiempo para establecer un punto de referencia de la implementación más rentable. La sobrecompensación puede dar lugar a implementaciones con un aprovisionamiento excesivo y menos optimizado. Sin embargo, puede ser una opción razonable si debe “migrar mediante lift-and-shift” recursos desde su entorno en las instalaciones hacia la nube y optimizarlos más adelante.

La inversión del esfuerzo adecuado en una estrategia de optimización de costos por adelantado le permite obtener los beneficios económicos de la nube con mayor facilidad, al garantizar una adhesión constante a las prácticas recomendadas y evitar un aprovisionamiento excesivo innecesario. En las siguientes secciones se proporcionan 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.

Práctica de administración financiera en la nube

La administración financiera en la nube requiere que los procesos financieros existentes evolucionen con el fin de establecer y operar con transparencia de costos, control, planificación y optimización para los entornos de AWS.

La aplicación de modelos tradicionales y estáticos de planificación en cascada, presupuestación de TI y evaluación de costos al uso dinámico en la nube puede crear riesgos, provocar una planificación inexacta y conllevar una menor visibilidad. En última instancia, esto da como resultado una oportunidad perdida para optimizar y controlar eficazmente los costos y obtener valor empresarial a largo plazo. Para evitar estos inconvenientes, administre activamente los costos a lo largo del traspaso a la nube, tanto si crea aplicaciones de forma nativa en la nube, como si migra sus cargas de trabajo a la nube o amplía su adopción de servicios en la nube.

La administración financiera en la nube (CFM) permite a las organizaciones financieras, de productos, tecnológicas y empresariales administrar, optimizar y planificar los costos a medida que aumentan su uso y escala en AWS. El objetivo principal de la CFM es permitir a los clientes alcanzar sus resultados empresariales de forma más rentable y acelerar la generación de valor empresarial y económico, al mismo tiempo que se encuentra el equilibrio adecuado entre agilidad y control.

Las soluciones de CFM ayudan a transformar su empresa mediante la transparencia, el control, la previsión y la optimización de costos. Estas soluciones también pueden permitir una cultura consciente de los costos que impulsa la rendición de cuentas en todos los equipos y funciones. Los equipos de Finanzas pueden ver de dónde vienen los costos, ejecutar operaciones con un mínimo de gastos inesperados, planificar el uso dinámico de la nube y ahorrar en gastos en la nube mientras los equipos escalan sus adopciones en la nube. Compartir esto con los equipos de Ingeniería puede proporcionar el contexto financiero necesario para la selección, el uso y la optimización de sus recursos.

La CFM de AWS ofrece un conjunto de capacidades para administrar, optimizar y planificar los costos en la nube a la vez que se mantiene la agilidad empresarial. La CFM es primordial no solo para administrar eficazmente los costos, sino también para verificar que las inversiones impulsen los resultados comerciales esperados. Estos son los cuatro pilares del marco de administración financiera en la nube de la Nube de AWS: ver, ahorrar, planificar y ejecutar. Cada uno de estos pilares tiene un conjunto de actividades y capacidades.

Los cuatro pilares de la administración financiera en la nube.

- Ver: ¿Cómo mide y supervisa actualmente su gasto en la nube y cómo crea responsabilidad al respecto? Si es la primera vez que usa AWS o tiene previsto utilizar AWS, ¿cuenta con un plan para establecer la visibilidad del costo y uso?

  Para comprender los costos de AWS y optimizar el gasto, debe saber de dónde provienen esos costos. Esto requiere una estructura deliberada de sus cuentas y recursos que ayude a su organización financiera a hacer un seguimiento de los flujos de gastos y responsabilizar a los equipos de su parte del resultado final.

  Servicios de AWS: AWS Control Tower, AWS Organizations, etiquetas de asignaciones de costos, políticas de etiquetas, Grupos de recursos de AWS, categorías de costos de AWS, AWS Cost Explorer, AWS Cost and Usage Report, instancias reservadas y Savings Plans

  Recursos: prácticas recomendadas de etiquetado de AWS, categorías de costos de AWS

- Ahorrar: ¿Qué herramientas de optimización de costos utiliza actualmente para optimizar los gastos? Si no utiliza AWS, ¿está familiarizado con las optimizaciones comunes basadas en el uso y en el modelo de precios?

  En el principio de ahorro, optimizamos los costos con recomendaciones sobre precios y recursos. Optimizar los costos comienza por tener una estrategia bien definida para su nuevo modelo operativo en la nube. Idealmente, esto debería comenzar lo antes posible en su traspaso a la nube, sentando las bases para una cultura consciente de los costos reforzada por los procesos y comportamientos correctos.

Existen muchas maneras diferentes de optimizar los costos en la nube. Una de ellas es seleccionar el modelo de compra adecuado (instancias reservadas y Savings Plans) o si su carga de trabajo es inmutable y está en contenedores para poder adoptar instancias de spot de Amazon EC2. Además, escale la carga de trabajo mediante grupos de Amazon EC2 Auto Scaling.

Servicios de AWS: instancias reservadas y Savings Plans, grupos de Amazon EC2 Auto Scaling, instancias de spot

Recursos: instancias reservadas, Savings Plans,, prácticas recomendadas para gestionar Amazon EC2

- Planificar: ¿Cómo planifica actualmente el uso y el gasto futuros de la nube? ¿Dispone de una metodología para cuantificar la generación de valor de una nueva migración? ¿Ha evolucionado sus procesos actuales de presupuestación y previsión para adoptar el uso variable de la nube?

  El principio de planificación significa mejorar su planificación con presupuestos y previsiones flexibles. Una vez que haya establecido la visibilidad y los controles de costos, es probable que desee planificar y establecer expectativas de gasto en proyectos en la nube. AWS ofrece la flexibilidad de crear procesos dinámicos de previsión y presupuestos para que pueda mantenerse informado sobre si los costos cumplen, o exceden, los límites presupuestarios.

  Servicios de AWS: AWS Cost Explorer, AWS Cost and Usage Report, AWS Budgets

  Recursos: previsión basada en el uso, alertas e informes de presupuestos de AWS

- Ejecutar: ¿Cuáles son algunos de los procesos y herramientas operativos que utiliza actualmente para administrar sus gastos en la nube y quién dirige esos esfuerzos? ¿Ha pensado en cómo funcionarán las cosas desde el punto de vista de las operaciones diarias una vez que comience a utilizar AWS?

  En realidad, el principio de ejecución consiste en administrar la facturación y el control de costos. Puede definir barreras de protección y establecer controles para garantizar que los gastos se ajusten a los presupuestos. AWS proporciona varias herramientas para ayudarlo a empezar.

  Servicios de AWS: consola de Administración de facturación y costos de AWS, AWS Identity and Access Management, políticas de control de servicios (SCP), AWS Service Catalog, AWS Cost Anomaly Detection, AWS Budgets

  Recursos: introducción a la consola de AWS Billing

Estas son las prácticas recomendadas para la administración financiera en la nube:

Prácticas recomendadas

- COST01-BP01: establecimiento de la responsabilidad de la optimización de costos
- COST01-BP02 Establecimiento de la colaboración entre los departamentos de finanzas y tecnología
- COST01-BP03 Establecimiento de presupuestos y previsiones de la nube
- COST01-BP04 Implementación de conciencia de costos en los procesos organizativos
- COST01-BP05 Creación de informes y notificaciones sobre la optimización de costos
- COST01-BP06 Supervisión proactiva de los costos
- COST01-BP07 Seguimiento de la información sobre las nuevas versiones de los servicios
- COST01-BP08 Creación de una cultura sensibilizada con los costos
- COST01-BP09: cuantificación del valor empresarial de la optimización de costos

COST01-BP01: establecimiento de la responsabilidad de la optimización de costos

Cree un equipo (Oficina de negocios en la nube, Centro de excelencia en la nube o equipo de FinOps) que se encargue de establecer y afianzar el concienciación sobre los costos en toda la organización. El responsable de la optimización de costos puede ser una persona o un equipo (requiere representantes de los equipos financieros, tecnológicos y empresariales) que comprenda toda la organización y las finanzas en la nube.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Esta es la introducción a un equipo de Oficina de negocios en la nube (CBO) o Centro de excelencia en la nube (CCOE) que se encargue de establecer y afianzar una cultura de concienciación sobre los costos en la computación en la nube. Esta función puede ser una persona existente, un equipo de la organización o un nuevo equipo formado por partes interesadas clave de los equipos de Finanzas, Tecnología y Organización de la organización.

Esta función (la persona o el equipo) prioriza y dedica el porcentaje necesario de su tiempo a actividades de administración y optimización de costos. En una organización pequeña, es posible

que esta función dedique menos tiempo a ello, si lo comparamos con una función a tiempo completo de una empresa grande.

Esta función debe tener carácter multidisciplinar, es decir, debe tener experiencia en administración de proyectos, ciencia de datos, análisis financiero y desarrollo de software o infraestructura. Para mejorar la eficiencia de la carga de trabajo, puede ejecutar optimizaciones de costos dentro de tres propietarios diferentes:

- Centralizado: a través de equipos designados, como el equipo de FinOps, el equipo de Administración financiera en la nube (CFM), la oficina de negocios en la nube (CBO) o el Centro de excelencia en la nube (CCoE), los clientes pueden diseñar e implementar mecanismos de gobernanza e impulsar las prácticas recomendadas en toda la empresa.
- Descentralizado: se influye en los equipos de Tecnología para que haga optimizaciones de costos.
- Híbrido: una combinación de equipos centralizados y descentralizados que pueden trabajar de forma conjunta para optimizar los costos.

La función se evalúa según su capacidad de ejecutar y alcanzar los objetivos de optimización de costos (por ejemplo, las métricas de eficiencia de las cargas de trabajo).

Debe conseguir el patrocinio de los ejecutivos para esta función, lo cual es un factor clave para el éxito. El patrocinador es considerado el campeón del consumo rentable de la nube y proporciona apoyo al equipo para garantizar que las actividades de optimización de costos se traten según el nivel de prioridad definido por la organización. De lo contrario, se podrán ignorar las directrices y no se dará prioridad a las oportunidades de ahorro. De forma conjunta, el patrocinador y el equipo garantizan que su organización haga un consumo eficiente de la nube y ofrezca valor empresarial.

Si tiene el plan de asistencia Business, Enterprise-On-Ramp o Enterprise y necesita ayuda para crear este equipo o función, contacte con los expertos de administración financiera en la nube (CFM) a través de su equipo de Cuentas.

Pasos para la implementación

- Definición de los miembros clave: todas las partes pertinentes de la organización deben contribuir y estar interesadas en la administración de costos. En general, los equipos de las organizaciones constan de equipos de Finanzas, propietarios de aplicaciones o productos, y equipos administrativos y técnicos (DevOps). Algunos tienen dedicación completa (técnicos y financieros) mientras que otros participan periódicamente, según sea necesario. Las personas o los equipos encargadas de la CFM necesitan el siguiente conjunto de habilidades:

  - Desarrollo de software: en el caso de que se creen scripts y automatización.
  - Conocimientos de ingeniería de infraestructuras: para implementar scripts, automatizar procesos y entender cómo se aprovisionan los servicios o recursos.
  - Perspicacia en las operaciones: la CFM consiste en operar en la nube de forma eficiente midiendo, supervisando, modificando, planificando y escalando el uso eficiente de la nube.
- Definición de los objetivos y las métricas: esta función debe proporcionar valor a la organización de distintas maneras. Estos objetivos se definen y evolucionan de forma continua a medida que evoluciona la organización. Estas son las actividades habituales: crear y ejecutar programas educativos sobre optimización de costos en la organización, desarrollar estándares para toda la organización (como la supervisión y la creación de informes de optimización de costos) y establecer objetivos de carga de trabajo sobre la optimización. Esta función también debe informar regularmente a la organización sobre la capacidad de optimizar costos de la organización.

  Puede definir indicadores clave de rendimiento (KPI) basados en el valor o el costo. Cuando se definen los KPI, se puede calcular el costo previsto en términos de eficiencia y el resultado empresarial esperado. Los KPI basados en el valor vinculan las métricas de costo y uso a los impulsores del valor empresarial y ayudan a racionalizar los cambios en el gasto de AWS. El primer paso para derivar los KPI basados en el valor es trabajar juntos, entre organizaciones, para seleccionar y acordar un conjunto estándar de KPI.
- Establecimiento de una cadencia periódica: el grupo (equipos de Finanzas, Tecnología y Negocios) debe reunirse de manera regular para revisar sus objetivos y métricas. Una cadencia típica implica revisar el estado de la organización, los programas que se ejecutan actualmente y las métricas generales financieras y de optimización. Después, se debe informar sobre las cargas de trabajo clave con mayor detalle.

  Durante estas revisiones periódicas, se puede revisar la eficiencia de la carga de trabajo (costo) y los resultados empresariales. Por ejemplo, un aumento del 20 % en el costo de una carga de trabajo puede coincidir con un mayor uso por parte del cliente. En este caso, este aumento del 20 % de los costos puede interpretarse como una inversión. Estas reuniones de cadencia periódica pueden ayudar a los equipos a identificar los KPI de valor que proporcionan significado a toda la organización.

Recursos

Documentos relacionados:

- Blog de CCoE de AWS
- Creating Cloud Business Office
- CCOE - Cloud Center of Excellence

Videos relacionados:

- Vanguard CCOE Success Story

Ejemplos relacionados:

- Using a Cloud Center of Excellence (CCOE) to Transform the Entire Enterprise
- Building a CCOE to transform the entire enterprise
- 7 Pitfalls to Avoid When Building CCOE

COST01-BP02 Establecimiento de la colaboración entre los departamentos de finanzas y tecnología

Debe implicar a los equipos de finanzas y tecnología en las conversaciones sobre costos y uso en todas las etapas del traspaso a la nube. Los equipos deben reunirse y tratar regularmente temas como los objetivos organizativos, el estado actual de los costos y el uso, y las prácticas contables y financieras.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Los equipos tecnológicos innovan más rápido en la nube gracias a que los ciclos de aprobación, adquisición e implementación de la infraestructura son más cortos. Esto puede implicar un ajuste para las organizaciones financieras acostumbradas a tanto a lanzar procesos que requieren mucho tiempo y consumen muchos recursos para obtener e implementar capital en entornos de centros de datos y en las instalaciones, como a asignar los costos solo al aprobar el proyecto.

Desde el punto de vista de las organizaciones financieras y de adquisiciones, el proceso de presupuestos de capital, las solicitudes de capital, las aprobaciones, las adquisiciones y la instalación de la infraestructura física es un proceso que se ha aprendido y estandarizado durante décadas:

- Los equipos de Ingeniería o TI suelen ser los solicitantes.
- Varios equipos de Finanzas actúan como aprobadores y compradores.
- Los equipos de Operaciones montan, apilan y entregan la infraestructura lista para usar.

Con la adopción de la nube, la adquisición y el consumo de infraestructura dejan de estar a expensas de una cadena de dependencias. En el modelo de la nube, los equipos de tecnología y productos ya no son solo creadores, sino operadores y propietarios de sus productos, responsables de la mayoría de las actividades históricamente asociadas a los equipos de finanzas y operaciones, incluidas las adquisiciones y la implementación.

Todo lo que se necesita para aprovisionar recursos en la nube es una cuenta y el conjunto de permisos adecuado. Esto es también lo que reduce el riesgo de la TI y de las finanzas, lo cual significa que, con unos pocos clics o llamadas a la API, los equipos pueden eliminar los recursos inactivos o innecesarios en la nube. Esto también permite a los equipos de tecnología innovar más rápidamente gracias a la agilidad y la capacidad de poner en marcha experimentos y luego desmantelarlos. Aunque la naturaleza variable del consumo de la nube puede afectar a la previsibilidad desde el punto de vista de los presupuestos y las previsiones de capital, la nube ofrece a las organizaciones la posibilidad de reducir el costo del exceso de aprovisionamiento, así como el costo de oportunidad asociado al subaprovisionamiento conservador.

Establezca una asociación entre los partes interesadas clave de Finanzas y Tecnología para lograr un entendimiento común de los objetivos organizativos y desarrollar mecanismos para obtener éxito financiero en el modelo de gasto variable de la computación en la nube. Los equipos pertinentes de su organización deben estar presentes en las conversaciones sobre costos y uso en todas las etapas del traspaso a la nube, incluidos los siguientes:

- Líderes de finanzas: los directores financieros, responsables financieros, planificadores financieros, analistas empresariales y responsables de adquisición, abastecimiento y cuentas por pagar deben entender el modelo de consumo de la nube, las opciones de compra y el proceso de facturación mensual. El departamento de Finanzas debe asociarse con los equipos de Tecnología para crear y compartir una historia de valor de TI y, así, ayudar a los equipos de Negocios a comprender cómo el gasto en tecnología está vinculado a los resultados empresariales. De esta manera, los gastos en tecnología no se consideran costos, sino inversiones. Dado que hay diferencias fundamentales entre la nube (por ejemplo, la velocidad del cambio en el uso, los precios de pago por uso, los precios por niveles, los modelos de precios y la información detallada sobre la facturación y el uso) y las operaciones en las instalaciones, resulta esencial que el equipo de Finanzas comprenda de qué manera puede afectar el uso de la nube a aspectos empresariales como los procesos de adquisición, el seguimiento de incentivos, la asignación de costos y los estados financieros.
- Líderes de tecnología: los líderes de tecnología (incluidos los propietarios de aplicaciones y productos) deben conocer los requisitos financieros (por ejemplo, las limitaciones presupuestarias) y los requisitos empresariales (por ejemplo, los acuerdos de nivel de servicio). Esto permite que la carga de trabajo se implemente para lograr los objetivos empresariales deseados.

La asociación entre los departamentos de Finanzas y Tecnología aporta los siguientes beneficios:

- Los equipos de finanzas y tecnología tienen visibilidad casi en tiempo real de los costos y el uso.
- Los equipos de finanzas y tecnología establecen un procedimiento operativo estándar para gestionar la variación del gasto en la nube.
- Las partes interesadas de Finanzas actúan como asesores estratégicos en cuanto a cómo se utiliza el capital para comprar descuentos por compromiso de compra (por ejemplo, instancias reservadas o Savings Plans de AWS) y cómo se utiliza la nube para hacer crecer la organización.
- Las cuentas por pagar y los procesos de adquisición existentes también se usan con la nube.
- Los equipos de Finanzas y Tecnología colaboran a la hora de prever los costos y el uso de AWS en el futuro para adaptar y diseñar los presupuestos organizativos.
- Mejor comunicación dentro de la organización al compartir el mismo lenguaje y tener un conocimiento común de los conceptos financieros.

Otras partes interesadas de su organización que deberían estar implicadas en las discusiones sobre costos y uso son:

- Propietarios de unidades de negocio: los propietarios de unidades de negocio deben comprender el modelo de negocio de la nube para poder establecer directrices para las unidades de negocio y la empresa entera. Este conocimiento de la nube resulta esencial para llevar a cabo previsiones de crecimiento y de uso de las cargas de trabajo, pero también para valorar diferentes opciones de compra a más largo plazo (como instancias reservadas o Savings Plans).
- Equipo de Ingeniería: establecer una asociación entre los equipos de Finanzas y Tecnología es esencial para crear una cultura sensibilizada con los costos que anime a los ingenieros a actuar en la administración financiera en la nube (CFM). Uno de los problemas habituales de los profesionales de la CFM o las operaciones financieras y de los equipos de Finanzas es conseguir que los ingenieros entiendan todo el negocio en la nube, sigan las prácticas recomendadas y adopten las medidas recomendadas.
- Terceros: si en su organización participan terceros (por ejemplo, consultores o herramientas), asegúrese de que también sigan sus objetivos empresariales y que lo demuestren a través de sus modelos de compromiso y el retorno de la inversión (ROI). Por lo general, los terceros contribuyen a la generación de informes y al análisis de las cargas de trabajo que administran y también aportan análisis de costos de cualquier carga de trabajo que diseñan.

Para implementar la CFM y tener éxito, se necesita colaboración entre los equipos de Finanzas, Tecnología y Negocios y un cambio en la forma en que se comunica y evalúa el gasto en la nube en toda la organización. Incluya a los equipos de Ingeniería para que puedan formar parte de estas conversaciones sobre costos y uso en todas las etapas y anímelos a seguir las prácticas recomendadas y a adoptar las medidas acordadas de forma apropiada.

Pasos para la implementación

- Definición de los miembros clave: compruebe que todos los miembros pertinentes de sus equipos de Finanzas y Tecnología participen en la asociación. Los miembros de Finanzas pertinentes serán aquellos que interactúen con la factura de la nube. Suelen ser los directores financieros, responsables financieros, planificadores financieros, analistas empresariales, responsables de adquisiciones y responsables de abastecimiento. Los miembros de Tecnología suelen ser los propietarios de las aplicaciones y los productos y los gerentes y representantes técnicos de todos los equipos que crean en la nube. Otros miembros pueden ser los propietarios de las unidades de negocio, como el departamento de Marketing, que influyen en el uso de los productos, y terceros como consultores, para ajustarse a los objetivos y mecanismos y para asistir en la generación de informes.

- Definición de los temas de análisis: defina los temas comunes de todos los equipos o que requieran una compresión compartida. Haga un seguimiento del costo desde el momento en que se genera hasta que se paga la factura. Tome nota de todos los miembros implicados y de los procesos organizativos que deben aplicarse. Comprenda cada paso o proceso por el que pasa y la información asociada, como los modelos de precios disponibles, los precios por niveles, los modelos de descuento, la creación de presupuestos y los requisitos financieros.
- Establecimiento de una cadencia periódica: para crear una asociación entre los equipos de Finanzas y Tecnología, establezca una cadencia de comunicación periódica para crear y mantener la coherencia. El grupo debe reunirse de forma periódica para tratar sus objetivos y métricas. Una cadencia típica implica revisar el estado de la organización, los programas que se ejecutan actualmente y las métricas generales financieras y de optimización. Después, se debe informar sobre las cargas de trabajo clave con mayor detalle.

Recursos

Documentos relacionados:

- AWS Blog de noticias de

COST01-BP03 Establecimiento de presupuestos y previsiones de la nube

Ajuste los procesos de presupuestos y previsión de la organización para que sean compatibles con la naturaleza altamente variable de los costos y el uso de la nube. Los procesos deben ser dinámicos y usar algoritmos basados en tendencias o el motor principal de la empresa, o en una combinación de ambos.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

En las configuraciones tradicionales de TI en las instalaciones, los clientes suelen enfrentarse al reto que supone la planificación de unos costos fijos que solo cambian en ocasiones, normalmente con la compra de nuevos servicios y hardware de TI para satisfacer los picos de demanda. En cambio, Nube de AWS adopta un enfoque diferente, en el que los clientes pagan por los recursos que utilizan de acuerdo con sus necesidades empresariales y de TI reales. En el entorno de la nube, la demanda puede fluctuar cada mes, cada día o incluso cada hora.

El uso de la nube aporta eficacia, rapidez y agilidad, lo que se traduce en un patrón de costo y uso muy variable. Los costos pueden disminuir o a veces aumentar en respuesta a la mayor eficacia de las cargas de trabajo o a la implementación de nuevas cargas de trabajo y características. Cuando las cargas de trabajo se escalan para atender a una base de clientes en expansión, el uso y los costos de la nube aumentan en consecuencia debido a la mayor accesibilidad de los recursos. Esta flexibilidad de los servicios en la nube se extiende a los costos y las previsiones, lo que da lugar a un cierto grado de elasticidad.

Es esencial ajustarse muy bien a estos impulsores de la demanda y necesidades empresariales tan cambiantes e intentar que la planificación sea lo más precisa posible. Los procesos presupuestarios tradicionales de la organización deben adaptarse a esta variabilidad.

Considere la posibilidad de crear modelos de costos a la hora de pronosticar el costo de las nuevas cargas de trabajo. Los modelos de costos le permiten disponer de información de referencia sobre los costos previstos de la nube, lo que lo ayuda a llevar a cabo análisis sobre el costo total de propiedad (TCO), el retorno de la inversión (ROI) y otros análisis financieros, fijar objetivos y expectativas con las partes interesadas e identificar oportunidades para optimizar los costos.

Su organización debe conocer las definiciones de costos y las agrupaciones aceptadas. El nivel de detalle de las previsiones puede variar en función de la estructura y los flujos de trabajo internos de su organización. Seleccione una granularidad que se adapte a sus requisitos específicos y a la configuración de su organización. Es importante entender a qué nivel se hace la previsión:

- AWS Organizations o cuenta de administración: la cuenta de administración es la que usa para crear AWS Organizations. De forma predeterminada, las organizaciones tienen una cuenta de administración.
- Cuenta de miembro o vinculada: una cuenta de Organizations es una Cuenta de AWS estándar que contiene sus recursos de AWS y las identidades que pueden acceder a esos recursos.
- Entorno: un entorno es un conjunto de recursos de AWS que ejecutan una versión de la aplicación. Se puede crear un entorno con varias cuentas vinculadas o de miembro.
- Proyecto: un proyecto es una combinación de objetivos o tareas establecidos que deben lograrse en un periodo fijo. Es importante tener en cuenta el ciclo de vida del proyecto durante la previsión.
- Servicios de AWS: grupos o categorías, como servicios de computación o almacenamiento, en los que puede agrupar servicios de AWS para su previsión.
- Agrupación personalizada: puede crear grupos personalizados en función de las necesidades de su organización, como unidades de negocio, centros de costos, equipos, etiquetas de asignación de costos, categorías de costos, cuentas vinculadas o una combinación de estas necesidades.

Identifique los impulsores empresariales que pueden repercutir en sus costos de uso y haga previsiones para cada uno de ellos por separado para calcular el uso esperado con antelación. Algunos de los impulsores podrían estar relacionados con los equipos de TI y de Productos de la organización. Los líderes de ventas, marketing y negocios conocen otros impulsores empresariales, como los eventos de marketing, las promociones, las expansiones geográficas, las fusiones y las adquisiciones, por lo que es importante colaborar con ellos y tenerlos en cuenta también.

Puede usar AWS Cost Explorer para previsiones basadas en tendencias en un intervalo de tiempo futuro definido en función de su gasto anterior. El motor de previsión de AWS Cost Explorer segmenta los datos históricos en función de los tipos de cargo (por ejemplo, instancias reservadas) y utiliza una combinación de machine learning y modelos basados en reglas para predecir el gasto en todos los tipos de cargo individualmente.

Una vez que haya establecido el proceso de previsión y creado los modelos, puede usar AWS Budgets para establecer presupuestos personalizados de forma pormenorizada especificando el periodo de tiempo, la periodicidad o el importe (fijo o variable) y agregando filtros como el servicio, la Región de AWS y las etiquetas. Por lo general, el presupuesto se prepara para un solo año y no cambia, por lo que todas las partes involucradas deben ajustarse a él estrictamente. Por el contrario, las previsiones son más flexibles, ya que permiten reajustes a lo largo del año y proporcionan proyecciones dinámicas durante un periodo de uno, dos o tres años. La elaboración de presupuestos y de previsiones desempeña un papel crucial a la hora de establecer las expectativas financieras entre las diversas partes interesadas de los equipos de Tecnología y Negocios. La precisión de las previsiones y de su implementación también impone responsabilidades a las partes interesadas, que son directamente responsables de los costos de aprovisionamiento en primer lugar, y también puede aumentar su concienciación general de los costos.

Para mantenerse informado sobre el cumplimiento de los presupuestos existentes, puede crear y programar el envío regular de informes de AWS Budgets por correo electrónico a usted y a otras partes interesadas. También puede crear alertas de AWS Budgets basadas en los costos reales, que son de naturaleza reactiva, o en los costos previstos, lo que permite tener tiempo para implementar mitigaciones contra posibles sobrecostos. Puede recibir alertas cuando el costo o el uso reales superen un nivel determinado o si se prevé que superen el importe presupuestado.

Ajuste los procesos de elaboración de presupuestos y previsiones existentes para que sean más dinámicos, ya sea utilizando un algoritmo basado en las tendencias (que usa los costos históricos como entradas) o algoritmos basados en impulsores (por ejemplo, lanzamientos de productos nuevos, una expansión regional o nuevos entornos para cargas de trabajo), ya que son ideales para un entorno de gasto dinámico y variable. Una vez que haya determinado su previsión basada en las

tendencias con el Explorador de costos o cualquier otra herramienta, use Calculadora de precios de AWS para estimar su caso de uso de AWS y los costos futuros en función del uso esperado (tráfico, solicitudes por segundo o instancias de Amazon EC2 necesarias).

Haga un seguimiento de la precisión de esa previsión, ya que los presupuestos deben establecerse en función de estos cálculos y estimaciones de la previsión. Supervise la precisión y la eficacia de las previsiones de costos de la nube integrados. Compare periódicamente los gastos reales con su previsión y haga los ajustes necesarios para mejorar la precisión de la previsión. Haga un seguimiento de la desviación de las previsiones y lleve a cabo un análisis de la causa raíz de la desviación notificada para tomar medidas y ajustar las previsiones.

Como se indica en COST01-BP02 Establecimiento de la colaboración entre los departamentos de finanzas y tecnología, es importante contar con asociaciones y cadencias entre los departamentos de TI y Finanzas y otras partes interesadas para garantizar que todos utilicen las mismas herramientas o procesos en aras de la coherencia. En los casos en que los presupuestos deban cambiar, aumente los puntos de contacto regulares para reaccionar a esos cambios más rápidamente.

Pasos para la implementación

- Defina el lenguaje de costos de la organización: cree un lenguaje común de costos de AWS dentro de la organización con múltiples dimensiones y agrupaciones. Asegúrese de que las partes interesadas comprenden la granularidad de las previsiones, los modelos de precios y el nivel de sus previsiones de costos.
- Analice las previsiones basadas en las tendencias: utilice herramientas de previsión basadas en tendencias, como AWS Cost Explorer y Amazon Forecast. Analice su costo de uso en múltiples dimensiones (por ejemplo, servicio, cuentas, etiquetas y categorías de costos).
- Análisis de las previsiones basadas en impulsores: identifique el efecto de los impulsores empresariales en el uso de la nube y haga previsiones para cada uno de ellos por separado para calcular el costo de uso esperado con antelación. Trabaje en estrecha colaboración con los propietarios de las unidades de negocio y las partes interesadas para conocer la repercusión en los nuevos impulsores y calcular los cambios de costos esperados a la hora de definir presupuestos precisos.
- Actualice los procesos existentes de elaboración de presupuestos y previsiones: en función de los métodos de previsión adoptados, como los basados en tendencias, los basados en impulsores empresariales o una combinación, defina sus procesos de elaboración de presupuestos y previsiones. Los presupuestos deben estar bien calculados, ser realistas y basarse en sus previsiones.

- Configure alertas y notificaciones: utilice las alertas de AWS Budgets y la detección de anomalías de costos para recibir alertas y notificaciones.
- Revisiones periódicas con las principales partes interesadas: por ejemplo, las partes interesadas en TI, finanzas, equipos de plataforma y otras áreas de la empresa deben alinearse con los cambios en la dirección y el uso de la empresa.

Recursos

Documentos relacionados:

- AWS Cost Explorer
- AWS Cost and Usage Report
- Forecasting with Cost Explorer
- Previsiones de Quick
- AWS Budgets

Videos relacionados:

- How can I use AWS Budgets to track my spending and usage
- AWS Cost Optimization Series: AWS Budgets

Ejemplos relacionados:

- Understand and build driver-based forecasting
- How to establish and drive a forecasting culture
- How to improve your cloud cost forecasting
- Using the right tools for your cloud cost forecasting

COST01-BP04 Implementación de conciencia de costos en los procesos organizativos

Implemente la conciencia de costos, cree transparencia y responsabilidad de los costos en los procesos nuevos y existentes que afecten al uso, y aproveche los procesos existentes para tomar conciencia de los costos. Implemente la conciencia de costos en la capacitación del personal.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

La conciencia de costos debe implementarse en los procesos organizativos nuevos y existentes. Es una de las capacidades fundamentales, requisito previo para otras prácticas recomendadas. Se recomienda reutilizar y modificar los procesos existentes cuando sea posible, lo que minimiza el impacto en la agilidad y la velocidad. Informe de los costos de la nube a los equipos de Tecnología y a los responsables de la toma de decisiones de los equipos de Negocios y Finanzas para concienciar sobre los costos, y establezca indicadores clave de rendimiento (KPI) de eficiencia para las partes interesadas de finanzas y negocios. Las siguientes recomendaciones ayudarán a implementar la conciencia de costos en su carga de trabajo:

- Verifique que la administración de los cambios incluya la medición de los costos para cuantificar el efecto financiero de sus cambios. Esto ayuda a abordar de forma proactiva las preocupaciones relacionadas con los costos y destacar el ahorro de costos.
- Verifique que la optimización de costos sea un componente central de sus capacidades operativas. Por ejemplo, puede aprovechar los procesos existentes de administración de incidentes para investigar e identificar las causas raíz de las anomalías de los costos y el uso o costos excesivos.
- Acelere el ahorro de costos y la materialización del valor de negocio a través de la automatización o las herramientas. Al pensar en el costo de implementación, enmarque la conversación para que incluya un componente de retorno de la inversión (ROI) para justificar la inversión de tiempo o dinero.
- Asigne los costos de la nube mediante la aplicación de devoluciones o reembolsos de los gastos en la nube, incluidos los gastos en las opciones de compra basadas en el compromiso, los servicios compartidos y las compras en el mercado para impulsar el consumo de la nube teniendo siempre presentes los costos.
- Amplíe los programas de capacitación y desarrollo existentes para que incluyan la conciencia de costos en toda la organización. Se recomienda incluir capacitación y certificaciones continuas. Con ello logrará tener una organización capaz de autoadministrar los costos y el uso.
- Aproveche las herramientas nativas gratuitas de AWS como AWS Cost Anomaly Detection, AWS Budgets e informes de AWS Budgets.

Cuando las organizaciones adoptan sistemáticamente prácticas de administración financiera en la nube (CFM), esos comportamientos se arraigan en la forma de trabajar y la toma de decisiones. El resultado es una cultura que tiene más en cuenta los costos, desde los desarrolladores que diseñan

una nueva aplicación nacida en la nube hasta los administradores financieros que analizan el retorno de estas nuevas inversiones en la nube.

Pasos para la implementación

- Identifique los procesos organizativos relevantes: cada unidad organizativa debe revisar sus procesos e identificar los procesos que afecten a los costos y el uso. Cualquier proceso que conlleve la creación o terminación de un recurso debe incluirse en la revisión. Debe buscar procesos que ayuden a tomar conciencia de los costos en su negocio, como la administración de incidentes y la capacitación.
- Establezca una cultura autosuficiente en materia de costos: asegúrese de que todas las partes interesadas pertinentes se alineen con la causa del cambio y el impacto como costo para que entiendan el costo de la nube. Esto permitirá a su organización establecer una cultura de innovación autosuficiente y sensibilizada con los costos.
- Actualice los procesos teniendo en cuenta los costos: cada proceso se modifica para adaptarse a los costos. El proceso puede requerir controles previos adicionales, como evaluar el impacto del costo, o controles posteriores que validen que se han producido los cambios esperados en el costo y el uso. La asistencia a procesos tales como la capacitación y la administración de incidentes puede ampliarse para incluir elementos de costo y uso.

Para obtener ayuda, contacte con los expertos de CFM a través de su equipo de Cuentas o explore los recursos y documentos relacionados a continuación.

Recursos

Documentos relacionados:

- Administración financiera en la nube de AWS

Ejemplos relacionados:

- Strategy for Efficient Cloud Cost Management
- Cost Control Blog Series #3: How to Handle Cost Shock
- A Beginner’s Guide to AWS Cost Management

COST01-BP05 Creación de informes y notificaciones sobre la optimización de costos

Establezca presupuestos para la nube y configure mecanismos para detectar anomalías en el uso. Configure las herramientas relacionadas para que proporcionen alertas de costo y uso respecto a objetivos predefinidos y reciba notificaciones cuando el uso supere esos objetivos. Organice reuniones periódicas para analizar la rentabilidad de las cargas de trabajo y promover la conciencia de costos.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

Debe informar regularmente sobre la optimización de costos y el uso en su organización. Puede implementar sesiones dedicadas a examinar el rendimiento de costos o incluir la optimización de costos en los ciclos de preparación de informes operativos habituales para sus cargas de trabajo. Utilice los servicios y las herramientas para supervisar el rendimiento de los costos con regularidad e implementar oportunidades de ahorro de costos.

Consulte sus costo y uso con varios filtros y grado de detalle con AWS Cost Explorer, que proporciona paneles e informes, como costos por servicio o por cuenta, costos diarios o costos del marketplace. Haga un seguimiento de la evolución del costo y el uso según los presupuestos creados con AWS Budgets Reports.

Utilice AWS Budgets para establecer presupuestos personalizados para hacer un seguimiento de sus costos y uso, y responder rápidamente a las alertas recibidas por correo electrónico o las notificaciones del Amazon Simple Notification Service (Amazon SNS) si supera su límite. Establezca su período presupuestario preferido para que sea diario, mensual, trimestral o anual, y cree límites presupuestarios específicos para mantenerse informado sobre cómo los costos y el uso reales o previstos avanzan hacia el límite de su presupuesto. También puede configurar alertas y acciones con respecto a esas alertas para que se ejecuten automáticamente o mediante un proceso de aprobación cuando se supere un objetivo presupuestario.

Implemente notificaciones sobre el costo y el uso para garantizar que los cambios en el costo y el uso se puedan corregir rápidamente en caso de que sean inesperados. AWS Cost Anomaly Detection le permite reducir los costos imprevistos y mejorar el control sin ralentizar la innovación. AWS Cost Anomaly Detection identifica los gastos anómalos y las causas raíz, lo que ayuda a

reducir el riesgo de sorpresas en la facturación. Con tres sencillos pasos, puede crear su propio monitor contextualizado y recibir alertas cuando se detecte cualquier gasto anómalo.

También puede usar Quick con datos de AWS Cost and Usage Report (CUR) para proporcionar informes altamente personalizados con datos más detallados. Quick le permite programar informes y recibir correos electrónicos de informes de costos periódicos para un historial de costos y uso u oportunidades de ahorro de costos. Consulte nuestra solución Cost Intelligence Dashboard (CID) basada en Quick, que ofrece visibilidad avanzada.

Use AWS Trusted Advisor, que proporciona orientación para verificar si los recursos aprovisionados están en consonancia con las prácticas recomendadas de AWS para la optimización de costos.

Compare sus recomendaciones de Savings Plans a través de gráficos visuales con sus costos y uso detallados. En los gráficos por hora se muestra el gasto bajo demanda junto con el compromiso de Savings Plans recomendado, lo que proporciona información sobre los ahorros, la cobertura de Savings Plans y la utilización de Savings Plans estimados. Esto ayuda a las organizaciones a comprender cómo sus Savings Plans se aplican a cada hora de gasto sin tener que invertir tiempo y recursos en la creación de modelos para analizar el gasto.

Crear periódicamente informes que contengan un resumen de los Savings Plans, las instancias reservadas y recomendaciones de tamaño adecuado de Amazon EC2 desde AWS Cost Explorer para empezar a reducir el costo asociado a las cargas de trabajo en estado estable, los recursos inactivos y los infrautilizados. Identifique y recupere el gasto asociado a los residuos de la nube para los recursos que se implementan. Los residuos de la nube se producen cuando se crean recursos de tamaño incorrecto o se observan patrones de uso diferentes a los esperados. Siga las prácticas recomendadas de AWS para reducir los residuos o pida a su socio y equipo de Cuentas que lo ayuden a optimizar y ahorrar en relación con los costos de la nube.

Genere informes con regularidad para mejorar las opciones de compra de sus recursos y reducir los costos unitarios de sus cargas de trabajo. Las opciones de compra, como los Savings Plans, las instancias reservadas o las instancias de spot de Amazon EC2, ofrecen el mayor ahorro de costos para las cargas de trabajo con tolerancia a errores y permiten a las partes interesadas (propietarios de la empresa y equipos financieros y técnicos) formar parte de estas conversaciones de compromiso.

Comparta los informes que contengan oportunidades o anuncios de nuevas versiones que puedan ayudarlo a reducir el costo total de propiedad (TCO) de la nube. Adopte nuevos servicios, regiones, características, soluciones o nuevas formas de lograr una mayor reducción de costos.

Pasos para la implementación

- Configuración de AWS Budgets: configure AWS Budgets en todas las cuentas de su carga de trabajo. Establezca un presupuesto para el gasto general de la cuenta y un presupuesto para la carga de trabajo con etiquetas.
  - Well-Architected Labs: Cost and Governance Usage
- Informe de la optimización de costos: defina un ciclo habitual para tratar y analizar la eficiencia de la carga de trabajo. Utilice las métricas establecidas y notifique las métricas alcanzadas y el costo para alcanzarlas. Identifique y corrija las tendencias negativas, así como las tendencias positivas que puede promover en su organización. La preparación de informes debe incluir a representantes de los equipos y propietarios de las aplicaciones, de las finanzas y los principales responsables de la toma de decisiones con respecto al gasto en la nube.

Recursos

Documentos relacionados:

- AWS Cost Explorer
- AWS Trusted Advisor
- AWS Budgets
- AWS Cost and Usage Report
- AWS Budgets Prácticas recomendadas de
- Análisis de Amazon S3

Ejemplos relacionados:

- Key ways to start optimizing your AWS cloud costs

COST01-BP06 Supervisión proactiva de los costos

Implemente herramientas y paneles para supervisar los costos de la carga de trabajo de forma proactiva. Revise periódicamente los costos con herramientas configuradas o listas para usar, no se limite a mirar los costos y las categorías cuando reciba las notificaciones. Supervisar y analizar los costos de forma proactiva ayuda a identificar las tendencias positivas y permite promoverlas en toda la organización.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

Se recomienda supervisar los costos y el uso de forma proactiva en la organización, no solo cuando surjan anomalías o excepciones. Disponer de unos paneles muy visibles en la oficina o en el entorno de trabajo permite a las personas clave tener acceso a la información que necesitan y transmite la idea de que la organización se centra en la optimización de costos. Los paneles visibles permiten promover de forma activa los resultados de éxito e implementarlos en toda la organización.

Cree una rutina diaria o frecuente para utilizar AWS Cost Explorer o cualquier otro panel, como Amazon Quick, para ver los costos y analizarlos de forma proactiva. Analice el uso y los costos de los servicios de AWS por cuenta de AWS, carga de trabajo o servicio específico de AWS con agrupaciones y filtros, y valide si son los esperados o no. Utilice la granularidad por hora y recurso y las etiquetas para filtrar e identificar los costos incurridos de los principales recursos. También puede crear sus propios informes con Cost Intelligence Dashboard, una solución de Amazon Quick creada por arquitectos de soluciones de AWS, y comparar sus presupuestos con el costo y el uso reales.

Pasos para la implementación

- Informe de la optimización de costos: defina un ciclo habitual para tratar y analizar la eficiencia de la carga de trabajo. Utilice las métricas establecidas y notifique las métricas alcanzadas y el costo para alcanzarlas. Identifique y corrija las tendencias negativas e identifique las tendencias positivas que quiere promover en su organización. La administración de informes debe implicar a los representantes de los equipos de Aplicaciones y los propietarios, así como de los departamentos de Finanzas y de Dirección.
- Creación y activación de presupuestos de AWS Budgets de granularidad diaria del costo y el uso a fin de tomar medidas oportunas para evitar posibles sobrecostos: AWS Budgets le permite configurar las notificaciones de alerta para mantenerse informado en caso de que alguno de sus tipos de presupuesto supere los umbrales preconfigurados. La mejor manera de aprovechar AWS Budgets es establecer los costos y el uso previstos como límites, de modo que todo lo que supere los presupuestos se considere un gasto excesivo.
- Creación de AWS Cost Anomaly Detection para la supervisión de costos: AWS Cost Anomaly Detection utiliza tecnología avanzada de machine learning para identificar los gastos anómalos y las causas que los originan para que pueda adoptar medidas rápidamente. Le permite configurar monitores de costos que definen los segmentos de gastos que desea evaluar (por ejemplo, servicios individuales de AWS, cuentas de miembro, etiquetas de asignación de costos y categorías de costos) y le permite establecer cuándo, dónde y cómo recibir sus notificaciones

de alerta. Para cada monitor, adjunte varias suscripciones de alerta para los propietarios de negocios y los equipos de tecnología, que incluyan un nombre, un umbral de impacto de costos y la frecuencia de las alertas (alertas individuales, resumen diario, resumen semanal) para cada suscripción.

- Use AWS Cost Explorer o integre de sus datos de AWS Cost and Usage Report (CUR) con los paneles de Amazon Quick para visualizar los costos de su organización: AWS Cost Explorer cuenta con una interfaz fácil de usar que le permite visualizar, comprender y administrar los costos y el uso de AWS a lo largo del tiempo. Cost Intelligence Dashboard es un panel personalizable y accesible para ayudar a crear la base de su propia herramienta de administración y optimización de costos.

Recursos

Documentos relacionados:

- AWS Budgets
- AWS Cost Explorer
- Daily Cost and Usage Budgets
- AWS Cost Anomaly Detection

Ejemplos relacionados:

- AWS Cost Anomaly Detection Alert with Slack

COST01-BP07 Seguimiento de la información sobre las nuevas versiones de los servicios

Consulte regularmente con expertos o socios de AWS qué servicios y características proporcionan un costo inferior. Revise los blogs de AWS y otras fuentes de información.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

AWS está agregando constantemente nuevas capacidades para que pueda aprovechar las últimas tecnologías para experimentar e innovar más rápidamente. Puede implementar nuevos

servicios y características de AWS para incrementar la rentabilidad de su carga de trabajo. Consulte periódicamente Administración de costos de AWS, el blog de novedades de AWS, el blog de administración de costos de AWS y Novedades de AWS para obtener información sobre las nuevas versiones de servicios y características. En las publicaciones de novedades se ofrece un breve resumen de todos los anuncios de servicios, características y ampliación de regiones de AWS a medida que se publican.

Pasos para la implementación

- Suscripción a los blogs: vaya a las páginas de los blogs de AWS y suscríbase al blog de novedades y a otros blogs pertinentes. Puede registrarse en la página de preferencias de comunicación con su dirección de correo electrónico.
- Suscripción a las novedades de AWS: consulte periódicamente el blog de novedades de AWS y Novedades de AWS para obtener información sobre los nuevos lanzamientos de servicios y características. Suscríbase a la fuente RSS o con su correo electrónico para seguir los anuncios y lanzamientos.
- Seguimiento de las reducciones de precios de AWS: la reducción periódica de los precios de todos nuestros servicios ha sido una forma habitual de AWS de trasladar a nuestros clientes las eficiencias económicas obtenidas gracias a nuestra escala. A 20 de septiembre de 2023, AWS ha reducido los precios 134 veces desde 2006. Si tiene alguna decisión comercial pendiente por cuestiones de precio, puede volver a revisarla después de las reducciones de precios y las nuevas integraciones de servicios. Puede obtener información sobre las iniciativas de reducción de precios anteriores, incluidas las instancias de Amazon Elastic Compute Cloud (Amazon EC2), en la categoría de reducción de precios del blog de novedades de AWS.
- Eventos y reuniones de AWS: asista a la cumbre local de AWS y a cualquier reunión local con otras organizaciones de su zona. Si no puede asistir en persona, intente asistir a los eventos virtuales para conocer mejor a los expertos de AWS y los casos empresariales de otros clientes.
- Reunión con su equipo de Cuentas: programe una cadencia regular con su equipo de cuentas, reúnase con él y trate sobre las tendencias del sector y los servicios de AWS. Hable con el administrador de cuentas, el arquitecto de soluciones y el equipo de Asistencia.

Recursos

Documentos relacionados:

- Administración de costos de AWS

- Novedades de AWS
- Blog de noticias de AWS

Ejemplos relacionados:

- Amazon EC2 – 15 Years of Optimizing and Saving Your IT Costs
- AWS News Blog - Price Reduction

COST01-BP08 Creación de una cultura sensibilizada con los costos

Implemente cambios o programas en la organización para crear una cultura sensibilizada con los costos. Se recomienda empezar discretamente, y a medida que crezcan las capacidades y el uso de la nube por parte de la empresa, implementar programas grandes y de gran alcance.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

Afianzar una cultura de conciencia de los costos permite mejorar la optimización de costos y la administración financiera en la nube (operaciones financieras, centro de excelencia en la nube, equipos de operaciones en la nube, etc.) a través de las prácticas recomendadas llevadas a cabo de forma orgánica y descentralizada en toda la organización. La conciencia de costos permite obtener grandes niveles de capacidad en la organización con un esfuerzo mínimo, en comparación con un enfoque centralizado y descendente.

La conciencia de costos en la computación en la nube, especialmente para los principales impulsores de costos en la computación en la nube, permite a los equipos comprender los resultados esperados de cualquier cambio en la perspectiva de los costos. Los equipos que acceden a los entornos de la nube deben conocer los modelos de precios y la diferencia entre los centros de datos tradicionales en las instalaciones y la computación en la nube.

La principal ventaja de una cultura sensibilizada con los costos es que los equipos de Tecnología los optimizan de forma proactiva y continua (por ejemplo, se consideran un requisito no funcional a la hora de diseñar nuevas cargas de trabajo o de hacer cambios en las existentes) en lugar de llevar a cabo optimizaciones de costos reactivas según sea necesario.

Aplicar unos pequeños cambios en la cultura puede tener un gran impacto en la eficiencia de las cargas de trabajo actuales y futuras. Ejemplos:

- Dar visibilidad y sensibilizar a los equipos de ingeniería para que comprendan lo que hacen y su impacto en términos de costos.
- Ludificar los costos y el uso en toda la organización. Esto se puede hacer con un panel visible para todo el personal o mediante un informe que compare los costos y el uso normalizados de los diferentes equipos (por ejemplo, costo por carga de trabajo y costo por transacción).
- Reconocer la rentabilidad. Premie los logros voluntarios o espontáneos de optimización de costos de forma pública o privada y aprenda de los errores para no repetirlos en el futuro.
- Crear requisitos organizativos descendentes para que las cargas de trabajo se ejecuten con presupuestos predefinidos.
- Cuestionar los requisitos empresariales de los cambios y el impacto de los costos de los cambios solicitados en la infraestructura de la arquitectura o la configuración de la carga de trabajo para asegurarse de que se paga solo lo que se necesita.
- Asegurarse de que el planificador del cambio es consciente de los cambios previstos que tienen un impacto en los costos y que las partes interesadas los confirmen para obtener resultados empresariales de forma rentable.

Pasos para la implementación

- Informe de los costos de la nube a los equipos de Tecnología: para aumentar la conciencia de costos y establecer KPI de eficiencia para las partes interesadas de las finanzas y la empresa.
- Informe a las partes interesadas o a los miembros del equipo sobre los cambios previstos: cree un punto en el orden del día para debatir los cambios previstos y el impacto del costo-beneficio en la carga de trabajo durante las reuniones semanales sobre cambios.
- Reunión con el equipo de Cuentas: establezca una cadencia de reuniones regular con su equipo de cuentas, y trate las tendencias del sector y los servicios de AWS. Hable con el administrador de cuentas, el arquitecto y el equipo de Asistencia.
- Comparta historias de éxito: comparta historias de éxito sobre la reducción de costos para cualquier carga de trabajo, Cuenta de AWS u organización para crear una actitud positiva y un estímulo en torno a la optimización de costos.
- Formación asegúrese de que los equipos técnicos o los miembros del equipo reciban formación para conocer los costos de los recursos en Nube de AWS.

- Eventos y reuniones de AWS: asista a las cumbres locales de AWS y a cualquier reunión local con otras organizaciones de su zona.
- Suscripción a los blogs: vaya a las páginas de los blogs de AWS y suscríbase al blog de novedades y a otros blogs pertinentes para seguir los nuevos lanzamientos, implementaciones, ejemplos y cambios que comparte AWS.

Recursos

Documentos relacionados:

- Blog de AWS
- Administración de costos de AWS
- Blog de noticias de AWS

Ejemplos relacionados:

- Administración financiera en la nube de AWS

COST01-BP09: cuantificación del valor empresarial de la optimización de costos

Cuantificar el valor empresarial de la optimización de costos le permite comprender todos los beneficios para su organización. Dado que la optimización de costos es una inversión necesaria, cuantificar el valor empresarial le permite explicar el retorno de la inversión a las partes interesadas. Cuantificar el valor empresarial lo puede ayudar a lograr mayor aceptación de las partes interesadas para hacer inversiones futuras en optimización de costos y, además, le proporciona un marco para medir los resultados de las actividades de optimización de costos de la organización.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

Cuantificar el valor empresarial implica determinar los beneficios que las empresas obtienen de las acciones y decisiones que toman. El valor empresarial puede ser tangible (por ejemplo, una reducción de los gastos o un aumento de los beneficios) o intangible (por ejemplo, la mejora de la reputación de la marca o el aumento de la satisfacción del cliente).

Para cuantificar el valor empresarial que se produce con la optimización de costos, hay que determinar cuánto valor o beneficio obtiene de sus esfuerzos por gastar de manera más eficiente. Por ejemplo, si una empresa gasta 100 000 USD en implementar una carga de trabajo en AWS y, posteriormente, la optimiza, el nuevo costo pasa a ser de solo 80 000 USD sin sacrificar la calidad ni el rendimiento. En este escenario, el valor empresarial cuantificado de la optimización de costos sería un ahorro de 20 000 USD. No obstante, más allá del ahorro, la empresa también podría cuantificar el valor en términos de tiempos de entrega más rápidos, una mayor satisfacción del cliente u otras métricas que se deriven de los esfuerzos de optimización de costos. Las partes interesadas deben tomar decisiones sobre el valor potencial de la optimización de costos, el costo de optimizar la carga de trabajo y el valor de retorno.

Además de informar sobre los ahorros de la optimización de costos, se recomienda cuantificar el valor adicional conseguido. Los beneficios de la optimización de costos se suelen cuantificar en términos de menos costos por resultado empresarial. Por ejemplo, puede cuantificar los ahorros de costos de Amazon Elastic Compute Cloud (Amazon EC2) al comprar Savings Plans, lo que reduce los costos y mantiene los niveles de producción de la carga de trabajo. Puede cuantificar reducciones de costos en el gasto de AWS cuando se eliminan las instancias de Amazon EC2 inactivas o cuando se eliminan volúmenes de Amazon Elastic Block Store (Amazon EBS) no asociados.

Sin embargo, la optimización de costos tiene muchos más beneficios, aparte de reducir o evitar costos. Plantéese capturar más datos para medir las mejoras en la eficiencia y el valor empresarial.

Pasos para la implementación

- Evaluación de los beneficios empresariales: este es el proceso de analizar y ajustar los costos de la Nube de AWS de manera que se maximice el beneficio recibido por cada dólar gastado. En lugar de centrarse en la reducción de costos sin valor empresarial, considere los beneficios empresariales y el retorno de la inversión de la optimización de costos, lo que puede aportar más valor al dinero que gasta. Se trata de gastar con prudencia y llevar a cabo inversiones y gastos en las áreas que tengan el mejor rendimiento.
- Análisis de los costos de AWS de las previsiones: las previsiones ayudan a las partes interesadas de finanzas a establecer expectativas con otras partes interesadas internas y externas de la organización y pueden mejorar la previsibilidad financiera de la organización. AWS Cost Explorer se puede utilizar para hacer una previsión de los costos y el uso.

Recursos

Documentos relacionados:

- Fundamentos económicos de la Nube de AWS
- Blog de AWS
- Administración de costos de AWS
- Blog de noticias de AWS
- Documento técnico Pilar de fiabilidad: Well-Architected
- Explorador de costos de AWS

Videos relacionados:

- Unlock Business Value with Windows on AWS

Ejemplos relacionados:

- Measuring and Maximizing the Business Value of Customer 360
- The Business Value of Adopting Amazon Web Services Managed Databases
- The Business Value of Amazon Web Services for Independent Software Vendors
- Business Value of Cloud Modernization
- The Business Value of Migration to Amazon Web Services

Conocimiento del gasto y del uso

Es fundamental comprender los costos y lo que impulsa a su organización para administrar eficazmente el costo y el uso e identificar las oportunidades de reducción de costos. Las organizaciones suelen operar múltiples cargas de trabajo que ejecutan varios equipos. Estos equipos pueden estar en diferentes unidades organizativas, cada una con su propio flujo de ingresos. La capacidad de atribuir los costos de los recursos a las cargas de trabajo, a la organización individual o a los propietarios de los productos impulsa un comportamiento de uso eficiente y contribuye a reducir los residuos. Una supervisión precisa del costo y del uso le permite comprender la rentabilidad de los productos y las unidades organizativas; de este modo, podrá tomar decisiones más informadas sobre dónde asignar los recursos en su organización. El conocimiento del uso en todos los niveles de la organización es clave para impulsar el cambio, ya que el cambio en el uso impulsa los cambios en el costo.

Considere la posibilidad de adoptar un enfoque multifacético para conocer su uso y sus gastos. Su equipo tiene que recopilar gastos, analizarlos e informar de ellos. Entre los factores clave que se deben tener en cuenta se incluyen los siguientes:

Temas

- Gobernanza
- Supervisión del costo y el uso
- Retirada de recursos

Gobernanza

Para administrar los costos en la nube, debe administrar su uso a través de las siguientes áreas de gobernanza:

Prácticas recomendadas

- COST02-BP01 Desarrollo de políticas basadas en los requisitos de su organización
- COST02-BP02 Implementación de objetivos y metas
- COST02-BP03 Implementación de una estructura de cuentas
- COST02-BP04 Implementación de grupos y roles
- COST02-BP05 Implementación de controles de costos
- COST02-BP06 Seguimiento del ciclo de vida de los proyectos

COST02-BP01 Desarrollo de políticas basadas en los requisitos de su organización

Desarrolle políticas que definan la forma en que su organización administra los recursos e inspecciónelas periódicamente. Las políticas deben abarcar los aspectos de costo de los recursos y las cargas de trabajo, como su creación, modificación y retirada durante la vida útil del recurso.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Es fundamental comprender los costos y lo que impulsa a su organización para administrar eficazmente el costo y el uso e identificar las oportunidades de reducción de costos. Las organizaciones suelen operar múltiples cargas de trabajo que ejecutan varios equipos. Estos equipos pueden estar en diferentes unidades organizativas, cada una con su propio flujo de ingresos. La capacidad de atribuir los costos de los recursos a las cargas de trabajo, a la organización individual o a los propietarios de los productos impulsa un comportamiento de uso eficiente y contribuye a reducir los residuos. La monitorización precisa de los costos y el uso le ayuda a comprender el nivel de optimización de una carga de trabajo, así como la rentabilidad de las unidades y los productos de la organización. Este conocimiento permite tomar decisiones más fundamentadas sobre dónde asignar los recursos dentro de la organización. El conocimiento del uso en todos los niveles de la organización es clave para impulsar el cambio, ya que el cambio en el uso impulsa los cambios en el costo. Considere la posibilidad de adoptar un enfoque multifacético para conocer su uso y sus gastos.

El primer paso para aplicar la gobernanza es utilizar los requisitos de su organización para desarrollar políticas de uso de la nube. Estas políticas definen cómo su organización utiliza la nube y cómo se administran los recursos. Las políticas deben tratar todos los aspectos de los recursos y las cargas de trabajo que tienen que ver con el costo o el uso, como su creación, modificación y retirada durante la vida útil del recurso. Verifique que se siguen e implementan las políticas y los procedimientos ante cualquier cambio en un entorno en la nube. Durante las reuniones de administración de cambios de TI, formule preguntas para averiguar el impacto en los costos de los cambios previstos, tanto si aumentan como si disminuyen, la justificación empresarial y el resultado esperado.

Las políticas deben ser sencillas para que se comprendan fácilmente y puedan implementarse con eficacia en toda la organización. Las políticas también deben ser fáciles de seguir e interpretar (para que se usen) y específicas (para que no haya malinterpretaciones entre los equipos). Además,

deben inspeccionarse periódicamente (igual que nuestros mecanismos) y actualizarse a medida que cambien las condiciones empresariales o las prioridades de los clientes, ya que esto podría hacer que la política quedara obsoleta.

Empiece con políticas amplias y generales, como la región geográfica que se usará o las horas del día en las que deben ejecutarse los recursos. Mejore gradualmente las políticas para las distintas unidades organizativas y cargas de trabajo. Entre las políticas más comunes se incluyen los servicios y las características que pueden utilizarse (por ejemplo, el almacenamiento de menor rendimiento en los entornos de prueba o de desarrollo), los tipos de recursos que pueden utilizar los distintos grupos (por ejemplo, el mayor tamaño de recurso en una cuenta de desarrollo es el medio) y durante cuánto tiempo estarán en uso estos recursos (temporalmente, a corto plazo o durante un periodo de tiempo específico).

Ejemplo de política

A continuación, tenemos un ejemplo de política que puede utilizar para crear sus propias políticas de gobernanza en la nube que se centren en la optimización de costos. Asegúrese de ajustar la política en función de los requisitos de su organización y de las peticiones de las partes interesadas.

- Nombre de la política: defina un nombre claro, como “Política de optimización de recursos y reducción de costos”.
- Propósito: explique por qué se debe utilizar esta política y cuál es el resultado esperado. El objetivo de esta política es verificar que se requiere un costo mínimo para implementar y ejecutar la carga de trabajo deseada con el fin de cumplir los requisitos empresariales.
- Ámbito: defina claramente quién debe usar esta política y cuándo debe usarse; por ejemplo, podría indicar que el equipo X de DevOps debe usar esta política en los clientes de la región us-east para el entorno X (de producción o no producción).

Declaración de la política

1. Seleccione us-east-1 o varias regiones de us-east en función del entorno y los requisitos empresariales de su carga de trabajo (desarrollo, pruebas de aceptación de los usuarios, preproducción o producción).
2. Programe instancias de Amazon EC2 y Amazon RDS para que se ejecuten entre las seis de la mañana y las ocho de la tarde (hora estándar del este [EST]).
3. Detenga todas las instancias de Amazon EC2 no utilizadas después de ocho horas y las instancias de Amazon RDS no utilizadas después de 24 horas de inactividad.

4. Termine todas las instancias de Amazon EC2 no utilizadas después de 24 horas de inactividad en entornos que no sean de producción. Recuérdele al propietario de la instancia de Amazon EC2 (basándose en las etiquetas) que revise las instancias de Amazon EC2 detenidas en producción e infórmele de que sus instancias de Amazon EC2 se terminarán en un plazo de 72 horas si no están en uso.
5. Utilice una familia y un tamaño de instancias genéricos, como m5.large, y luego cambie el tamaño de la instancia en función del uso de la CPU y la memoria mediante AWS Compute Optimizer.
6. Priorice el uso del escalado automático para ajustar dinámicamente la cantidad de instancias en ejecución en función del tráfico.
7. Utilice instancias de spot para cargas de trabajo no críticas.
8. Revise los requisitos de capacidad para confirmar Savings Plans o instancias reservadas para cargas de trabajo predecibles e informe al equipo de Administración financiera en la nube.
9. Utilice políticas de ciclo de vida de Amazon S3 para mover los datos a los que se accede con poca frecuencia a niveles de almacenamiento más económicos. Si no se ha definido ninguna política de retención, utilice Amazon S3 Intelligent-Tiering para mover los objetos al nivel de archivado automáticamente.
10. Supervise el uso de los recursos y configure alarmas para activar eventos de escalado mediante Amazon CloudWatch.
11. Para cada Cuenta de AWS, utilice AWS Budgets para establecer presupuestos de costos y uso para su cuenta en función del centro de costos y las unidades de negocios.
12. Si usa AWS Budgets para establecer presupuestos de costos y uso para su cuenta, puede resultarle más fácil controlar sus gastos y evitar facturas inesperadas, lo que te permitirá controlar mejor sus costos.

Procedimiento: proporcione procedimientos detallados para implementar esta política o consulte otros documentos en los que se describa cómo implementar cada declaración de la política. En esta sección, se deben proporcionar instrucciones paso a paso para cumplir los requisitos de la política.

Para implementar esta política, puede utilizar diversas herramientas o reglas de AWS Config de terceros para comprobar si se cumple la declaración de la política y activar medidas de corrección automatizadas mediante funciones de AWS Lambda. También puede utilizar AWS Organizations para hacer cumplir la política. Además, debe revisar periódicamente el uso de sus recursos y ajustar la política según sea necesario para comprobar que sigue satisfaciendo las necesidades de su empresa.

Pasos para la implementación

- Reunión con las partes interesadas: para desarrollar políticas, pida a las partes interesadas (oficinas de la empresa en la nube, ingenieros o responsables de la toma de decisiones funcionales para la aplicación de las políticas) de su organización que especifiquen sus requisitos y los documenten. Adopte un enfoque iterativo; para ello, empiece con un enfoque amplio y vaya reduciendo hasta llegar a las unidades más pequeñas en cada paso. Entre los miembros del equipo se encuentran los que tienen un interés directo en la carga de trabajo, como las unidades organizativas o los propietarios de las aplicaciones, además de los grupos de asistencia, como los equipos de Seguridad y Finanzas.
- Obtención de la confirmación: asegúrese de que los equipos se ponen de acuerdo en las políticas sobre quién puede acceder e implementar en la Nube de AWS. Asegúrese de que siguen las políticas de su organización y confirme que sus creaciones de recursos se ajustan a las políticas y procedimientos acordados.
- Creación de sesiones de formación de incorporación: pida a los nuevos miembros de la organización que completen los cursos de capacitación de incorporación para concienciar de los costos y que conozcan los requisitos de la organización. Es posible que asuman políticas diferentes debido a su experiencia anterior o que no piensen en ellas en absoluto.
- Definición de ubicaciones para la carga de trabajo: defina dónde opera su carga de trabajo, incluido el país y la zona dentro del país. Esta información se utiliza para la asignación de Regiones de AWS y zonas de disponibilidad.
- Definición y agrupación de los servicios y los recursos: defina los servicios que requieren las cargas de trabajo. Para cada servicio, especifique los tipos, el tamaño y el número de recursos necesarios. Defina grupos para los recursos por función, como servidores de aplicaciones o almacenamiento de bases de datos. Los recursos pueden pertenecer a varios grupos.
- Definición y agrupación de los usuarios por función: defina a los usuarios que interactúan con la carga de trabajo; para ello, céntrese en lo que hacen y en cómo utilizan la carga de trabajo, no en quiénes son o en su puesto en la organización. Agrupe usuarios o funciones similares. Puede utilizar las políticas administradas por AWS como guía.
- Definición de las acciones: mediante las ubicaciones, los recursos y los usuarios identificados anteriormente, defina las acciones que requiere cada uno de ellos para lograr los resultados de la carga de trabajo a lo largo de su vida útil (desarrollo, funcionamiento y retirada). Identifique las acciones en función de los grupos, y no de los elementos individuales de los grupos, en cada ubicación. Empiece a grandes rasgos con la lectura o la escritura y, después, vaya reduciendo hasta llegar a las acciones específicas para cada servicio.

- Definición del periodo de revisión: las cargas de trabajo y los requisitos organizativos pueden cambiar con el tiempo. Defina el calendario de revisión de la carga de trabajo para asegurarse de que se mantiene alineado con las prioridades organizativas.
- Documentación de las políticas: verifique que las políticas que se han definido sean accesibles tal y como lo requiere su organización. Estas políticas se utilizan para implementar, mantener y auditar el acceso de sus entornos.

Recursos

Documentos relacionados:

- Change Management in the Cloud
- Políticas administradas de AWS para funciones de trabajo
- Estrategia de facturación de varias cuentas de AWS
- Actions, Resources, and Condition Keys for AWS Services
- Administración y gobernanza en AWS
- Control access to Regiones de AWS using IAM policies
- Regiones y zonas de disponibilidad de la infraestructura global

Videos relacionados:

- AWS Management and Governance at Scale

COST02-BP02 Implementación de objetivos y metas

Implemente objetivos de costos y uso para la carga de trabajo. Los objetivos son una guía de resultados previstos para la organización. Las metas proporcionan resultados medibles específicos que se deben alcanzar para las cargas de trabajo.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Desarrolle objetivos y metas de costos y uso para su organización. Para una organización en crecimiento en AWS, es importante establecer objetivos de optimización de costos y hacer un

seguimiento de ellos. Estos objetivos o indicadores clave de rendimiento(KPI) pueden incluir aspectos como el porcentaje del gasto bajo demanda o la adopción de ciertos servicios optimizados, como instancias de AWS Graviton o tipos de volúmenes gp3 de EBS. Establezca objetivos cuantificables y factibles que lo ayuden a medir las mejoras de la eficiencia, lo cual es importante para las operaciones empresariales. Los objetivos son una guía sobre los resultados esperados para su organización.

Las metas proporcionan resultados cuantificables específicos que se deben alcanzar. En resumen, un objetivo es la dirección en la que quiere ir y la meta es hasta dónde ir en esa dirección y cuándo debe lograrse ese objetivo (mediante una orientación específica, cuantificable, asignable, realista y oportuna, o SMART, por sus siglas en inglés). Un ejemplo de un objetivo es que el uso de la plataforma debería incrementarse de forma significativa con tan solo un ligero incremento (no lineal) del costo. Un ejemplo de meta es un incremento del 20 % del uso de la plataforma con un incremento de menos del 5 % de los costos. Otro objetivo común es que las cargas de trabajo deben ser más eficientes cada seis meses. La meta correspondiente sería que las métricas de costo por empresa disminuyan un 5 % cada seis meses. Use las métricas correctas y establezca KPI calculados para su organización. Puede empezar con KPI básicos e ir evolucionando en función de las necesidades de la empresa.

Un objetivo de la optimización de costos es aumentar la eficiencia de la carga de trabajo, lo que supone reducir el costo por resultado empresarial de la carga de trabajo con el tiempo. Implemente este objetivo para todas las cargas de trabajo y fíjese una meta, como un aumento del cinco por ciento de la eficacia cada seis meses a un año. En la nube, puede lograrlo estableciendo capacidades en la optimización de costos, así como con el lanzamiento de nuevos servicios y características.

Las metas son los puntos de referencia cuantificables que desea alcanzar para cumplir sus objetivos y los puntos de referencia le permiten comparar sus resultados reales con una meta. Establezca puntos de referencia con KPI para el costo unitario de los servicios de computación (por ejemplo, adopción de Spot, adopción de Graviton, últimos tipos de instancias y cobertura bajo demanda), los servicios de almacenamiento (por ejemplo, adopción de EBS GP3, instantáneas obsoletas de EBS y almacenamiento de Amazon S3 Standard) o el uso de los servicios de bases de datos (por ejemplo, motores de código abierto de RDS, adopción de Graviton y cobertura bajo demanda). Estos puntos de referencia y KPI pueden ayudarlo a verificar que utiliza los servicios de AWS de la manera más rentable.

En la siguiente tabla se proporciona una lista de métricas de AWS estándar como referencia. Cada organización puede tener diferentes valores meta para estos KPI.

Categoría	KPI (%)	Descripción
Computación	Cobertura de uso de EC2	Instancias de EC2 (en costo u horas) que utilizan Savings Plans + instancias reservadas + spot en comparación con el total (en costo u horas) de las instancias de EC2
Computación	Cómputo del uso de Savings Plans o instancias reservadas	Horas utilizadas de Savings Plans o instancias reservadas en comparación con el total de horas disponibles de Savings Plans o instancias reservadas
Computación	Costo de EC2/hora	Costo de EC2 dividido por el número de instancias de EC2 que se ejecutan en esa hora
Computación	Costo de vCPU	Costo por vCPU para todas las instancias
Computación	Última generación de instancias	Porcentaje de instancias en Graviton (u otros tipos de instancias de generación moderna)
Base de datos	Cobertura de RDS	Instancias de RDS (en costo u horas) que utilizan instancias reservadas en comparación con el total (en costo u horas) de las instancias de RDS
Base de datos	Utilización de RDS	Horas utilizadas de instancias reservadas en comparación con el total de horas disponibles de instancias reservadas

Categoría	KPI (%)	Descripción
Base de datos	Tiempo de actividad de RDS	El costo de RDS dividido por el número de instancias de RDS que se ejecuten en esa hora
Base de datos	Última generación de instancias	Porcentaje de instancias en Graviton (u otros tipos modernos de instancias)
Almacenamiento	Uso del almacenamiento	Costo de almacenamiento optimizado (por ejemplo, Glacier, archivo profundo o acceso poco frecuente) dividido por el costo total de almacenamiento
Etiquetado	Recursos sin etiquetar	Explorador de costos: 1. Filtre créditos, descuentos, impuestos, reembolsos y mercados, y copie el último costo mensual. 2. Seleccione Mostrar solo los recursos sin etiquetar en el Explorador de costos. 3. Divida la cantidad de recursos sin etiquetar por el costo mensual.

Utilizando esta tabla, incluya los valores meta o de referencia, que deberán calcularse en función de los objetivos de la organización. Debe medir determinadas métricas de su negocio y conocer el resultado empresarial de esa carga de trabajo para definir unos indicadores clave de rendimiento (KPI) precisos y realistas. Cuando evalúe las métricas de rendimiento dentro de una organización, distinga entre diferentes tipos de métricas que sirven para propósitos distintos. Estas métricas

miden principalmente el rendimiento y la eficiencia de la infraestructura técnica en lugar de la repercusión global en el negocio directamente. Por ejemplo, podrían hacer un seguimiento de los tiempos de respuesta del servidor, la latencia de la red o el tiempo de actividad del sistema. Estas métricas son cruciales para evaluar cómo admite la infraestructura las operaciones técnicas de la organización. Sin embargo, no proporcionan información directa de objetivos empresariales más amplios, como la satisfacción del cliente, el crecimiento de los ingresos o la cuota de mercado. Para tener un conocimiento global del rendimiento empresarial, complemente estas métricas de eficiencia con métricas empresariales estratégicas que estén directamente relacionadas con los resultados empresariales.

Establezca una visibilidad casi en tiempo real de los KPI y las oportunidades de ahorro relacionadas y haga un seguimiento del progreso a lo largo del tiempo. Para empezar a definir los objetivos de los KPI y hacer un seguimiento de ellos, recomendamos utilizar el panel de KPI de Cloud Intelligence Dashboards (CID). En función de los datos del Informe de costos y uso (CUR), el panel de KPI proporciona una serie de KPI de optimización de costos recomendados con la capacidad de establecer objetivos personalizados y hacer un seguimiento de su progreso a lo largo del tiempo.

Si dispone de otras soluciones para establecer los objetivos de los KPI y hacer un seguimiento de ellos, asegúrese de que todas las partes interesadas de la administración financiera en la nube de su organización adopten esos métodos.

Pasos para la implementación

- Defina los niveles de uso esperados: para empezar, céntrese en los niveles de uso. Contacte con los propietarios de aplicaciones, los equipos de Marketing y otros equipos importantes de la empresa para conocer los niveles de uso esperados de la carga de trabajo. ¿Cómo podría cambiar la demanda de los clientes con el tiempo y qué podría cambiar debido a los incrementos de temporada o a las campañas de marketing?
- Defina los recursos y los costos de las cargas de trabajo: una vez definidos los niveles de uso, se deben cuantificar los cambios en los recursos de las cargas de trabajo necesarios para ajustarse a dichos niveles de uso. Es posible que tenga que incrementar el tamaño o el número de recursos para un componente de la carga de trabajo, incrementar la transferencia de datos o cambiar los componentes de las cargas de trabajo por un servicio distinto en un nivel determinado. Especifique los costos en cada uno de estos puntos importantes y prediga el cambio en el costo cuando se produzca un cambio en el uso.
- Defina los objetivos empresariales: debe combinar el resultado de los cambios previstos en el uso y los costos con los cambios previstos en la tecnología, o cualquier programa que esté ejecutando, y establecer objetivos para la carga de trabajo. Los objetivos deben centrarse en el

uso y los costos, además de en la relación entre ambos. Los objetivos deben ser sencillos y de alto nivel. Además, deben ayudar a otras personas a saber lo que espera la empresa en cuanto a los resultados (por ejemplo, asegurarse de que los recursos sin usar estén por debajo de un determinado nivel de costo). No tiene que definir objetivos para cada tipo de recurso no utilizado ni definir costos que provoquen pérdidas en los objetivos y las metas. Verifique que haya programas organizativos (por ejemplo, desarrollo de capacidades a través de cursos de formación) si se prevén cambios en los costos sin cambios en el uso.

- Definición de objetivos: debe especificar un objetivo cuantificable para cada de uno de los objetivos definidos. Si el objetivo es incrementar la eficiencia de la carga de trabajo, la meta debería cuantificar la mejora (normalmente en resultados empresariales por cada dólar gastado) y cuándo debe producirse. Por ejemplo, podría fijar un objetivo para minimizar el desperdicio debido a un exceso de aprovisionamiento. Con este objetivo, su meta puede ser que el desperdicio debido a un exceso de aprovisionamiento de computación en el primer nivel de las cargas de trabajo de producción no supere el 10 % del costo de computación del nivel. Además, una segunda meta podría ser que el desperdicio debido a un exceso de aprovisionamiento de computación en el segundo nivel de las cargas de trabajo de producción no supere el 5 % del costo de computación del nivel.

Recursos

Documentos relacionados:

- Políticas administradas de AWS para funciones de trabajo
- Estrategia de facturación de varias cuentas de AWS
- Control access to Regiones de AWS using IAM policies
- S.M.A.R.T. Goals
- How to track your cost optimization KPIs with the CID KPI Dashboard

Videos relacionados:

- Well-Architected Labs: Goals and Targets (Level 100)

Ejemplos relacionados:

- What is a unit metric?
- Selecting a unit metric to support your business

- Unit metrics in practice – lessons learned
- How unit metrics help create alignment between business functions

COST02-BP03 Implementación de una estructura de cuentas

Implemente una estructura de cuentas adaptada a su organización. Esto le ayudará a asignar y administrar los costos en toda la organización.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

AWS Organizations le permite crear varias Cuentas de AWS que pueden ayudarle a controlar de forma centralizada su entorno a medida que escala sus cargas de trabajo en AWS. Puede modelar su jerarquía organizativa si agrupa las Cuentas de AWS en una estructura de unidades organizativas (OU) y crea varias Cuentas de AWS en cada OU. Para crear una estructura de cuentas, primero debe decidir cuál de sus Cuentas de AWS será la de administración. Después de eso, puede crear nuevas Cuentas de AWS o seleccionar cuentas existentes como cuentas de miembro en función de la estructura de cuentas que haya diseñado según las prácticas recomendadas de cuentas de administración y las prácticas recomendadas de cuentas de miembro.

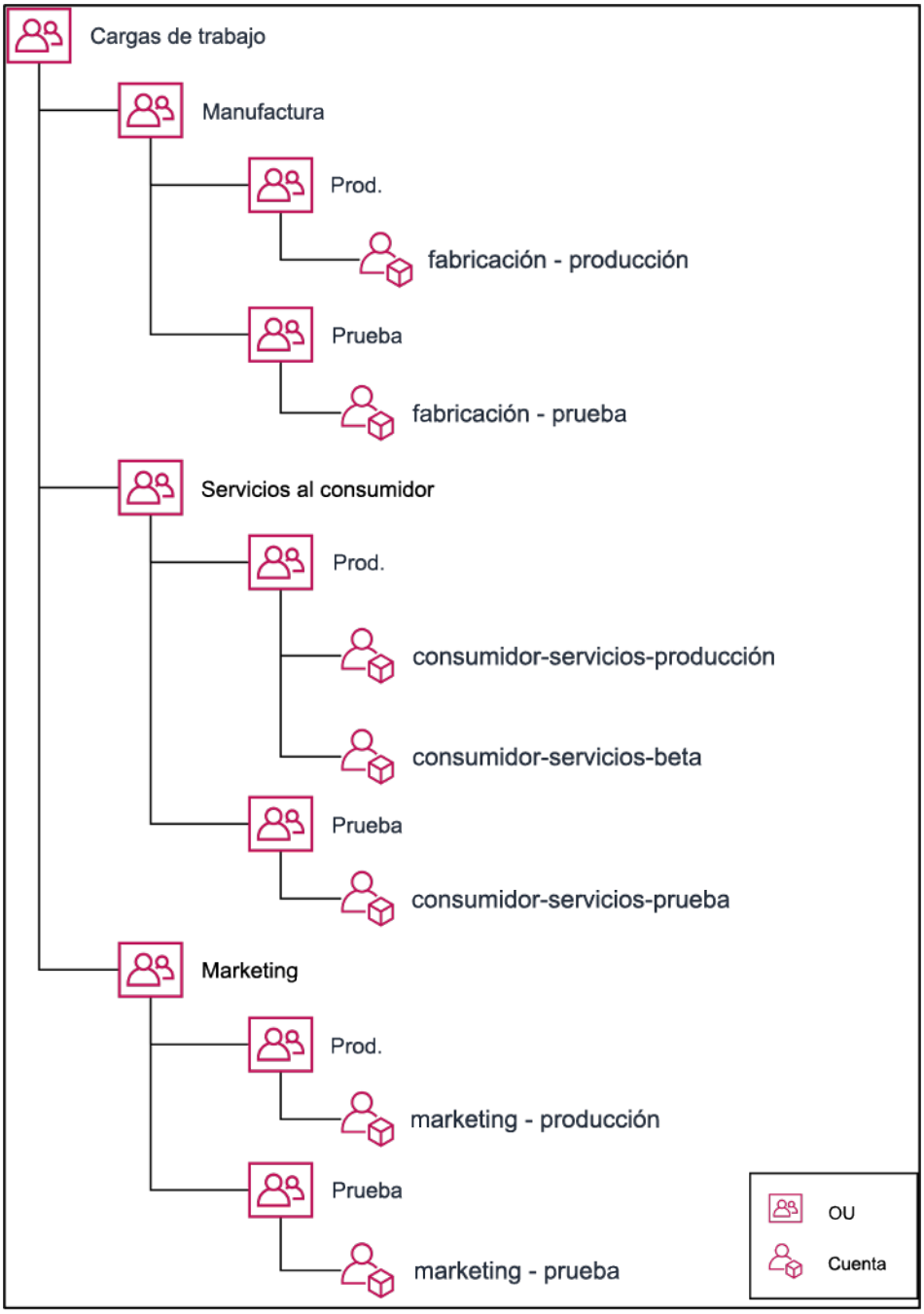
Se aconseja tener siempre al menos una cuenta de administración con una cuenta de miembro vinculada, sin importar el tamaño de la organización o su uso. Los recursos de las cargas de trabajo deberían estar solo en las cuentas de miembro y no se debería crear ningún recurso en la cuenta de administración. En cuanto a la pregunta sobre la cantidad de Cuentas de AWS que se debe tener, no existe una sola respuesta correcta para todas las situaciones. Primero debe evaluar sus modelos operativos y de costos, tanto actuales como futuros, para asegurarse de que la estructura de sus Cuentas de AWS refleje los objetivos de su organización. Algunas empresas crean varias Cuentas de AWS por motivos empresariales, por ejemplo:

- Se requiere aislamiento administrativo o fiscal y de facturación entre unidades organizativas, centros de costos o cargas de trabajo específicas.
- Los límites de servicio de AWS están configurados para ser específicos para cargas de trabajo concretas.
- Existe un requisito de aislamiento y separación entre cargas de trabajo y recursos.

En AWS Organizations, la facturación unificada crea la estructura entre una o más cuentas de miembro y la cuenta de administración. Las cuentas de miembro le permiten aislar y distinguir los costos y el uso por grupos. Una práctica común es tener cuentas de miembro independientes para cada unidad organizativa (como finanzas, marketing y ventas), para cada ciclo de vida del entorno (como desarrollo, prueba y producción) o para cada carga de trabajo (carga de trabajo a, b y c) y luego agregar dichas cuentas vinculadas mediante la facturación unificada.

La facturación unificada le permite unificar el pago de varias Cuentas de AWS de miembro en una sola cuenta de administración y proporcionar a la vez visibilidad de la actividad de cada cuenta vinculada. A medida que se agregan costos y uso a la cuenta de administración, puede maximizar los descuentos de volumen de servicio y el uso de los descuentos por compromiso (Savings Plans e instancias reservadas) para obtener los mayores descuentos.

En el siguiente diagrama se muestra cómo puede utilizar AWS Organizations con unidades organizativas (OU) para agrupar varias cuentas y colocar múltiples Cuentas de AWS en cada OU. Se recomienda utilizar OU para diversos casos de uso y cargas de trabajo, lo que proporciona patrones para organizar las cuentas.



Ejemplo de agrupación de varias Cuentas de AWS en unidades organizativas.

AWS Control Tower puede configurar rápidamente varias cuentas de AWS, lo que garantiza que la gobernanza esté alineada con los requisitos de la organización.

Pasos para la implementación

- Definición de los requisitos de separación: los requisitos de separación son una combinación de múltiples factores, como la seguridad, la fiabilidad y los modelos financieros. Ocúpese de cada

factor por orden y especifique si la carga de trabajo o el entorno de la carga de trabajo deberían separarse de otras cargas de trabajo. La seguridad promueve la adhesión a los requisitos de acceso y datos. La fiabilidad administra los límites de tal forma que los entornos y las cargas de trabajo no afecten a los demás. Consulte periódicamente los pilares de seguridad y fiabilidad del Marco de Well-Architected y siga las prácticas recomendadas. Los componentes financieros crean una separación financiera estricta (centro de costo diferente, propietarios de la carga de trabajo y responsabilidad). Algunos ejemplos comunes de separación son la ejecución de cargas de trabajo de producción y prueba en cuentas separadas o el uso de una cuenta separada para que la factura y los datos de facturación se puedan proporcionar a las unidades de negocios o departamentos individuales de la organización o parte interesada propietaria de la cuenta.

- Definición de los requisitos de agrupación: los requisitos de agrupación no anulan los requisitos de separación, pero se utilizan para ayudar en la administración. Agrupe entornos o cargas de trabajo similares que no requieran separación. Un ejemplo es agrupar múltiples entornos de prueba o desarrollo de una o varias cargas de trabajo.
- Definición de la estructura de cuentas: use estas separaciones y agrupaciones, especifique una cuenta para cada grupo y asegúrese de que se cumplan los requisitos de separación. Estas cuentas son sus cuentas de miembro o vinculadas. Al agrupar estas cuentas de miembro en una única cuenta de administración o de pagador, combina el uso, lo que le permite disfrutar de descuentos de mayor volumen en todas las cuentas y le proporciona una sola factura para todas las cuentas. Es posible separar los datos de facturación y proporcionar a cada cuenta de miembro una vista individual de sus datos de facturación. Si una cuenta de miembro no debe tener los datos de facturación o de uso visibles para las demás cuentas, o si se requiere una factura distinta de AWS, defina múltiples cuentas de administración o de pagador. En este caso, cada cuenta de miembro tiene su propia cuenta de administración o de pagador. Los recursos deberían colocarse siempre en las cuentas de miembro o vinculadas. Las cuentas de administración o de pagador solo deben usarse para tareas de administración.

Recursos

Documentos relacionados:

- Uso de etiquetas de asignación de costos
- Políticas administradas de AWS para funciones de trabajo
- Estrategia de facturación de varias cuentas de AWS
- Control access to Regiones de AWS using IAM policies
- AWS Control Tower

- AWS Organizations
- Prácticas recomendadas de cuentas de administración y cuentas de miembro
- Organización de su entorno de AWS con varias cuentas
- Turning on shared reserved instances and Savings Plans discounts
- Consolidated billing
- Consolidated billing

Ejemplos relacionados:

- Splitting the CUR and Sharing Access

Videos relacionados:

- Introducing AWS Organizations
- Set Up a Multi-Account AWS Environment that Uses Best Practices for AWS Organizations

Ejemplos relacionados:

- Defining an AWS Multi-Account Strategy for telecommunications companies
- Best Practices for Optimizing Cuentas de AWS
- Best Practices for Organizational Units with AWS Organizations

COST02-BP04 Implementación de grupos y roles

Implemente grupos y roles que se ajusten a sus políticas y controle quién puede crear, modificar o retirar instancias y recursos en cada grupo. Por ejemplo, implemente grupos de desarrollo, de pruebas y de producción. Esto se aplica a los servicios de AWS y a las soluciones de terceros.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

Los roles y los grupos de usuarios son elementos fundamentales en el diseño y la implementación de sistemas seguros y eficaces. Los roles y los grupos ayudan a las organizaciones a equilibrar la necesidad de control con el requisito de flexibilidad y productividad, lo que facilita, en última instancia, los objetivos de la organización y las necesidades de los usuarios. Como se recomienda

en la sección de Administración de identidad y acceso del pilar de seguridad del Marco de AWS Well-Architected, necesita contar con una sólida gestión de identidades y permisos para proporcionar acceso a los recursos correctos a las personas adecuadas en las condiciones adecuadas. Los usuarios reciben solo el acceso necesario para completar sus tareas. Esto minimiza el riesgo asociado con el acceso no autorizado o el uso indebido.

Después de desarrollar las políticas, puede crear grupos lógicos y roles de usuario dentro de la organización. Esto le permite asignar permisos, controlar el uso y ayudar a implementar mecanismos de control de acceso sólidos, lo que evita el acceso no autorizado a la información confidencial. Empiece con grupos de personas generales. Esto suele corresponderse con unidades organizativas y roles de trabajos (por ejemplo, un administrador de sistemas del departamento de TI, responsable financiero o analistas empresariales). Los grupos permiten clasificar a las personas que llevan a cabo tareas similares y necesitan accesos similares. Los roles definen lo que debe hacer un grupo. Es más fácil administrar permisos de grupos y roles que permisos de usuarios individuales. Los roles y los grupos asignan permisos de manera uniforme y sistemática a todos los usuarios, lo que evita errores e incoherencias.

Cuando cambia el rol de un usuario, los administradores pueden ajustar el acceso por rol o grupo, en lugar de volver a configurar cuentas de usuarios individuales. Por ejemplo, un administrador de sistemas de TI requiere acceso para crear todos los recursos, pero un miembro del equipo de Análisis solo lo necesita para crear recursos de análisis.

Pasos para la implementación

- Implementación de grupos: use los grupos de usuarios definidos en sus políticas organizativas e implemente los grupos correspondientes, si es necesario. Para conocer las prácticas recomendadas sobre los usuarios, los grupos y la autenticación, consulte el pilar de seguridad del Marco de AWS Well-Architected.
- Implementación de roles y políticas: use las acciones definidas en sus políticas organizativas y cree los roles y las políticas de acceso necesarios. Para conocer las prácticas recomendadas sobre funciones y políticas, consulte el pilar de seguridad del Marco de AWS Well-Architected.

Recursos

Documentos relacionados:

- Políticas administradas de AWS para funciones de trabajo
- Estrategia de facturación de varias cuentas de AWS

- Pilar de seguridad: Marco de AWS Well-Architected
- AWS Identity and Access Management (IAM)
- Políticas de AWS Identity and Access Management

Videos relacionados:

- Why use Identity and Access Management

Ejemplos relacionados:

- Control access to Regiones de AWS using IAM policies
- Starting your Cloud Financial Management journey: Cloud cost operations

COST02-BP05 Implementación de controles de costos

Implemente controles basados en las políticas de la organización y en los grupos y los roles definidos. De este modo, se certifica que los costos solo se producen según los requisitos de la organización, como controlar el acceso a regiones o tipos de recursos.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

Un primer paso común en la implementación de los controles de costos es establecer notificaciones cuando se producen eventos de costos o de uso fuera de las políticas. Puede actuar con rapidez y verificar si es necesaria una acción correctiva, sin restringir ni afectar negativamente a las cargas de trabajo o la nueva actividad. Una vez que conozca los límites de la carga de trabajo y del entorno, podrá aplicar la gobernanza. AWS Budgets le permite configurar notificaciones y definir presupuestos mensuales para sus costos, uso y descuentos por compromiso de AWS (Savings Plans e instancias reservadas). Puede crear presupuestos en un nivel de costo agregado (por ejemplo, todos los costos) o en un nivel más detallado en el que incluya solo dimensiones específicas como, por ejemplo, cuentas vinculadas, servicios, etiquetas o zonas de disponibilidad.

Una vez que haya establecido sus límites presupuestarios con AWS Budgets, use AWS Cost Anomaly Detection para reducir los costos imprevistos. AWS Cost Anomaly Detection es un servicio de administración de costos que utiliza machine learning para supervisar de forma continua el costo y el uso y, así, detectar gastos inusuales. Le ayuda a identificar los gastos anómalos y las causas

que los originan para que pueda adoptar medidas rápidamente. En primer lugar, cree un monitor de costos en AWS Cost Anomaly Detection, y, a continuación, elija su preferencia de alerta mediante el establecimiento de un umbral en dólares (como una alerta sobre anomalías con un impacto superior a 1000 USD). Una vez que reciba las alertas, podrá analizar la causa raíz que provoca la anomalía y el impacto en los costos. También puede supervisar y hacer sus propios análisis de anomalías en AWS Cost Explorer.

Aplique las políticas de gobernanza en AWS mediante AWS Identity and Access Management y políticas de control de servicios (SCP) de AWS Organizations. IAM le permite administrar el acceso a los servicios y recursos de AWS de forma segura. Mediante IAM, puede controlar quién puede crear o administrar los recursos de AWS, el tipo de recursos que se pueden crear y dónde se pueden crear. Esto minimiza la posibilidad de que se creen recursos fuera de la política definida. Utilice los roles y grupos creados anteriormente y asigne políticas de IAM para aplicar el uso correcto. La SCP ofrece un control centralizado de los permisos máximos disponibles para todas las cuentas de su organización, lo que mantiene sus cuentas según las directrices de control de acceso. Las SCP están disponibles solo en una organización que tenga todas las características activadas. Puede configurar las SCP para denegar o permitir acciones en las cuentas de los miembros de forma predeterminada. Para obtener más información sobre la implementación de la administración del acceso, consulte el documento técnico Pilar de seguridad: Marco de Well-Architected.

También se puede implementar la gobernanza mediante la administración de cuotas de servicio de AWS. Si garantiza que las cuotas de servicio se configuran con los gastos generales mínimos y se mantienen correctamente, puede minimizar la creación de recursos que no necesite su organización. Para lograrlo, debe conocer la velocidad con la que pueden cambiar sus requisitos, comprender los proyectos en curso (tanto la creación como la retirada de recursos) y tener en cuenta la rapidez con la que se pueden implementar los cambios de cuota. Las cuotas de servicio se pueden utilizar para aumentar las cuotas cuando sea necesario.

Pasos para la implementación

- Implementación de notificaciones sobre los gastos: utilice las políticas definidas de su organización y cree AWS Budgets para recibir notificaciones cuando los gastos estén fuera de sus políticas. Configure varios presupuestos de costos, uno para cada cuenta, que le notifiquen el gasto global de la cuenta. Configure presupuestos de costos adicionales en cada cuenta para unidades más pequeñas en ella. Estas unidades varían en función de la estructura de la cuenta. Algunos ejemplos comunes son las Regiones de AWS, las cargas de trabajo (mediante etiquetas) o los servicios de AWS. Configure una lista de distribución de correo electrónico como destinatario de las notificaciones y no una cuenta de correo electrónico individual. Puede configurar un

presupuesto real en caso de que se supere una cantidad o utilizar un presupuesto previsto para notificar el uso previsto. También puede preconfigurar acciones presupuestarias de AWS que pueden aplicar políticas de IAM o SCP específicas, o detener las instancias de Amazon EC2 y Amazon RDS de destino. Las acciones presupuestarias se pueden ejecutar automáticamente o requerir la aprobación del flujo de trabajo.

- Implementación de notificaciones sobre los gastos anómalos: use AWS Cost Anomaly Detection para reducir los costos imprevistos de su organización y analizar la causa fundamental de los posibles gastos anómalos. Una vez que haya creado el monitor de costos para identificar los gastos inusuales con el detalle que especifique y haya configurado las notificaciones en AWS Cost Anomaly Detection, le enviará una alerta cuando se detecten gastos inusuales. Esto le permitirá analizar la causa raíz detrás de la anomalía y comprender el impacto en el costo. Utilice las categorías de costos de AWS durante la configuración de AWS Cost Anomaly Detection para identificar qué equipo de proyecto o de unidad de negocio puede analizar la causa raíz del costo inesperado y tomar las medidas necesarias a tiempo.
- Implementación de controles de uso: mediante las políticas definidas de su organización, implemente políticas y roles de IAM para especificar qué acciones pueden hacer los usuarios y cuáles no. En una política de AWS pueden incluirse múltiples políticas de la organización. De la misma manera en que ha definido las políticas, empiece de manera amplia y, a continuación, aplique controles más detallados en cada paso. Los límites de servicio son también un control eficaz del uso. Implemente los límites de servicio correctos en todas las cuentas.

Recursos

Documentos relacionados:

- Políticas administradas de AWS para funciones de trabajo
- Estrategia de facturación de varias cuentas de AWS
- Control access to Regiones de AWS using IAM policies
- AWS Budgets
- AWS Cost Anomaly Detection
- Controle los costos de AWS

Videos relacionados:

- How can I use AWS Budgets to track my spending and usage

Ejemplos relacionados:

- Ejemplos de políticas de administración de acceso de IAM
- Example service control policies
- AWS Acciones de Budgets
- Creación de una política de IAM para controlar el acceso a los recursos de Amazon EC2 mediante etiquetas
- Restricción del acceso de IAM Identity a recursos específicos de Amazon EC2
- Slack integrations for Cost Anomaly Detection using Amazon Q Developer in chat applications

COST02-BP06 Seguimiento del ciclo de vida de los proyectos

Controle, mida y audite el ciclo de vida de los proyectos, equipos y entornos para evitar el uso y el pago de recursos innecesarios.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

Al hacer un seguimiento eficaz del ciclo de vida del proyecto, las organizaciones pueden controlar mejor los costos al planificar, administrar y optimizar los recursos. La información que se obtiene con el seguimiento es muy valiosa para tomar decisiones fundamentadas que contribuyen a la rentabilidad y al éxito general del proyecto.

El seguimiento de todo el ciclo de vida de la carga de trabajo le ayuda a comprender cuándo dejan de ser necesarias las cargas de trabajo o sus componentes. Puede que parezca que se usan las cargas de trabajo y los componentes existentes, pero cuando AWS publica nuevos servicios o características, estos pueden retirarse o adoptarse. Compruebe las etapas anteriores de las cargas de trabajo. Cuando una carga de trabajo ya no está en producción, los entornos previos se pueden retirar o reducirse en gran medida hasta que se necesiten de nuevo.

Puede etiquetar los recursos con un marco de tiempo o un recordatorio para fijar el momento en que se revisó la carga de trabajo. Por ejemplo, si el entorno de desarrollo se revisó por última vez hace meses, podría ser un buen momento para revisarlo de nuevo para analizar si se pueden adoptar nuevos servicios o si el entorno está en uso. Puede agrupar y etiquetar sus aplicaciones con myApplications activado en AWS para gestionar y hacer un seguimiento de los metadatos, como la criticidad, el entorno, la última revisión y el centro de costos. Puede hacer un seguimiento del ciclo de

vida de su carga de trabajo y supervisar y administrar el costo, el estado, la postura de seguridad y el rendimiento de sus aplicaciones.

AWS proporciona varios servicios de administración y gobernanza que puede usar para hacer un seguimiento del ciclo de vida de la entidad. Puede usar AWS Config o AWS Systems Manager para proporcionar un inventario detallado de sus recursos de AWS y su configuración. Se recomienda efectuar una integración con sus sistemas de administración de proyectos o recursos existentes para hacer un seguimiento de los proyectos y productos activos en su organización. Mediante la combinación del sistema actual con el amplio conjunto de eventos y métricas que ofrece AWS, podrá crear una vista de eventos importantes del ciclo de vida y administrar de forma proactiva los recursos a fin de reducir costos innecesarios.

Al igual que en la administración del ciclo de vida de aplicaciones (ALM), el seguimiento del ciclo de vida de un proyecto debe implicar la colaboración de varios procesos, herramientas y equipos, como el diseño y el desarrollo, las pruebas, la producción, el soporte y la redundancia de la carga de trabajo.

Al supervisar cuidadosamente cada fase del ciclo de vida de un proyecto, las organizaciones obtienen información crucial y mejoran el control, lo que facilita la planificación, implementación y finalización exitosas del proyecto. En esta cuidadosa supervisión, se verifica que los proyectos no solo cumplan los estándares de calidad, sino que se entreguen a tiempo y dentro del presupuesto, lo que fomenta el ahorro de costos.

Para obtener más información sobre la implementación del seguimiento del ciclo de vida de las entidades, consulte el documento técnico del Pilar de excelencia operativa de AWS Well-Architected.

Pasos para la implementación

- Establecimiento del proceso de supervisión del ciclo de vida de los proyectos: el equipo del Centro de excelencia en la nube debe establecer el proceso de supervisión del ciclo de vida de los proyectos. Establezca un enfoque estructurado y sistemático para supervisar las cargas de trabajo a fin de mejorar el control, la visibilidad y el rendimiento de los proyectos. Haga que el proceso de supervisión sea transparente y colaborativo y esté centrado en la mejora continua para maximizar su eficacia y valor.
- Revisiones de la carga de trabajo: según lo definan las políticas de su organización, establezca un ritmo regular para auditar sus proyectos existentes y hacer revisiones de la carga de trabajo. El esfuerzo dedicado a la auditoría debería ser proporcional al riesgo, el valor o el costo aproximados de la organización. Las principales áreas que debería incluir en la auditoría son el riesgo de incidente o interrupción en la organización, el valor o la contribución a la organización (medida en

ingresos o reputación de la marca), el costo de la carga de trabajo (medido como costo total de los recursos y costos operativos) y el uso de la carga de trabajo (medido en número de resultados organizativos por unidad de tiempo). Si estas áreas cambian durante el ciclo de vida, se deberá ajustar la carga de trabajo, por ejemplo, mediante una retirada total o parcial.

Recursos

Documentos relacionados:

- Guidance for Tagging on AWS
- ¿Qué es la administración del ciclo de vida de las aplicaciones (ALM)?
- Políticas administradas de AWS para funciones de trabajo

Ejemplos relacionados:

- Control access to Regiones de AWS using IAM policies

Herramientas relacionadas

- AWS Config
- AWS Systems Manager
- AWS Budgets
- AWS Organizations
- AWS CloudFormation

Supervisión del costo y el uso

Permita que los equipos tomen medidas en relación con sus costos y su uso mediante una visibilidad detallada de la carga de trabajo. La optimización de costos comienza con una comprensión detallada del desglose en costo y uso, la capacidad de modelar y pronosticar el gasto, el uso y las características futuros, y la implementación de mecanismos suficientes para alinear el costo y el uso con los objetivos de su organización. A continuación se indican áreas obligatorias para supervisar el costo y el uso:

Prácticas recomendadas

- COST03-BP01 Configuración de los orígenes de información detallados
- COST03-BP02 Incorporación de información de la organización a los costos y el uso
- COST03-BP03 Identificación de las categorías de atribución de costos
- COST03-BP04 Establecimiento de métricas de la organización
- COST03-BP05 Configuración de herramientas de administración de facturación y costos
- COST03-BP06 Asignación de costos según las métricas de la carga de trabajo

COST03-BP01 Configuración de los orígenes de información detallados

Configure herramientas de generación de informes y administración de costos para mejorar el análisis y la transparencia de los datos de costos y uso. Configure la carga de trabajo para crear entradas de registro que faciliten el seguimiento y la división de los costos y el uso.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

La información de facturación detallada, como la especificidad por horas en las herramientas de administración de costos, permite a las organizaciones hacer un seguimiento más detallado de su consumo y les ayuda a identificar algunos de los motivos del aumento de costos. Estos orígenes de datos proporcionan la visión más veraz del costo y uso en toda la organización.

Puede usar Exportaciones de datos de AWS para crear exportaciones del AWS Cost and Usage Report (CUR) 2.0. Esta es la nueva forma recomendada de recibir los datos detallados de costo y uso de AWS. Proporciona especificidad de uso diario o por horas, tarifas, costos y atributos de uso de todos los servicios de AWS de pago (la misma información que el CUR), junto con algunas mejoras. Todas las dimensiones posibles están en el CUR, como, por ejemplo, el etiquetado, la ubicación, los atributos de recursos y los ID de cuentas.

Hay tres tipos de exportación según el tipo de exportación que desee crear: una exportación de datos estándar, una exportación a un panel de costos y uso con la integración de Quick o una exportación de datos heredados.

- Exportación de datos estándar: exportación personalizada de una tabla que se suministra a Amazon S3 de forma periódica.
- Panel de costos y uso: exportación e integración a Quick para implementar un panel de costos y uso prediseñado.

- Exportación de datos heredados: exportación del AWS Cost and Usage Report (CUR) heredado.

Puede crear exportaciones de datos con las siguientes personalizaciones:

- Inclusión de los ID de recurso
- División de los datos de asignación de costos
- Especificidad por horas
- Control de versiones
- Tipo de compresión y formato de archivo

En el caso de cargas de trabajo que ejecuten contenedores en Amazon ECS o Amazon EKS, habilite los datos de asignación de costos divididos para asignar los costos de los contenedores a unidades de negocio y equipos individuales, en función del modo en que las cargas de trabajo de los contenedores consumen los recursos compartidos de computación y memoria. Los datos de asignación de costos divididos también ingresan datos de costos y uso de los nuevos recursos de contenedor en el AWS Cost and Usage Report. Los datos de asignación de costos divididos se calculan mediante el cálculo del costo de los servicios y tareas individuales de ECS que se ejecutan en el clúster.

Un panel de costos y uso exporta la tabla del panel de costos y uso a un bucket de S3 de forma periódica e implementa un panel de costos y uso prediseñado en Quick. Use esta opción si desea implementar rápidamente un panel de sus datos de costos y uso sin que se pueda personalizar.

Si lo desea, puede seguir exportando CUR en modo heredado, donde puede integrar otros servicios de procesamiento, como AWS Glue para preparar los datos para el análisis y ejecutar análisis de datos con Amazon Athena mediante SQL para consultar los datos.

Pasos para la implementación

- Creación de exportaciones de datos: cree exportaciones personalizadas con los datos que desee y controle el esquema de sus exportaciones. Cree exportaciones de datos de administración de facturación y costos mediante SQL básico y visualice dichos datos mediante la integración con Quick. También puede exportar los datos en modo estándar para analizarlos con otras herramientas de procesamiento, como, por ejemplo, Amazon Athena.
- Configuración del informe de costo y uso: configure al menos un informe de costo y uso con la consola de facturación. Configure un informe detallado por horas que incluya todos los

identificadores y los ID de recurso. También puede crear otros informes con distintos niveles de especificidad para proporcionar información resumida de nivel superior.

- Configuración de la especificidad horaria en el Explorador de costos: para acceder a los datos de costo y uso con especificidad de los últimos 14 días, plantéese habilitar los datos por hora y recursos en la consola de facturación.
- Configuración de los registros de la aplicación: verifique que su aplicación registre cada resultado empresarial que ofrece para que se pueda hacer el seguimiento y la medición. Asegúrese de que la especificidad de estos datos sea, como mínimo, por horas, para que coincidan con los datos de costo y uso. Para obtener más información sobre el registro y la supervisión, consulte el pilar de excelencia operativa de Well-Architected.

Recursos

Documentos relacionados:

- Exportaciones de datos de AWS
- AWS Glue
- Quick
- Administración financiera en la nube con AWS
- Etiquetado de recursos de AWS
- Analyzing your costs with Cost Explorer
- Managing AWS Cost and Usage Reports

Ejemplos relacionados:

- AWS Account Setup
- Data Exports for AWS Billing and Cost Management
- AWS Cost Explorer Common Use Cases

COST03-BP02 Incorporación de información de la organización a los costos y el uso

Defina un esquema de etiquetado basado en su organización, los atributos de carga de trabajo y las categorías de asignación de costos para poder filtrar y buscar recursos o supervisar el costo y el uso

en las herramientas de administración de costos. Implemente un etiquetado coherente en todos los recursos, siempre que sea posible, por finalidad, equipo, entorno u otros criterios pertinentes para su empresa.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

Implemente el etiquetado en AWS para agregar información de la organización a sus recursos, que luego se agregará a su información de costo y uso. Una etiqueta es un par clave-valor: la clave está definida y debe ser única en toda la organización, mientras que el valor es único para un grupo de recursos. Un ejemplo de par clave-valor es una clave `Environment`, con un valor `Production`. Todos los recursos del entorno de producción tendrán este par clave-valor. El etiquetado le permite categorizar y controlar los costos con información de la organización relevante y útil. Puede aplicar etiquetas que representen categorías de la organización (como centros de costos, nombres de aplicaciones, proyectos o propietarios) e identificar cargas de trabajo y características de cargas de trabajo (por ejemplo, de prueba o producción) para clasificar sus costos y uso en toda la organización.

Cuando aplica etiquetas a sus recursos de AWS (como instancias Amazon Elastic Compute Cloud o buckets de Amazon Simple Storage Service) y las activa, AWS agrega esta información a los informes de uso y costos. Puede ejecutar informes y análisis en recursos con etiquetas o sin ellas para permitir un mayor cumplimiento de las políticas de administración de costos internos y garantizar una atribución precisa.

Crear e implementar un estándar de etiquetado de AWS en las cuentas de su organización le ayudará a administrar y controlar sus entornos de AWS de manera coherente y uniforme. Utilice las políticas de etiquetas en AWS Organizations para definir reglas sobre cómo usar las etiquetas en los recursos de AWS de sus cuentas en AWS Organizations. Las políticas de etiquetas le permiten adoptar un enfoque estandarizado para los recursos de etiquetado de AWS.

El Editor de etiquetas de AWS le permite agregar, eliminar y administrar etiquetas de varios recursos. Con el editor de etiquetas, busque los recursos que desee etiquetar y, a continuación, administre las etiquetas de los recursos que aparecen en los resultados de la búsqueda.

Categorías de costos de AWS le permite asignar un significado organizativo a sus costos, sin necesidad de etiquetar los recursos. Puede asignar la información de costos y uso a estructuras organizativas internas únicas. Debe definir reglas de categorías para asignar y categorizar los costos mediante dimensiones de facturación, como cuentas y etiquetas. Esto proporciona otro nivel de

capacidad de administración, además del etiquetado. También puede asignar cuentas específicas y etiquetas a varios proyectos.

Pasos para la implementación

- Definición de un esquema de etiquetado: reúna a todas las partes interesadas de su empresa para definir un esquema. En general, son personas con cargos técnicos, financieros o administrativos. Defina una lista de etiquetas que deben tener todos los recursos, así como una lista de las etiquetas que deberían tener los recursos. Compruebe que los nombres y los valores de las etiquetas sean coherentes en toda la organización.
- Etiquete recursos: se las categorías de atributos de costos definidas para colocar etiquetas en todos los recursos en las cargas de trabajo según las categorías. Use herramientas como la CLI, el Editor de etiquetas o AWS Systems Manager para incrementar la eficiencia.
- Implementación de Categorías de costos de AWS: puede crear Categorías de costos sin implementar el etiquetado. Las categorías de costos usan las dimensiones de costos y uso existentes. Cree reglas de categorías a partir de su esquema e impleméntelas en las categorías de costos.
- Etiquetado automático: para comprobar que mantiene altos niveles de etiquetado en todos los recursos, automatice el etiquetado para que los recursos reciban etiquetas automáticamente en cuanto se creen. Utilice servicios como AWS CloudFormation para verificar que los recursos estén etiquetados al crearlos. También puede crear una solución personalizada para etiquetar automáticamente con funciones de Lambda o usar un microservicio personalizado que escanee la carga de trabajo periódicamente y elimine cualquier recurso que no tenga etiqueta, lo que es ideal para los entornos de prueba y de desarrollo.
- Supervisión e informes mediante el etiquetado: para garantizar que mantiene altos niveles de etiquetado en toda la organización, reporte y supervise las etiquetas de sus cargas de trabajo. Puede usar el AWS Cost Explorer para ver el costo de los recursos etiquetados o sin etiquetar, o bien usar servicios como el Editor de etiquetas. Revise periódicamente el número de recursos no etiquetados y agregue etiquetas hasta alcanzar el nivel de etiquetado que desee.

Recursos

Documentos relacionados:

- Tagging Best Practices
- AWS CloudFormation Resource Tag

- AWS Cost Categories
- Etiquetado de recursos de AWS
- Analyzing your costs with AWS Budgets
- Analyzing your costs with Cost Explorer
- Managing AWS Cost and Usage Reports

Videos relacionados:

- How can I tag my AWS resources to divide up my bill by cost center or project
- Tagging AWS Resources

COST03-BP03 Identificación de las categorías de atribución de costos

Identifique las categorías de la organización como las unidades empresariales, los departamentos o los proyectos que podrían utilizarse para asignar los costos dentro de su organización a las entidades consumidoras internas. Utilice esas categorías para imponer la responsabilidad del gasto, concienciar sobre los costos y fomentar comportamientos de consumo eficaces.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

El proceso de categorización de los costos es crucial en la elaboración de presupuestos, la contabilidad, los informes financieros, la toma de decisiones, las evaluaciones comparativas y la administración de proyectos. Al clasificar y categorizar los gastos, los equipos pueden comprender mejor los tipos de costos en los que incurrirán durante su traspaso a la nube, lo que les ayuda a tomar decisiones fundamentadas y a administrar los presupuestos de manera eficaz.

La responsabilidad de los gastos en la nube es un buen incentivo para conseguir una administración disciplinada de la demanda y los costos. Como resultado, las organizaciones que destinan la mayor parte de su gasto en la nube a unidades empresariales o equipos que consumen recursos ahorran mucho más en costos en la nube. Además, la asignación del gasto en la nube ayuda a las organizaciones a adoptar más prácticas recomendadas de gobernanza en la nube centralizada.

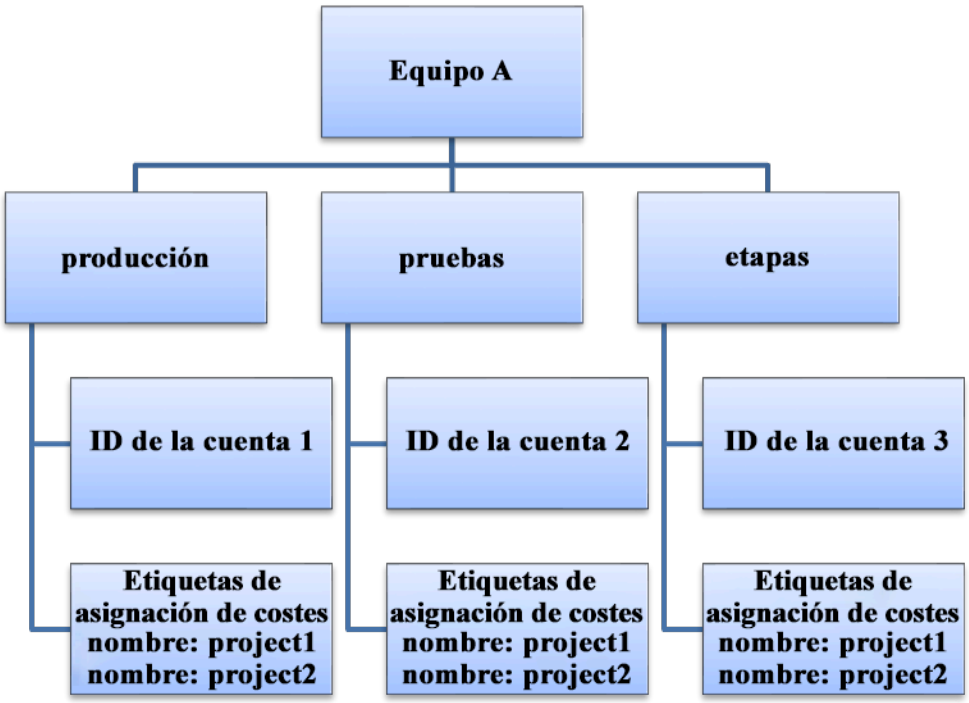
Colabore con el equipo financiero y otras partes interesadas pertinentes para comprender los requisitos sobre cómo deben asignarse los costos en la organización durante las llamadas de cadencia periódicas. Los costos de las cargas de trabajo deben asignarse a todo el ciclo de

vida, como las fases de desarrollo, pruebas, producción y retirada. Debe saber qué costos de la organización proceden de la formación, el desarrollo del personal y la creación de ideas. Puede ser útil para asignar correctamente las cuentas que se usan para los presupuestos de formación y desarrollo, en lugar de presupuestos genéricos de costos de TI.

Tras definir las categorías de atribución de costos con las partes interesadas de la organización, utilice Categorías de costos de AWS para agrupar la información de costos y uso en categorías significativas en Nube de AWS, como el costo de un proyecto específico o Cuentas de AWS de departamentos o unidades de negocio. Puede crear categorías personalizadas y asignar su información de costos y uso a estas categorías en función de las reglas que defina mediante varias dimensiones, tales como: cuenta, etiqueta, servicio o tipo de cargo. Tras configurar las categorías de costos, puede ver la información de costos y uso por estas categorías, lo que permite a la organización tomar mejores decisiones estratégicas y de compra. Estas categorías también se pueden ver en el AWS Cost Explorer, AWS Budgets y el AWS Cost and Usage Report.

Por ejemplo, cree categorías de costos para sus unidades empresariales (equipo de DevOps) y, en cada categoría, cree varias reglas (para cada subcategoría) con múltiples dimensiones (Cuentas de AWS, etiquetas de asignación de costos, servicios o tipo de cargo) basadas en las agrupaciones definidas. Con Cost Categories, puede organizar los costos usando un motor basado en reglas. Las reglas que configura organizan los costos en categorías. En estas reglas, puede llevar a cabo el filtrado mediante varias dimensiones para cada categoría, como Cuentas de AWS, servicios de AWS o tipos de cargos específicos. Puede utilizar estas categorías en los múltiples productos de la consola de Administración de costos y Administración de facturación y costos de AWS. Esto incluye AWS Cost Explorer, AWS Budgets, AWS Cost and Usage Report y AWS Cost Anomaly Detection.

En el siguiente diagrama se muestra, a modo de ejemplo, cómo agrupar la información de costos y uso de la organización si tiene varios equipos (categoría de costos) con varios entornos (reglas) y cada entorno tiene varios recursos o activos (dimensiones).



Organigrama de costos y uso

También puede crear agrupaciones de costos mediante categorías de costos. Después de crear las categorías de costos (deje que transcurran hasta 24 horas desde la creación de una categoría de costos para que sus registros de uso se actualicen con valores), estas aparecen en AWS Cost Explorer, AWS Budgets, AWS Cost and Usage Report y AWS Cost Anomaly Detection. En AWS Cost Explorer y AWS Budgets, una categoría de costo aparece como una dimensión de facturación adicional. Puede utilizarla para aplicar filtros para el valor de la categoría de costos específico o para efectuar agrupaciones por categorías de costos.

Pasos para la implementación

- Definición de las categorías de la organización: reúnase con las partes interesadas internas y las unidades empresariales para definir las categorías que reflejen la estructura y los requisitos de su organización. Estas categorías deben reflejar directamente la estructura de las categorías financieras existentes, como unidad empresarial, presupuestaria, centro de costos o departamento. Consulte los resultados de la nube para su empresa, como la formación o la educación, pues también son categorías de la organización.
- Definición de las categorías funcionales: reúnase con las partes interesadas internas y las unidades empresariales para definir las categorías que reflejen las funciones de su empresa.

Pueden ser los nombres de las aplicaciones o las cargas de trabajo y el tipo de entorno, como producción, pruebas o desarrollo.

- Definición de categorías de costos de AWS: cree categorías de costos para organizar la información de costos y uso mediante Categorías de costos de AWS y asigne el costo y el uso de AWS a categorías significativas. Se pueden asignar varias categorías a un recurso y un recurso puede estar en muchas categorías distintas, por lo que se recomienda definir tantas categorías como sea necesario para que pueda administrar sus costos en la estructura de categorías con Categorías de costos de AWS.

Recursos

Documentos relacionados:

- Etiquetado de recursos de AWS
- Uso de etiquetas de asignación de costos de
- Analyzing your costs with AWS Budgets
- Analyzing your costs with Cost Explorer
- Managing AWS Cost and Usage Reports
- AWS Cost Categories
- Administración de costos con AWS Cost Categories
- Creating cost categories
- Tagging cost categories
- Splitting charges within cost categories
- AWS Características de las categorías de costos de

Ejemplos relacionados:

- Organize your cost and usage data with AWS Cost Categories
- Administración de costos con AWS Cost Categories

COST03-BP04 Establecimiento de métricas de la organización

Establezca las métricas de la organización necesarias para esta carga de trabajo. Algunos ejemplos de métricas de cargas de trabajo son los informes de clientes producidos o las páginas web que se entregan a los clientes.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Comprenda cómo se mide el rendimiento de su carga de trabajo en relación con el éxito empresarial. Cada carga de trabajo suele tener un pequeño conjunto de resultados principales que indican el rendimiento. Si tiene una carga de trabajo compleja con muchos componentes, puede priorizar la lista o definir las métricas de cada componente y hacer un seguimiento de ellas. Colabore con sus equipos para entender qué métricas debe utilizar. Esta unidad se usará para comprender la eficiencia de la carga de trabajo o el costo de cada resultado empresarial.

Pasos para la implementación

- Definición de los resultados de las cargas de trabajo: reúnase con las partes interesadas de la empresa y defina los resultados de las cargas de trabajo. Son una medida principal del uso de los clientes y deben ser métricas empresariales y no técnicas. Debe haber un pequeño número de métricas generales (menos de cinco) por carga de trabajo. Si la carga de trabajo produce varios resultados para diferentes casos de uso, agrúpelos en una sola métrica.
- Definición de los resultados de los componentes de las cargas de trabajo: de manera opcional, si tiene una carga de trabajo grande y compleja, o puede dividir fácilmente su carga de trabajo en componentes (como microservicios) con entradas y salidas bien definidas, establezca métricas para cada componente. El esfuerzo debe reflejar el valor y el costo del componente. Empiece por los componentes más grandes y continúe con los más pequeños.

Recursos

Documentos relacionados:

- Etiquetado de recursos de AWS
- Analyzing your costs with AWS Budgets
- Analyzing your costs with Cost Explorer

- Managing AWS Cost and Usage Reports

COST03-BP05 Configuración de herramientas de administración de facturación y costos

Configure las herramientas de administración de costos que cumplan las políticas de su organización para administrar y optimizar el gasto en la nube. Esto incluye servicios, herramientas y recursos para organizar los datos de costos y uso y hacer un seguimiento de ellos, mejorar el control mediante una facturación consolidada y permisos de acceso, mejorar la planificación mediante presupuestos y previsiones, recibir notificaciones o alertas y reducir los costos con optimizaciones de recursos y precios.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Para establecer una responsabilidad sólida, analice primero la estrategia de cuentas como parte de la estrategia de asignación de costos. Si lo hace bien, es posible que no necesite nada más. En caso contrario, puede haber desconocimiento y problemas adicionales.

Para fomentar la responsabilidad del gasto en la nube, conceda a los usuarios acceso a herramientas que les proporcionen visibilidad de sus costos y uso. AWS recomienda configurar todas las cargas de trabajo y equipos para los siguientes fines:

- Organización: establezca su base de referencia de asignación de costos y gobernanza con su propia estrategia de etiquetado y taxonomía. Cree varias cuentas de AWS con herramientas como AWS Control Tower o AWS Organization. Etiquete los recursos que admite AWS y clasifíquelos de una forma fácil de entender en función de la estructura de su organización (unidades empresariales, departamentos o proyectos). Etiquete los nombres de las cuentas para centros de costos específicos y asígnelos a Categorías de costos de AWS para agrupar las cuentas de las unidades empresariales con sus centros de costos, de modo que el propietario de la unidad empresarial pueda ver el consumo de varias cuentas en un solo lugar.
- Acceso: haga un seguimiento de la información de facturación de toda la organización en la facturación unificada. Verifique que las partes interesadas y los propietarios empresariales adecuados tengan acceso a ella.
- Control: cree mecanismos de gobernanza efectivos con las barreras de protección adecuadas para evitar escenarios inesperados cuando se utilicen las políticas de control de servicio (SCP),

las políticas de etiquetas, las políticas de IAM y las alertas de presupuestos. Por ejemplo, puede permitir que los equipos creen recursos específicos solo en las regiones que prefieran mediante mecanismos de control eficaces e impedir la creación de recursos sin una etiqueta específica (como el centro de costos).

- Estado actual: configure un panel que muestre los niveles actuales de costo y uso. El panel debe estar disponible en un lugar muy visible del entorno de trabajo, de forma similar a un panel de operaciones. Puede exportar los datos y utilizar el Panel de costos y uso del Centro de optimización de costos de AWS o cualquier producto compatible para conseguir esta visibilidad. Es posible que tenga que crear diferentes paneles para diferentes perfiles. Por ejemplo, el panel de administrador podría ser diferente al panel de ingeniero.
- Notificaciones: envíe notificaciones cuando el costo o el uso superen los límites definidos y se produzcan anomalías con AWS Budgets o la Detección de anomalías en los costos de AWS.
- Informes: resuma toda la información sobre costos y uso. Aumente la concienciación y la responsabilidad de su gasto en la nube con datos de costos detallados y atribuibles. Cree informes que sean relevantes para el equipo que los consume y que contengan recomendaciones.
- Seguimiento: muestre el costo y uso actuales con respecto a los objetivos o las metas configurados.
- Análisis: permita a los miembros del equipo ejecutar análisis detallados y personalizados con diferentes filtros (recurso, cuenta, etiqueta, etc.) con distintos filtros (recurso, cuenta, etiqueta, etc.).
- Inspección: manténgase al día de las oportunidades de implementación de recursos y optimización de costos. Reciba notificaciones mediante Amazon CloudWatch, Amazon SNS o Amazon SES para las implementaciones de recursos en la organización. Revise las recomendaciones de optimización de costos con AWS Trusted Advisor o AWS Compute Optimizer.
- Informes de tendencias: muestre la variabilidad del costo y uso durante el periodo requerido con el nivel de detalle necesario.
- Previsiones: muestre los costos futuros estimados y calcule el uso de sus recursos y el gasto con paneles de previsión que puede crear manualmente.

Puede utilizar el Centro de optimización de costos de AWS para conocer las posibles oportunidades de ahorro de costos consolidadas desde una ubicación centralizada y crear exportaciones de datos para su integración con Amazon Athena. También puede usar el Centro de optimización de costos de AWS para implementar el Panel de costos y uso, que utiliza Quick para ejecutar un análisis de costos interactivo y un intercambio seguro de información sobre los costos.

Si no tiene las habilidades o el ancho de banda esenciales en su organización, puede trabajar con AWS ProServ, AWS Managed Services (AMS) o socios de AWS. También puede usar herramientas de terceros, pero asegúrese de validar la propuesta de valor.

Pasos para la implementación

- Concesión del acceso a las herramientas por equipo: configure sus cuentas y cree grupos que tengan acceso a los informes de costos y uso necesarios para su consumo y utilice AWS Identity and Access Management para controlar el acceso a herramientas como el AWS Cost Explorer. Estos grupos deben incluir representantes de todos los equipos que posean o administren una aplicación. De este modo, se certifica que cada equipo tiene acceso a su información de costos y uso para hacer el seguimiento de su consumo.
- Organización de las etiquetas y categorías de costos: organice sus costos entre equipos, unidades empresariales, aplicaciones, entornos y proyectos. Use etiquetas de recursos para organizar los costos por etiquetas de asignación de costos. Cree categorías de costos en función de las dimensiones mediante etiquetas, cuentas, servicios, etc., para asignar los costos.
- Configuración de AWS Budgets: configure AWS Budgets en todas las cuentas para sus cargas de trabajo. Establezca presupuestos para el gasto general de la cuenta y presupuestos para las cargas de trabajo con etiquetas y categorías de costos. Configure las notificaciones en AWS Budgets para recibir alertas cuando supere los importes presupuestados o cuando los costos estimados superen sus presupuestos.
- Configuración de la Detección de anomalías en los costos de AWS: use la Detección de anomalías en los costos de AWS para sus cuentas, servicios básicos o categorías de costos que haya creado para supervisar el costo y el uso, y detectar gastos fuera de lo habitual. Puede recibir las alertas individualmente en informes agregados y en un correo electrónico o un tema de Amazon SNS que le permita analizar y determinar la causa raíz de la anomalía, e identificar el factor que está provocando el aumento de los costos.
- Uso de herramientas de análisis de costos: configure AWS Cost Explorer para su carga de trabajo y cuentas para visualizar los datos de costos y hacer un análisis posterior. Cree un panel para la carga de trabajo que haga un seguimiento del gasto general, las métricas clave de uso de la carga de trabajo y la previsión de los costos futuros a partir de sus datos históricos de costos.
- Utilice herramientas de análisis para ahorrar costos: utilice Centro de optimización de costos de AWS para identificar las oportunidades de ahorro con recomendaciones personalizadas, como eliminar los recursos no utilizados, ajustar el tamaño, Savings Plans, hacer reservas y hacer recomendaciones para optimizar el cálculo.

- Configure herramientas avanzadas: si lo desea, puede crear imágenes para facilitar el análisis interactivo y el intercambio de información sobre los costos. Con la exportación de datos del Centro de optimización de costos de AWS, puede crear un Panel de costos y uso con tecnología de Quick para la organización que aporte detalles y grado de detalle adicional. También puede implementar funciones de análisis avanzadas mediante el uso de exportaciones de datos en Amazon Athena para consultas avanzadas y crear paneles en Quick. Trabaje con socios de AWS para adoptar soluciones de administración en la nube para la supervisión y la optimización consolidadas de las facturas en la nube.

Recursos

Documentos relacionados:

- What is Administración de facturación y costos de AWS and Cost Management?
- Establecer su entorno de acuerdo a las prácticas recomendadas de AWS
- Best Practices for Tagging AWS Resources
- Etiquetado de los recursos de AWS
- AWS Cost Categories
- Analyzing your costs with AWS Budgets
- Analyzing your costs with AWS Cost Explorer
- What is AWS Data Exports?

Videos relacionados:

- Deploying Cloud Intelligence Dashboards
- Get Alerts on any FinOps or Cost Optimization Metric or KPI

Ejemplos relacionados:

- Panel de costos y uso con tecnología de Quick
- AWS Cost and Usage Governance Workshop

COST03-BP06 Asignación de costos según las métricas de la carga de trabajo

Asigne los costos de la carga de trabajo según las métricas de uso o los resultados empresariales para medir la eficiencia de los costos. Implemente un proceso para analizar los datos de costos y uso con servicios de análisis que pueden proporcionar información y capacidad de recuperación.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

La optimización de costos significa aportar resultados empresariales al menor precio, lo que solo se puede conseguir asignando los costos de la carga de trabajo por métricas de carga de trabajo (medidas por eficiencia de la carga de trabajo). Supervise las métricas de la carga de trabajo definidas mediante archivos de registro u otro tipo de supervisión de la aplicación. Combine estos datos con los costos de la carga de trabajo, que pueden obtenerse al consultar los costos que tienen un valor de etiqueta o identificador de cuenta específicos. Lleve a cabo este análisis en el nivel de hora. Por lo general, su eficiencia cambia si tiene componentes de costos estáticos (por ejemplo, una base de datos backend que se ejecuta permanentemente) con un índice de solicitudes variable (por ejemplo, el uso alcanza su punto máximo entre las nueve de la mañana y las cinco de la tarde, pero hay pocas solicitudes por la noche). Comprender la relación entre los costos variables y fijos le ayuda a centrar sus actividades de optimización.

Crear métricas de carga de trabajo para recursos compartidos puede ser difícil en comparación con recursos como las aplicaciones en contenedores de Amazon Elastic Container Service (Amazon ECS) y Amazon API Gateway. Sin embargo, hay ciertas formas de clasificar el uso y hacer un seguimiento de los costos. Si necesita hacer un seguimiento de los recursos compartidos de Amazon ECS y AWS Batch, puede habilitar los datos de asignación de costos divididos en AWS Cost Explorer. Al dividir los datos de asignación de costos, puede comprender y optimizar el costo y el uso de sus aplicaciones en contenedores y volver a asignar los costos de las aplicaciones a entidades empresariales individuales en función de cómo se consumen los recursos compartidos de computación y memoria.

Pasos para la implementación

- Asignación de costos a métricas de cargas de trabajo: use las métricas definidas y las etiquetas configuradas, y cree una métrica que combine el resultado de la carga de trabajo y el costo de la carga de trabajo. Utilice servicios de análisis como Amazon Athena y Amazon Quick para crear un panel de eficiencia para la carga de trabajo general y cualquier componente.

Recursos

Documentos relacionados:

- Etiquetado de recursos de AWS
- Analyzing your costs with AWS Budgets
- Analyzing your costs with Cost Explorer
- Managing AWS Cost and Usage Reports

Ejemplos relacionados:

- Improve cost visibility of Amazon ECS and AWS Batch with AWS Split Cost Allocation Data

Retirada de recursos

Después de administrar una lista de proyectos, empleados y recursos tecnológicos a lo largo del tiempo, puede identificar recursos que ya no se utilizan y proyectos que ya no tienen propietario.

Prácticas recomendadas

- COST04-BP01 Seguimiento de los recursos a lo largo de su ciclo de vida
- COST04-BP02 Implementación de un proceso de retirada
- COST04-BP03 Retirada de recursos
- COST04-BP04 Retirada automática de los recursos
- COST04-BP05 Aplicación de políticas de retención de datos

COST04-BP01 Seguimiento de los recursos a lo largo de su ciclo de vida

Defina e implemente un método para hacer un seguimiento de los recursos y sus asociaciones con los sistemas a lo largo de su ciclo de vida. Puede usar etiquetas para identificar la carga de trabajo o la función del recurso.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Retire los recursos de la carga de trabajo que ya no necesite. Por ejemplo, después de hacer pruebas, los recursos empleados se pueden eliminar. El seguimiento de los recursos con etiquetas

(y la ejecución de informes en dichas etiquetas) puede ayudarle a identificar los elementos que se deben eliminar, ya que no estarán en uso o caducará su licencia. Usar etiquetas es una forma efectiva de hacer un seguimiento de los recursos. Se puede etiquetar el recurso con su función o una fecha conocida en la que se puede retirar. Puede ejecutar informes en estas etiquetas. Los valores de ejemplo para el etiquetado de características son `feature-X testing` para identificar la finalidad del recurso en términos del ciclo de vida de la carga de trabajo. Otro ejemplo es usar `LifeSpan` o `TTL` para los recursos, como el nombre y el valor de la clave de la etiqueta que se va a eliminar, para definir el periodo o el tiempo específico de retirada.

Pasos para la implementación

- Implementación de un esquema de etiquetado: implemente un esquema de etiquetado que identifique la carga de trabajo a la que pertenece el recurso y verifique que todos los recursos de la carga de trabajo estén etiquetados en consecuencia. El etiquetado le ayuda a categorizar los recursos por finalidad, equipo, entorno u otros criterios pertinentes para su empresa. Para obtener más información sobre los casos de uso, las estrategias y las técnicas de etiquetado, consulte Prácticas recomendadas para el etiquetado de los recursos de AWS.
- Implementación de la supervisión del rendimiento o la producción de la carga de trabajo: implemente la supervisión o las alarmas del rendimiento de la carga de trabajo para que se inicien en las solicitudes de entrada o en las terminaciones de salida. Configúrela para que proporcione notificaciones cuando las solicitudes de carga de trabajo o los resultados lleguen a cero, lo que significa que ya no se usan los recursos de la carga de trabajo. Incorpore un factor de tiempo si la carga de trabajo baja a cero de forma periódica en condiciones normales. Para obtener más información sobre los recursos no utilizados o infrautilizados, consulte AWS Trusted Advisor Cost Optimization checks.
- Agrupación de recursos de AWS: cree grupos de recursos para sus recursos de AWS. Puede utilizar Grupos de recursos de AWS para organizar y administrar los recursos de AWS que se encuentran en la misma Región de AWS. Puede agregar etiquetas a la mayoría de sus recursos como ayuda para identificarlos y clasificarlos en su organización. Utilice el editor de etiquetas para agregar etiquetas a los recursos compatibles de forma masiva. Considere la posibilidad de utilizar AWS Service Catalog para crear, administrar y distribuir carteras de productos aprobados para los usuarios finales y administrar el ciclo de vida del producto.

Recursos

Documentos relacionados:

- AWS Auto Scaling
- AWS Trusted Advisor
- AWS Trusted Advisor Cost Optimization Checks
- Etiquetado de recursos de AWS
- Publicar métricas personalizadas

Videos relacionados:

- How to optimize costs using AWS Trusted Advisor

Ejemplos relacionados:

- ¿Cómo organizo mis recursos de AWS?
- Optimización de costos con AWS Trusted Advisor

COST04-BP02 Implementación de un proceso de retirada

Implemente un proceso para identificar y retirar los recursos sin usar.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Implemente un proceso estandarizado en toda la organización para identificar y eliminar los recursos que no se utilizan. El proceso debería definir la frecuencia con que se hacen las búsquedas y los procesos para retirar el recurso para verificar que se cumplan todos los requisitos de la organización.

Pasos para la implementación

- Creación e implementación de un proceso de retirada: trabaje con los desarrolladores y propietarios de las cargas de trabajo para diseñar un proceso de retirada de la carga de trabajo y sus recursos. El proceso debería incluir un método para verificar si se usa la carga de trabajo y también si se usa cada recurso de la carga de trabajo. Detalle los pasos necesarios para retirar el recurso del servicio a la vez que garantiza el cumplimiento de cualquier requisito normativo. Se debe incluir cualquier recurso asociado, como licencias o almacenamiento asociado. Notifique a los propietarios de las cargas de trabajo que se ha iniciado el proceso de retirada.

  Siga estos pasos de retirada como guía sobre lo que se debe comprobar como parte del proceso:

- Identificación de los recursos que se van a retirar: identifique los recursos que son aptos para la retirada de su Nube de AWS. Registre toda la información necesaria y programe la retirada. En su cronología, asegúrese de tener en cuenta si surgen (y cuándo surgen) problemas inesperados durante el proceso.
- Coordinación y comunicación: trabaje con los propietarios de la carga de trabajo para confirmar el recurso que se va a retirar
- Registro de los metadatos y creación de copias de seguridad: registre los metadatos (como las IP públicas, la región, la AZ, la VPC, la subred y los grupos de seguridad) y cree copias de seguridad (como las instantáneas de Amazon Elastic Block Store o la captura de AMI, la exportación de claves y la exportación de certificados) si es necesario para los recursos del entorno de producción o si son recursos críticos.
- Validación de la infraestructura como código: determine si los recursos se implementaron con CloudFormation, Terraform, Kit de desarrollo de la nube de AWS (AWS CDK) o cualquier otra herramienta de implementación de infraestructura como código para poder volver a implementarlos si es necesario.
- Prevención del acceso: aplique controles restrictivos durante un periodo de tiempo para evitar el uso de los recursos y, al mismo tiempo, determinar si el recurso es necesario. Verifique que el entorno del recurso se pueda revertir a su estado original si es necesario.
- Ejecución del proceso de desmantelamiento interno: siga las tareas administrativas y el proceso de retirada de su organización, como eliminar el recurso del dominio de la organización, eliminar el registro de DNS y eliminar el recurso de la herramienta de administración de la configuración, la herramienta de supervisión, la herramienta de automatización y las herramientas de seguridad.

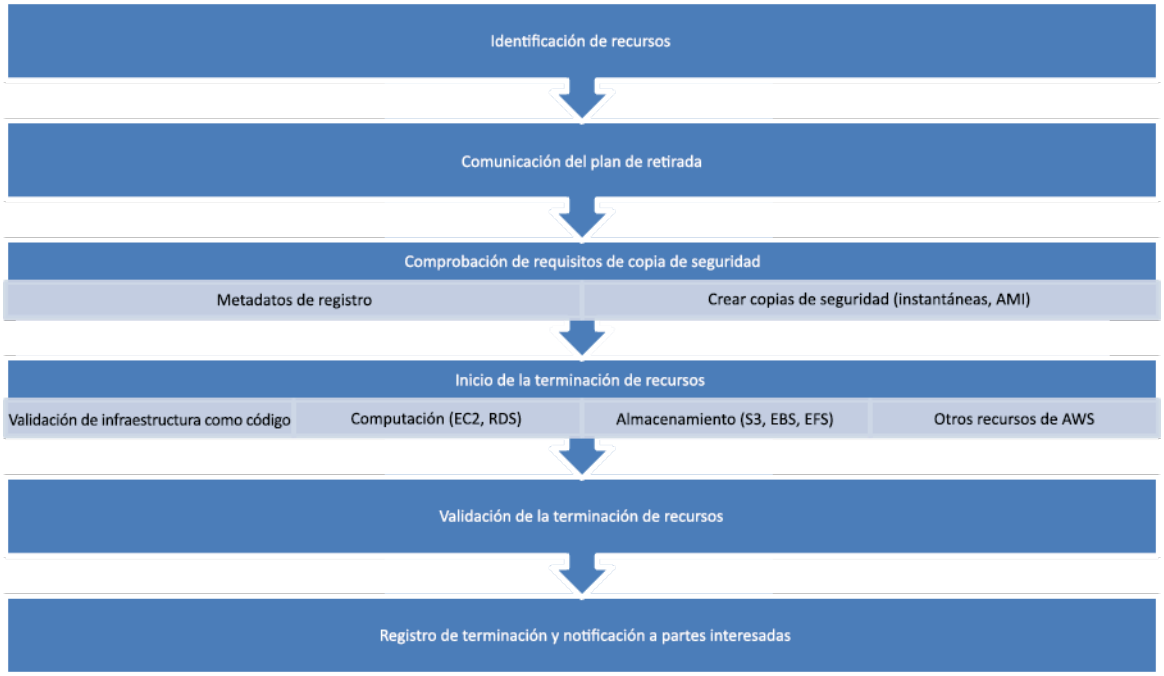
Si el recurso es una instancia de Amazon EC2, consulte la siguiente lista. Para obtener más información, consulte ¿Cómo puedo eliminar o terminar mis recursos de Amazon EC2?

- Detenga o termine todos sus equilibradores de carga e instancias de Amazon EC2. Las instancias de Amazon EC2 son visibles en la consola por poco tiempo después de su terminación. No se facturan las instancias que no se encuentran en estado de ejecución.
- Elimine la infraestructura de escalado automático.
- Libere todos los host dedicados.
- Elimine todos los volúmenes de Amazon EBS y las instantáneas de Amazon EBS.
- Libere todas las direcciones IP elásticas.
- Anule el registro de todas las imágenes de máquina de Amazon (AMI).

- Termine todos los entornos de AWS Elastic Beanstalk.

Si el recurso es un objeto en el almacenamiento de Amazon Glacier y si elimina un archivo antes de cumplir la duración de almacenamiento mínima, se le cobrará una tarifa prorrateada por eliminación anticipada. La duración de almacenamiento mínima de Amazon Glacier depende de la clase de almacenamiento utilizada. Para obtener un resumen de la duración mínima de almacenamiento para cada clase de almacenamiento, consulte Rendimiento en las clases de almacenamiento de Amazon S3. Para obtener información sobre cómo se calculan las tarifas por eliminación anticipada, consulte Precios de Amazon S3.

En el sencillo diagrama de flujo del proceso de retirada que figura a continuación se describen las etapas de retirada. Antes de retirar los recursos, verifique que la organización no use los que ha identificado para su retirada.



Flujo de retirada de recursos.

Recursos

Documentos relacionados:

- AWS Auto Scaling
- AWS Trusted Advisor
- AWS CloudTrail

Videos relacionados:

- Delete CloudFormation stack but retain some resources
- Find out which user launched Amazon EC2 instance

Ejemplos relacionados:

- ¿Cómo puedo eliminar o terminar mis recursos de Amazon EC2?
- ¿Cómo puedo saber qué usuario ha iniciado una instancia de Amazon EC2 en mi cuenta?

COST04-BP03 Retirada de recursos

Retire los recursos que algunos eventos inician, como las auditorías periódicas o los cambios en el uso. La retirada se suele llevar a cabo periódicamente y es manual o automática.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

La frecuencia y el esfuerzo dedicados a buscar recursos que no se utilizan deberían reflejar el ahorro potencial, de manera que una cuenta con pocos costos debería analizarse con menos frecuencia que una cuenta con costos mayores. Las búsquedas y los eventos de retirada pueden iniciarse por cambios de estado de la carga de trabajo, como el fin de la vida útil de un producto o su reemplazo. También pueden iniciarse por eventos externos, como cambios en las condiciones de mercado o la finalización de un producto.

Pasos para la implementación

- Recursos de retirada: se trata de la fase de amortización de los recursos de AWS que ya no se necesitan o que tienen un contrato de licencia que está a punto de finalizar. Complete todas las comprobaciones finales hechas antes de pasar a la fase de eliminación y retirada de recursos para evitar interrupciones no deseadas, como la captura de instantáneas o la creación de copias de seguridad. Use el proceso de retirada para retirar los recursos identificados como no utilizados.

Recursos

Documentos relacionados:

- AWS Auto Scaling
- AWS Trusted Advisor

COST04-BP04 Retirada automática de los recursos

Diseñe su carga de trabajo para que gestione de manera sencilla la finalización de recursos a medida que identifica y retira recursos que no son críticos, recursos innecesarios o recursos con poco uso.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

Use la automatización para reducir o eliminar los costos asociados al proceso de retirada. El diseño de la carga de trabajo para que aplique procesos de retirada automáticos reducirá los costos generales de la carga de trabajo durante su vida. Puede utilizar Amazon EC2 Auto Scaling o Escalado automático de aplicaciones para llevar a cabo el proceso de retirada. También puede implementar código personalizado mediante la API o el SDK para retirar automáticamente los recursos de la carga de trabajo.

Las aplicaciones modernas se crean primero sin servidor, una estrategia que prioriza la adopción de servicios sin servidor. AWS desarrolló servicios sin servidor para los tres niveles de su pila: computación, integración y almacenes de datos. El uso de la arquitectura sin servidor le permitirá ahorrar costos durante periodos de poco tráfico, con escalado y desescalado verticales de forma automática.

Pasos para la implementación

- Implementación de Amazon EC2 Auto Scaling o Escalado automático de aplicaciones: para los recursos compatibles, configúrelos con Amazon EC2 Auto Scaling o Escalado automático de aplicaciones. Estos servicios pueden ayudarle a optimizar el uso y la rentabilidad a la hora de consumir servicios de AWS. Cuando baje la demanda, estos servicios eliminarán automáticamente cualquier exceso de capacidad de recursos para evitar un gasto excesivo.
- Configuración de CloudWatch para finalizar instancias: las instancias se pueden configurar para que finalicen mediante alarmas de CloudWatch. Use las métricas del proceso de retirada para implementar una alarma con una acción de Amazon Elastic Compute Cloud. Verifique la operación en un entorno que no sea de producción antes de la implementación.

- Implementación del código dentro de la carga de trabajo: puede usar el AWS SDK o la AWS CLI para retirar los recursos de la carga de trabajo. Implemente código en la aplicación que se integre con AWS y finalice o elimine recursos que ya no se usan.
- Uso de servicios sin servidor: priorice la creación de arquitecturas sin servidor y una arquitectura basada en eventos en AWS para crear y ejecutar sus aplicaciones. AWS ofrece varios servicios de tecnología sin servidor que, de forma inherente, optimizan automáticamente el uso de los recursos y automatizan la retirada (reducción y escalado horizontales). Con las aplicaciones sin servidor, el uso de los recursos se optimiza automáticamente y nunca pagará por un exceso de aprovisionamiento.

Recursos

Documentos relacionados:

- Amazon EC2 Auto Scaling
- Getting Started with Amazon EC2 Auto Scaling
- Aplicación de escalado automático
- AWS Trusted Advisor
- Sin servidor en AWS
- Creación de alarmas para parar, terminar, reiniciar o recuperar una instancia
- Agregación de acciones de terminación a las alarmas de Amazon CloudWatch

Ejemplos relacionados:

- Programación de la eliminación automática de las pilas de AWS CloudFormation

COST04-BP05 Aplicación de políticas de retención de datos

Defina políticas de retención de datos en los recursos admitidos para gestionar la eliminación de objetos según los requisitos de su organización. Identifique y elimine los recursos y objetos innecesarios o huérfanos que ya no sean necesarios.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Utilice las políticas de retención de datos y de ciclo de vida para reducir los costos asociados al proceso de retirada y los costos de almacenamiento de los recursos identificados. Definir sus

políticas de retención de datos y de ciclo de vida para llevar a cabo la migración y eliminación automatizadas de clases de almacenamiento reducirá los costos generales de almacenamiento durante su vida útil. Puede utilizar Amazon Data Lifecycle Manager para automatizar la creación y eliminación de instantáneas de Amazon Elastic Block Store e imágenes de máquina de Amazon (AMI) basadas en Amazon EBS, y utilizar Amazon S3 Intelligent-Tiering o una configuración del ciclo de vida de Amazon S3 para administrar el ciclo de vida de sus objetos de Amazon S3. También puede implementar código personalizado mediante la API o el SDK para crear políticas de ciclo de vida y reglas de políticas para que los objetos se eliminen automáticamente.

Pasos para la implementación

- Uso de Amazon Data Lifecycle Manager: utilice políticas de ciclo de vida en Amazon Data Lifecycle Manager para automatizar la eliminación de las instantáneas de Amazon EBS y las AMI basadas en Amazon EBS.
- Configuración del ciclo de vida en un bucket: utilice la configuración del ciclo de vida de Amazon S3 en un bucket para definir las acciones que Amazon S3 debe llevar a cabo durante el ciclo de vida de un objeto, así como la eliminación al final del ciclo de vida del objeto, en función de los requisitos de su empresa.

Recursos

Documentos relacionados:

- AWS Trusted Advisor
- Administrador de vida útil de datos de Amazon
- Configuración de un ciclo de vida en un bucket de Amazon S3

Videos relacionados:

- Automate Amazon EBS Snapshots with Amazon Data Lifecycle Manager
- ¿Cómo puedo vaciar un bucket de Amazon S3 mediante una regla de configuración del ciclo de vida?

Ejemplos relacionados:

- ¿Cómo puedo vaciar un bucket de Amazon S3 mediante una regla de configuración del ciclo de vida?

Recursos rentables

El uso de los servicios, recursos y configuraciones adecuados para las cargas de trabajo es clave para ahorrar costos. Tenga en cuenta lo siguiente al crear recursos rentables:

Con la ayuda de los arquitectos de soluciones de AWS, las soluciones de AWS, las arquitecturas de referencia de AWS y los socios de APN, podrá elegir una arquitectura según la información que ha recibido.

Temas

- Evaluación del costo al seleccionar los servicios
- Selección del tipo, tamaño y número correctos de recursos
- Selección del mejor modelo de precios
- Planificación para la transferencia de datos

Evaluación del costo al seleccionar los servicios

Prácticas recomendadas

- COST05-BP01 Identificación de los requisitos de la organización en relación con el costo
- COST05-BP02 Análisis de todos los componentes de la carga de trabajo
- COST05-BP03 Ejecución de un análisis exhaustivo de cada componente
- COST05-BP04 Selección de software con licencias rentables
- COST05-BP05 Selección de los componentes de la carga de trabajo para optimizar los costos de acuerdo con las prioridades de la organización
- COST05-BP06 Análisis de costos para diferentes usos a lo largo del tiempo

COST05-BP01 Identificación de los requisitos de la organización en relación con el costo

Trabaje con los miembros del equipo para definir el equilibrio entre la optimización de costos y otros pilares, como el rendimiento y la fiabilidad, de la carga de trabajo.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

En la mayoría de las organizaciones, el departamento de tecnología de la información (TI) está compuesto por varios equipos pequeños, cada uno con su propia agenda y área de enfoque, que refleja las especialidades y habilidades de los miembros de su equipo. Debe conocer los objetivos, prioridades y metas generales de su organización y cómo cada departamento o proyecto contribuye a estos objetivos. La clasificación de todos los recursos esenciales, incluidos el personal, el equipo, la tecnología, los materiales y los servicios externos, es crucial para lograr los objetivos de la organización y una planificación del presupuesto exhaustiva. La adopción de este enfoque sistemático para la identificación y comprensión de los costos es fundamental para establecer un plan de costos realista y sólido para la organización.

A la hora de seleccionar los servicios para su carga de trabajo, es fundamental que entienda las prioridades de su organización. Cree un equilibrio entre la optimización de costos y otros pilares del Marco de AWS Well-Architected, como el rendimiento y la fiabilidad. Este proceso debe llevarse a cabo de manera sistemática y regular para reflejar los cambios en los objetivos de la organización, las condiciones del mercado y la dinámica operativa. Una carga de trabajo totalmente optimizada en cuanto a costos es la solución que más se ajusta a los requisitos de su organización, no necesariamente la de menor costo. Reúnase con todos los equipos de su organización (por ejemplo, de productos, empresarial, técnico y financiero) para recopilar información. Evalúe el impacto de las compensaciones que se hacen entre intereses opuestos o enfoques alternativos para ayudar a tomar decisiones fundamentadas a la hora de determinar dónde centrar los esfuerzos o elegir una vía de acción.

Por ejemplo, comercializar más rápido las nuevas características puede primar sobre la optimización de los costos, o se podría elegir una base de datos relacional para los datos no relacionales para 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 su aplicación.

Pasos para la implementación

- Identificación de los requerimientos para costos de su organización: reúnase con los miembros de los equipos de su organización, incluidos los de administración de productos, propietarios de aplicaciones, equipos de desarrollo y operativos, departamentos de administración y roles en finanzas. Priorice los pilares de Well-Architected para esta carga de trabajo y sus componentes. El resultado debería ser una lista ordenada de los pilares. También puede agregar una ponderación a cada pilar para indicar cuánto enfoque adicional tiene, o las similitudes de un enfoque entre dos pilares.

- Corrección de la deuda técnica y documentación: durante la revisión de la carga de trabajo, corrija la deuda técnica. Documente una tarea pendiente para revisitar la carga de trabajo en el futuro, con el objetivo de refactorizarla o rediseñarla para optimizarla aún más. Es esencial comunicar claramente a otras partes interesadas las compensaciones que se han hecho.

Recursos

Prácticas recomendadas relacionadas:

- REL11-BP07 Diseño de su producto para cumplir objetivos de disponibilidad y acuerdos de nivel de servicio (SLA) de tiempo de actividad
- OPS01-BP06 Evaluación de las compensaciones

Documentos relacionados:

- Calculadora del costo total de propiedad (TCO) de AWS
- Clases de almacenamiento de Amazon S3
- Productos de la nube

COST05-BP02 Análisis de todos los componentes de la carga de trabajo

Asegúrese de que se analice cada componente de la carga de trabajo, independientemente del tamaño o del costo actuales. El esfuerzo de revisión debería reflejar el beneficio potencial, como los costos actuales y previstos.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Los componentes de la carga de trabajo, que están diseñados para ofrecer valor empresarial a la organización, pueden abarcar varios servicios. Para cada componente, se pueden elegir servicios específicos de Nube de AWS para abordar las necesidades empresariales. Esta selección podría estar influenciada por factores como la familiaridad con estos servicios o la experiencia previa con ellos.

Después de identificar los requisitos de su organización, tal como se menciona en COST05-BP01 Identificación de los requisitos de la organización en relación con el costo y haga un análisis

exhaustivo de todos los componentes de su carga de trabajo. Analice cada componente teniendo en cuenta los costos y tamaños actuales y previstos. Compare el costo del análisis con cualquier posible ahorro en la carga de trabajo a lo largo de su ciclo de vida. El esfuerzo que se dedique a analizar todos los componentes de esta carga de trabajo debe compensar los posibles ahorros o mejoras que se tiene previsto conseguir con la optimización de ese componente específico. Por ejemplo, si el costo del recurso propuesto es de 10 USD al mes y, según las cargas previstas, no superaría los 15 USD al mes, dedicar un día de esfuerzo a reducir los costos un 50 % (5 USD al mes) no debería superar el beneficio potencial durante la vida del sistema. Utilice una estimación basada en datos más eficiente y rápida para conseguir el mejor resultado global para este componente.

Las cargas de trabajo pueden cambiar con el tiempo y el conjunto adecuado de servicios podría no ser óptimo si la arquitectura o el uso de la carga de trabajo cambia. En el análisis para seleccionar los servicios, se deben incluir estados de carga de trabajo actuales y futuros y niveles de uso. Implementar un servicio para un estado o uso de la carga de trabajo futura puede reducir los costos globales al reducir o eliminar el esfuerzo requerido para hacer cambios en el futuro. Por ejemplo, es posible que sea adecuado utilizar EMR sin servidor en un principio. Sin embargo, a medida que aumenta el consumo de ese servicio, la transición a EMR en EC2 podría reducir los costos de ese componente de la carga de trabajo.

AWS Cost Explorer y el AWS Cost and Usage Report (CUR) pueden analizar el costo de una prueba de concepto (PoC) o un entorno en ejecución. También puede utilizar Calculadora de precios de AWS para estimar los costos de la carga de trabajo.

Escriba el flujo de trabajo que deben seguir los equipos técnicos para revisar sus cargas de trabajo. Procure que este flujo de trabajo sea sencillo, pero que abarque todos los pasos necesarios para asegurarse de que los equipos conozcan cada componente de la carga de trabajo y sus precios. Luego, su organización puede seguir y personalizar este flujo de trabajo en función de las necesidades específicas de cada equipo.

1. Enumeración de cada servicio que se use para la carga de trabajo: este es un buen punto de partida. Identifique todos los servicios que se están utilizando actualmente y el origen de los costos.
2. Comprensión de cómo funcionan los precios de esos servicios: comprenda el modelo de precios de cada servicio. Los distintos servicios de AWS tienen diferentes modelos de precios en función de factores como el volumen de uso, la transferencia de datos y los precios de características específicas.
3. Concentración en los servicios que conllevan costos de la carga de trabajo inesperados y que no se ajustan al uso y los resultados empresariales esperados: identifique los valores atípicos

o los servicios en los que el costo no sea proporcional al valor o al uso mediante AWS Cost Explorer o AWS Cost and Usage Report. Es importante correlacionar los costos con los resultados empresariales para priorizar los esfuerzos de optimización.

4. AWS Cost Explorer, Registros de CloudWatch, VPC Flow Logs y Lente de almacenamiento de Amazon S3 para comprender la causa principal de esos altos costos: estas herramientas son fundamentales para diagnosticar los altos costos. Cada servicio ofrece una perspectiva diferente para ver y analizar el uso y los costos. Por ejemplo, el Explorador de costos ayuda a determinar las tendencias generales de los costos, Registros de CloudWatch proporciona información operativa, Registros de flujos de VPC muestran el tráfico IP y la Lente de almacenamiento de Amazon S3 es útil para analizar el almacenamiento.
5. Uso de AWS Budgets para establecer presupuestos para una determinada cantidad de servicios o cuentas: establecer presupuestos es una forma proactiva de administrar los costos. Utilice AWS Budgets para establecer umbrales presupuestarios personalizados y recibir alertas cuando los costos superen esos umbrales.
6. Configuración de las alarmas de Amazon CloudWatch para enviar alertas de facturación y uso: configure la supervisión y las alertas para las métricas de costo y uso. Las alarmas de CloudWatch pueden avisarle cuando se alcanzan ciertos umbrales, lo que mejora el tiempo de respuesta a la intervención.

Consiga una mejora y un ahorro financiero importantes a lo largo del tiempo mediante una revisión estratégica de todos los componentes de la carga de trabajo e independientemente de sus atributos actuales. El esfuerzo invertido en este proceso de revisión debe ser deliberado y deben estudiarse cuidadosamente las ventajas que podrían conseguirse.

Pasos para la implementación

- Enumeración de los componentes de la carga de trabajo: cree una lista de los componentes de la carga de trabajo. Utilice esta lista para comprobar que se hayan analizado todos los componentes. El esfuerzo que le dedique debería reflejar la importancia de la carga de trabajo, tal como definen las prioridades de la organización. Agrupar los recursos mejora la eficiencia funcional (por ejemplo, el almacenamiento de la base de datos de producción si hay varias bases de datos).
- Priorización de la lista de componentes: tome la lista de componentes y priorícela por orden de esfuerzo. En general, se ordena por el costo del componente, es decir, de más caro a menos caro, o por la importancia definida en las prioridades de la organización.
- Análisis: para cada componente de la lista, revise las opciones y los servicios disponibles y elija la opción que mejor se adapte a las prioridades de su organización.

Recursos

Documentos relacionados:

- Calculadora de precios de AWS
- AWS Cost Explorer
- Clases de almacenamiento de Amazon S3
- Productos en la Nube de AWS

Videos relacionados:

- AWS Cost Optimization Series: CloudWatch

COST05-BP03 Ejecución de un análisis exhaustivo de cada componente

Consulte el costo total que supone para la organización cada componente. Calcule el costo total de propiedad teniendo en cuenta el costo de las operaciones y la administración, sobre todo cuando utilice servicios administrados por el proveedor de servicios en la nube. El esfuerzo de revisión debe reflejar los posibles beneficios (por ejemplo, el tiempo empleado en analizar es proporcional al costo de los componentes).

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Considere el ahorro de tiempo que permitirá a su equipo centrarse en la retirada de la deuda técnica, la innovación, las características que agregan valor y la creación de lo que diferencia a la empresa. Por ejemplo, puede que necesite migrar mediante lift-and-shift (también conocido como “volver a alojar”) sus bases de datos de su entorno en las instalaciones a la nube lo más rápidamente posible y optimizarlas más tarde. Merece la pena explorar el ahorro que puede suponer el uso de servicios administrados en AWS que puedan eliminar o reducir los costos de las licencias. Los servicios administrados en AWS eliminan la carga operativa y administrativa del mantenimiento de un servicio, como la aplicación de parches o la actualización del sistema operativo, y le permiten centrarse en la innovación y la empresa.

Dado que los servicios administrados operan a la escala de la nube, pueden ofrecer un costo menor por transacción o servicio. Puede llevar a cabo optimizaciones potenciales para obtener alguna

ventaja tangible, sin cambiar la arquitectura principal de la aplicación. Por ejemplo, es posible que desee reducir la cantidad de tiempo que dedica a administrar instancias de bases de datos mediante la migración a una plataforma de base de datos como servicio, como Amazon Relational Database Service (Amazon RDS), o mediante la migración de su aplicación a una plataforma totalmente administrada, como AWS Elastic Beanstalk.

Normalmente, los servicios administrados tienen atributos que puede configurar para garantizar una capacidad suficiente. Debe configurar y supervisar estos atributos para que su exceso de capacidad se mantenga al mínimo y el rendimiento se maximice. Puede modificar los atributos de AWS Managed Services mediante la Consola de administración de AWS o las API y los SDK de AWS para adaptar las necesidades de recursos a la demanda cambiante. Por ejemplo, puede aumentar o disminuir la cantidad de nodos en un clúster de Amazon EMR (o en un clúster de Amazon Redshift) para reducir o escalar horizontalmente.

También puede empaquetar varias instancias en un recurso de AWS para activar un uso de mayor densidad. Por ejemplo, puede aprovisionar varias bases de datos pequeñas en una sola instancia de base de datos de Amazon Relational Database Service (Amazon RDS). A medida que aumenta el uso, puede migrar una de las bases de datos a una instancia de base de datos de Amazon RDS dedicada mediante un proceso de restauración y una instantánea.

Cuando aprovisione cargas de trabajo mediante servicios administrados, debe conocer los requisitos para ajustar la capacidad del servicio. Estos requisitos suelen ser tiempo, esfuerzo y cualquier impacto en el funcionamiento normal de la carga de trabajo. El recurso aprovisionado debe dejar tiempo para que se produzca cualquier cambio, por lo que debe aprovisionar la sobrecarga necesaria para permitirlo. El esfuerzo continuo requerido para modificar los servicios se puede reducir a prácticamente cero mediante el uso de las API y los SDK que se integran con el sistema y las herramientas de supervisión, como Amazon CloudWatch.

Amazon RDS, Amazon Redshift y Amazon ElastiCache proporcionan un servicio de bases de datos administrado. Amazon Athena, Amazon EMR y Amazon OpenSearch Service proporcionan un servicio de análisis administrado.

AMS es un servicio que utiliza la infraestructura de AWS en nombre de los socios y clientes de la empresa. Proporciona un entorno seguro y conforme a las normativas en el que puede implementar sus cargas de trabajo. AMS utiliza modelos operativos de nube empresarial con automatización para permitirle satisfacer los requisitos de su organización, trasladarse a la nube más rápidamente y reducir los costos de administración continua.

Pasos para la implementación

- Análisis exhaustivo: mediante la lista de componentes, examine cada uno de ellos de mayor a menor prioridad. En el caso de los componentes con mayor prioridad y más costosos, lleve a cabo un análisis adicional y evalúe todas las opciones disponibles y su impacto a largo plazo. En el caso de los componentes con menor prioridad, evalúe si los cambios en el uso modificarían la prioridad del componente y, a continuación, haga un análisis del esfuerzo adecuado.
- Comparación de los recursos administrados y no administrados: considere el costo operativo de los recursos que administra y compárelos con los recursos administrados de AWS. Por ejemplo, revise sus bases de datos que se ejecutan en instancias de Amazon EC2 y compárelas con las opciones de Amazon RDS (un servicio administrado de AWS) o compare Amazon EMR con la ejecución de Apache Spark en Amazon EC2. Cuando cambie de una carga de trabajo autoadministrada a una completamente administrada por AWS, investigue cuidadosamente sus opciones. Los tres factores más importantes que tener en cuenta son el tipo de servicio administrado que desea utilizar, el proceso que utilizará para migrar los datos y comprender el modelo de responsabilidad compartida de AWS.

Recursos

Documentos relacionados:

- AWS Calculadora del costo total de propiedad (TCO) de
- Clases de almacenamiento de Amazon S3
- Productos en la Nube de AWS
- Modelo de responsabilidad compartida de AWS

Videos relacionados:

- Why move to a managed database?
- What is Amazon EMR and how can I use it for processing data?

Ejemplos relacionados:

- Why to move to a managed database
- Consolidate data from identical SQL Server databases into a single Amazon RDS for SQL Server database using AWS DMS
- Deliver data at scale to Amazon Managed Streaming for Apache Kafka (Amazon MSK)

- Migrate an ASP.NET web application to AWS Elastic Beanstalk

COST05-BP04 Selección de software con licencias rentables

El software de código abierto elimina los costos de licencias de software, lo que puede repercutir enormemente en los costos de las cargas de trabajo. Si se requiere software con licencia, evite licencias vinculadas a atributos arbitrarios como las CPU y busque licencias vinculadas a los resultados. El costo de estas licencias está más vinculado al beneficio que aportan.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

El código abierto se originó en el contexto del desarrollo de software e indica que el software cumple ciertos criterios para su distribución gratuita. El software de código abierto consta de código fuente que cualquiera puede inspeccionar, modificar y mejorar. En función de los requisitos empresariales, la habilidad de los ingenieros, el uso previsto u otras dependencias tecnológicas, las organizaciones pueden considerar la posibilidad de utilizar software de código abierto en AWS para minimizar los costos de sus licencias. En otras palabras, el costo de las licencias de software se puede reducir mediante el uso de software de código abierto. Esto puede repercutir de forma significativa en los costos de la carga de trabajo a medida que esta aumente.

Determine los beneficios del software con licencia teniendo en cuenta el costo total para optimizar su carga de trabajo. Haga simulaciones de los cambios en las licencias y estudie cómo afectaría a los costos de la carga de trabajo. Si un proveedor cambia el costo de la licencia de la base de datos, investigue cómo afecta eso a la eficiencia general de la carga de trabajo. Consulte el historial de anuncios de precios de sus proveedores para ver las tendencias en los cambios de las licencias en sus productos. Los costos de licencia también pueden variar sin tener en cuenta el rendimiento o el uso, como las licencias que varían según el hardware (licencias vinculadas a la CPU). Estas licencias deberían evitarse porque sus costos pueden incrementarse rápidamente sin que haya unos resultados correspondientes.

Por ejemplo, utilizar una instancia de Amazon EC2 en us-east-1 con un sistema operativo Linux le permite reducir los costos en aproximadamente un 45 %, en comparación con la ejecución de otra instancia de Amazon EC2 que se ejecute en Windows.

Calculadora de precios de AWS ofrece una forma integral de comparar los costos de varios recursos con diferentes opciones de licencia, como las instancias de Amazon RDS y los diferentes motores de bases de datos. Además, AWS Cost Explorer proporciona una perspectiva muy valiosa de los costos

de las cargas de trabajo existentes, especialmente aquellas que vienen con diferentes licencias. Para la administración de licencias, AWS License Manager ofrece un método simplificado para supervisar y administrar las licencias de software. Los clientes pueden implementar y poner en funcionamiento su software de código abierto preferido en Nube de AWS.

Pasos para la implementación

- Análisis de las opciones de licencia: revise las condiciones de licencia del software disponible. Busque versiones de código abierto que dispongan de las funciones requeridas y si los beneficios del software con licencia superan su costo. Si las condiciones son favorables, el costo del software se compensa con el beneficio que aporta.
- Análisis del proveedor de software: revise el historial de cambios en los precios y las licencias del proveedor. Busque cambios que no se alineen con los resultados, tales como términos punitivos si se ejecuta hardware o se trabaja con plataformas de proveedores específicos. Además, fíjese en cómo hacen las auditorías y las sanciones que se podrían aplicar.

Recursos

Documentos relacionados:

- Open Source at AWS
- Calculadora del costo total de propiedad (TCO) de AWS
- Clases de almacenamiento de Amazon S
- Productos de la nube

Ejemplos relacionados:

- Open Source Blogs
- AWS Open Source Blogs
- Evaluación de optimización y licencias

COST05-BP05 Selección de los componentes de la carga de trabajo para optimizar los costos de acuerdo con las prioridades de la organización

Tenga en cuenta el costo al seleccionar los componentes de su carga de trabajo. Esto incluye el uso de servicios administrados y por aplicación o de una arquitectura sin servidor, de contenedores

o basada en eventos para reducir el costo global. Minimice los costos de licencia con software de código abierto, software que no tenga costos de licencia o alternativas para reducir el costo.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

Tenga en cuenta el costo de los servicios y las opciones a la hora de seleccionar los componentes. Esto incluye el uso de servicios administrados y a nivel de aplicación, como Amazon Relational Database Service (Amazon RDS), Amazon DynamoDB, Amazon Simple Notification Service (Amazon SNS) y Amazon Simple Email Service (Amazon SES) para reducir los costos organizativos generales.

Utilice contenedores y tecnología sin servidor para la computación, como AWS Lambda y Amazon Simple Storage Service (Amazon S3) para sitios web estáticos. Si es posible, coloque la aplicación en contenedores y utilice servicios de contenedores administrados de AWS, como Amazon Elastic Container Service (Amazon ECS) o Amazon Elastic Kubernetes Service (Amazon EKS).

Minimice los costos de licencia con software de código abierto o software que no tenga costos de licencia (por ejemplo, Amazon Linux para cargas de trabajo de computación o migre bases de datos a Amazon Aurora).

Puede utilizar servicios sin servidor o por aplicación, como Lambda, Amazon Simple Queue Service (Amazon SQS), Amazon SNS y Amazon SES. Estos servicios eliminan la necesidad de administrar un recurso y proporcionan la función de ejecución de código, servicios de colas y entrega de mensajes. La otra ventaja es que reducen horizontalmente el rendimiento y el costo de acuerdo con el uso, por lo que permiten la asignación y atribución de costos de forma eficiente.

El uso de una arquitectura basada en eventos también es posible con servicios sin servidor. Las arquitecturas basadas en eventos se basan en la inserción, por lo que todo sucede bajo demanda a medida que el evento se presenta en el enrutador. De esta forma, no pagará por un sondeo continuo para comprobar si hay algún evento. Esto se traduce en un menor consumo de ancho de banda de la red, un menor uso de la CPU, una menor capacidad inactiva de la flota y menos establecimientos de protocolo de enlace SSL/TLS.

Para obtener más información sobre la tecnología sin servidor, consulte el documento técnico sobre lentes de Well-Architected Serverless Application.

Pasos para la implementación

- Seleccione cada servicio para optimizar el costo: se la lista de prioridades y el análisis para seleccionar la opción que se adapte mejor a las prioridades de la organización. En lugar de aumentar la capacidad para satisfacer la demanda, considere otras opciones que puedan ofrecerle un mejor rendimiento con un costo menor. Por ejemplo, si debe revisar el tráfico previsto para sus bases de datos en AWS, considere la posibilidad de aumentar el tamaño de la instancia o de utilizar servicios de Amazon ElastiCache (Redis o Memcached) a fin de proporcionar mecanismos de caché para sus bases de datos.
- Evaluación de la arquitectura basada en eventos: el uso de una arquitectura sin servidor también le permite crear una arquitectura basada en eventos para aplicaciones distribuidas basadas en microservicios, lo que le ayuda a crear soluciones escalables, resilientes, ágiles y rentables.

Recursos

Documentos relacionados:

- Calculadora del costo total de propiedad (TCO) de AWS
- AWS sin servidor
- ¿Qué es la arquitectura basada en eventos (EDA)?
- Clases de almacenamiento de Amazon S3
- Productos de la nube
- Amazon ElastiCache (Redis OSS)

Ejemplos relacionados:

- Getting started with event-driven architecture
- Arquitectura basada en eventos
- How Statsig runs 100x more cost-effectively using Amazon ElastiCache (Redis OSS)
- Prácticas recomendadas para trabajar con funciones de AWS Lambda

COST05-BP06 Análisis de costos para diferentes usos a lo largo del tiempo

Las cargas de trabajo pueden cambiar con el tiempo. Algunos servicios o características son más rentables en diferentes niveles de uso. Al analizar cada componente a lo largo del tiempo, así como el uso previsto, la carga de trabajo se mantiene rentable durante su vida útil.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

A medida que AWS lanza nuevos servicios y características, los servicios óptimos para su carga de trabajo también pueden cambiar. El esfuerzo necesario debe reflejar las ventajas potenciales. La frecuencia de revisión de la carga de trabajo depende de los requisitos de la organización. Si se trata de una carga de trabajo con un costo importante, implementar nuevos servicios antes maximizará el ahorro, por lo que llevar a cabo la revisión con mayor frecuencia puede ser de gran ayuda. Otro aspecto inicial que revisar es el cambio en los patrones de uso. Unos cambios significativos en el uso pueden indicar que unos servicios alternativos serían óptimos.

Si necesita trasladar datos a la Nube de AWS, puede seleccionar cualquier amplia variedad de servicios que ofrece AWS y herramientas de socios para ayudarle a migrar sus conjuntos de datos, ya sean archivos, bases de datos, imágenes de máquinas, volúmenes de bloques o, incluso, copias de seguridad en cinta. Por ejemplo, para trasladar una gran cantidad de datos con destino y origen en AWS o procesar datos en la periferia, puede utilizar uno de los dispositivos personalizados de AWS para trasladar de forma rentable petabytes de datos fuera de línea. Otro ejemplo: para tasas de transferencia de datos más elevadas, un servicio de conexión directa puede resultar más barato que una VPN que proporcione la coherencia de conectividad necesaria para su empresa.

Revise su actividad de escalado basándose en el análisis de costos para diferentes usos a lo largo del tiempo. Analice el resultado para ver si la política de escalado puede ajustarse para agregar instancias con varios tipos de instancia y opciones de compra. Revise la configuración para ver si es posible reducir el mínimo para atender las solicitudes de los usuarios, pero con una flota de menor tamaño, y agregue más recursos para satisfacer la elevada demanda prevista.

Para efectuar un análisis de costos para diferentes usos a lo largo del tiempo, converse con las partes interesadas de su organización y utilice la característica de previsión de AWS Cost Explorer para predecir el impacto potencial de los cambios en el servicio. Supervise los desencadenadores del nivel de uso mediante AWS Budgets, alarmas de facturación de CloudWatch y AWS Cost Anomaly Detection para identificar y lanzar antes los servicios más rentables.

Pasos para la implementación

- Definición de los patrones de uso previstos: trabaje con su organización, como los propietarios de marketing y de productos, para documentar cuáles serán los patrones de uso esperados y previstos para la carga de trabajo. Hable con las partes interesadas de la empresa sobre el aumento de costo y uso, tanto históricos como previstos, y asegúrese de que el aumento se ajusta a los requisitos de la empresa. Identifique los días naturales, las semanas o los meses en los que espera que más usuarios utilicen sus recursos de AWS, lo que indica que debe aumentar la capacidad de los recursos existentes o adoptar servicios adicionales para reducir el costo y aumentar el rendimiento.
- Análisis de costos según el uso previsto: use los patrones de uso definidos para llevar a cabo un análisis en cada uno de estos puntos. El esfuerzo de análisis debería reflejar el resultado potencial. Por ejemplo, si el cambio de uso es grande, debería hacerse un análisis exhaustivo para verificar los costos y los cambios. En otras palabras, cuando el costo aumenta, el uso también debería aumentar para la empresa.

Recursos

Documentos relacionados:

- Calculadora del costo total de propiedad (TCO) de AWS
- Clases de almacenamiento de Amazon S3
- Productos de la nube
- Amazon EC2 Auto Scaling
- Migración de datos a la nube
- AWS Snow Family

Videos relacionados:

- AWS OpsHub for Snow Family

Selección del tipo, tamaño y número correctos de recursos

Al seleccionar el mejor tipo, tamaño y número de recursos, se cumplen los requisitos técnicos con el recurso de menor costo. Las acciones de ajuste correcto del tamaño tienen en cuenta todos los recursos de una carga de trabajo, todos los atributos de cada recurso individual y el esfuerzo de la operación de ajuste correcto del tamaño. El ajuste correcto del tamaño puede ser un proceso

iterativo que se inicia por los cambios que se produzcan en los patrones de uso y los factores externos, tales como la caída de los precios de AWS o los nuevos tipos de recursos de AWS. El ajuste correcto del tamaño también puede ser único si el costo del esfuerzo hecho para esta actividad supera los posibles ahorros a lo largo de la vida de la carga de trabajo.

En AWS existen varios enfoques diferentes:

Prácticas recomendadas

- COST06-BP01 Modelado de costos
- COST06-BP02 Selección del tipo, tamaño y número de recursos en función de los datos
- COST06-BP03 Selección automática del tipo, tamaño y número de recursos en función de las métricas
- COST06-BP04 Consideración del uso de los recursos compartidos

COST06-BP01 Modelado de costos

Identifique los requisitos de la organización (como las necesidades empresariales y los compromisos existentes) y lleve a cabo un modelado de costos (costos generales) de la carga de trabajo y de cada uno de sus componentes. Lleve a cabo actividades de referencia para la carga de trabajo bajo diferentes cargas previstas y compare los costos. El esfuerzo para llevar a cabo el modelado debería reflejar la ventaja potencial. Por ejemplo, el tiempo dedicado debe ser proporcional al costo del componente.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Lleve a cabo la creación de modelos de costos para la carga de trabajo y cada uno de sus componentes para comprender el equilibrio entre los recursos. Busque el tamaño adecuado para cada recurso de la carga de trabajo según un determinado nivel de rendimiento. Comprender las consideraciones de costos puede servir de base al caso empresarial de su organización y al proceso de toma de decisiones cuando se evalúen los resultados de obtención de valor para la implementación planificada de la carga de trabajo.

Lleve a cabo actividades de referencia para la carga de trabajo bajo diferentes cargas previstas y compare los costos. El esfuerzo de creación de modelos debe reflejar los posibles beneficios; por ejemplo, el tiempo dedicado es proporcional al costo de los componentes o al ahorro previsto. Para

conocer las mejores prácticas, consulte la sección de revisión del pilar de eficiencia de rendimiento del Marco de AWS Well-Architected.

Por ejemplo, para crear modelos de costos para una carga de trabajo compuesta por recursos informáticos, AWS Compute Optimizer puede ayudar a modelar los costos de las cargas de trabajo en ejecución. Proporciona recomendaciones de tamaño ideal para los recursos de computación en función del historial de uso. Asegúrese de que se implementen agentes de CloudWatch en las instancias de Amazon EC2 para recopilar métricas de memoria que le ayuden con recomendaciones más precisas en AWS Compute Optimizer. Se trata del origen de datos ideal para los recursos de computación porque es un servicio gratuito que usa el machine learning para llevar a cabo numerosas recomendaciones en función de los niveles de riesgo.

Hay varios servicios que puede utilizar con registros personalizados como orígenes de datos para el dimensionamiento correcto de las operaciones de otros servicios y componentes de la carga de trabajo como AWS Trusted Advisor, Amazon CloudWatch y Registros de Amazon CloudWatch. AWS Trusted Advisor comprueba los recursos y marca los que se utilizan poco, lo que puede ayudarle a dimensionar correctamente sus recursos y a crear modelos de costos.

Estas son recomendaciones de datos y métricas de creación de modelos de costos:

- La supervisión debe reflejar fielmente la experiencia del usuario. Seleccione la granularidad correcta del periodo y elija cuidadosamente el percentil 99 o el percentil máximo en lugar del promedio.
- Seleccione el nivel de detalle correcto para el periodo de análisis necesario a fin de cubrir cualquier ciclo de carga de trabajo. Por ejemplo, si se lleva a cabo un análisis de dos semanas, es posible que esté pasando por alto un ciclo mensual de alto uso, lo que podría generar un aprovisionamiento insuficiente.
- Elija los servicios de AWS adecuados para la carga de trabajo prevista; para ello, tenga en cuenta sus compromisos existentes, los modelos de precios seleccionados para otras cargas de trabajo y la capacidad de innovar más rápidamente y centrarse en el valor empresarial principal.

Pasos para la implementación

- Modelado de costos: implemente la carga de trabajo o una prueba de concepto en una cuenta separada con los tipos y tamaños de recurso específicos de la prueba. Ejecute la carga de trabajo con los datos de la prueba y registre los resultados de la salida, así como los datos de costos del momento en que se ejecutó la prueba. Después, vuelva a implementar la carga de trabajo o cambie los tipos y tamaños de recurso y vuelva a ejecutar la prueba. Incluya las tarifas de

licencia de cualquier producto que pueda utilizar con estos recursos y los costos estimados de las operaciones (mano de obra o ingenieros) para implementar y administrar estos recursos durante la creación del modelado de costos. Considere el modelado de costos para un periodo (por hora, por día, por mes, por año o por trienio).

Recursos

Documentos relacionados:

- AWS Auto Scaling
- Identificación de oportunidades para ajustar el tamaño
- Características de Amazon CloudWatch
- Cost Optimization: Amazon EC2 Right Sizing
- AWS Compute Optimizer
- Calculadora de precios de AWS

Ejemplos relacionados:

- Creación de modelos de costos basado en datos
- ¿Cómo calculo el costo de las configuraciones planificadas de recursos de AWS?
- Choose the right AWS tools

COST06-BP02 Selección del tipo, tamaño y número de recursos en función de los datos

Seleccione el tamaño o tipo de recurso en función de los datos sobre las características de la carga de trabajo y de los recursos. Por ejemplo, computación, memoria, rendimiento o uso intensivo de escritura. Para llevar a cabo esta selección, suele utilizarse una versión anterior (en las instalaciones) de la carga de trabajo, la documentación u otras fuentes de información sobre la carga de trabajo.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

Amazon EC2 ofrece una amplia selección de tipos de instancias con diferentes niveles de capacidad de CPU, memoria, almacenamiento y red para adaptarse a diferentes casos de uso. Estos tipos de

instancias ofrecen diferentes combinaciones de capacidades de CPU, memoria, almacenamiento y red, lo que le proporciona versatilidad a la hora de seleccionar la combinación de recursos adecuada para sus proyectos. Cada tipo de instancia se ofrece en varios tamaños, de modo que puede ajustar sus recursos en función de las demandas de su carga de trabajo. Para determinar qué tipo de instancia necesita, recopile datos sobre los requisitos del sistema de la aplicación o el software que tiene pensado ejecutar en su instancia. Estos datos deben incluir lo siguiente:

- Sistema operativo
- Número de núcleos de CPU
- Núcleos de GPU
- Cantidad de memoria del sistema (RAM)
- Tipo y espacio de almacenamiento
- Requisitos de ancho de banda de la red

Identifique el propósito de los requisitos de computación y qué instancia se necesita y, a continuación, examine las distintas familias de instancias de Amazon EC2. Amazon ofrece las siguientes familias de tipos de instancias:

- Uso general
- Computación optimizada
- Optimizada para memoria
- Optimización de almacenamiento
- Computación acelerada
- Optimizadas para HPC

Para obtener información más profunda sobre los propósitos y casos de uso específicos que puede cumplir una familia de instancias concreta de Amazon EC2, consulte Tipos de instancias de AWS.

La recopilación de requisitos del sistema es fundamental para seleccionar la familia de instancias y el tipo de instancia específicos que mejor se ajusten a sus necesidades. Los nombres de los tipos de instancias están compuestos por el nombre de la familia y el tamaño de la instancia. Por ejemplo, la instancia t2.micro pertenece a la familia T2 y tiene el tamaño micro.

Seleccione el tamaño o el tipo de recurso en función de las características de la carga de trabajo y de los recursos (por ejemplo: computación, memoria, rendimiento o uso intensivo de escritura). Para

llevar a cabo esta selección, suele utilizarse la creación de modelos de costos, una versión anterior de la carga de trabajo (por ejemplo, una versión en las instalaciones), mediante documentación u otras fuentes de información sobre la carga de trabajo (documentos técnicos o soluciones publicadas). El uso de calculadoras de precios o herramientas de administración de costos de AWS puede ayudar a tomar decisiones informadas sobre los tipos, tamaños y configuraciones de las instancias.

Pasos para la implementación

- Selección de los recursos en función de los datos: utilice los datos de modelado de costos para seleccionar el nivel de uso previsto de la carga de trabajo y elija el tipo y el tamaño de los recursos especificados. En función de los datos del modelado de costos, determine el número de CPU virtuales, la memoria total (GiB), el volumen del almacén de instancias local (GB), los volúmenes de Amazon EBS y el nivel de rendimiento de la red, teniendo en cuenta la velocidad de transferencia de datos necesaria para la instancia. Haga siempre selecciones basadas en análisis detallados y datos precisos para optimizar el rendimiento al tiempo que administra los costos de forma eficaz.

Recursos

Documentos relacionados:

- AWS Tipos de instancias
- AWS Auto Scaling
- Características de Amazon CloudWatch
- Optimización de costos: Ajuste correcto del tamaño de EC2

Videos relacionados:

- Selecting the right Amazon EC2 instance for your workloads
- Right size your service

Ejemplos relacionados:

- It just got easier to discover and compare Amazon EC2 instance types

COST06-BP03 Selección automática del tipo, tamaño y número de recursos en función de las métricas

Use métricas de la carga de trabajo actual para seleccionar el tamaño y tipo correctos para optimizar el costo. Aprovisione de forma adecuada el rendimiento, el tamaño y el almacenamiento para los servicios de computación, almacenamiento, datos y redes. Esto puede hacerse con un bucle de retroalimentación, como el escalado automático, o mediante un código personalizado en la carga de trabajo.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

Cree un bucle de retroalimentación en la carga de trabajo que use métricas activas de la carga de trabajo en ejecución para hacer cambios en dicha carga de trabajo. Puede utilizar un servicio gestionado, por ejemplo AWS Auto Scaling, que configure para que lleve a cabo las operaciones de dimensionamiento adecuadas. AWS también proporciona API, SDK y funciones que permiten modificar los recursos con un esfuerzo mínimo. Puede programar una carga de trabajo para que detenga e inicie una instancia de Amazon EC2 a fin de poder hacer un cambio en el tamaño o el tipo de instancia. Esto permite obtener el tamaño adecuado y, además, permite eliminar casi todo el costo operativo necesario para hacer el cambio.

Algunos servicios de AWS incluyen una selección automática de tipos o tamaños, como Amazon Simple Storage Service Intelligent-Tiering. Amazon S3 Intelligent-Tiering mueve automáticamente los datos entre dos niveles de acceso (frecuente y poco frecuente) en función de sus patrones de uso.

Pasos para la implementación

- Configuración de las métricas de la carga de trabajo para aumentar la observabilidad: capture las métricas clave de la carga de trabajo. Estas métricas son indicativas de la experiencia del cliente, como el resultado de la carga de trabajo, y alinean las diferencias que hay entre los tipos y los tamaños de los recursos, como la CPU y el uso de memoria. En el caso del recurso de computación, analice los datos de rendimiento para determinar el tamaño adecuado de sus instancias de Amazon EC2. Identifique las instancias inactivas y las infrautilizadas. Las métricas clave que hay que tener en cuenta son el uso de la CPU y de la memoria (por ejemplo, el 40 % de uso de la CPU el 90 % del tiempo, como se explica en Rightsizing with AWS Compute Optimizer and Memory Utilization Enabled). Identifique las instancias con un uso de CPU y de memoria máximos inferiores al 40 % durante un periodo de cuatro semanas. Estas son las instancias que

hay que dimensionar correctamente para reducir costos. Para los recursos de almacenamiento como Amazon S3, puede utilizar la Lente de almacenamiento de Amazon S3, que le permite ver 28 métricas de varias categorías por bucket y 14 días de datos históricos en el panel de control de forma predeterminada. Puede filtrar su panel de Lente de almacenamiento de Amazon S3 por resumen y optimización de costos o eventos para analizar métricas específicas.

- Consulta de las recomendaciones de ajuste de tamaño: utilice las recomendaciones de ajuste de tamaño de AWS Compute Optimizer y la herramienta de ajuste de tamaño de Amazon EC2 de la consola de gestión de costos, o revise el tamaño correcto de sus recursos para hacer ajustes en su carga de trabajo de AWS Trusted Advisor. Es importante utilizar las herramientas adecuadas a la hora de dimensionar los distintos recursos y seguir las pautas de dimensionamiento correctas, ya sea una instancia de Amazon EC2, clases de almacenamiento de AWS o tipos de instancias de Amazon RDS. En el caso de los recursos de almacenamiento, puede utilizar la Lente de almacenamiento de Amazon S3, que le ofrece visibilidad sobre el uso del almacenamiento de objetos, las tendencias de actividad y le proporciona recomendaciones prácticas para optimizar los costos y aplicar las prácticas recomendadas de protección de datos. Con las recomendaciones contextuales que la Lente de almacenamiento de Amazon S3 obtiene del análisis de las métricas de su organización, puede tomar medidas inmediatas para optimizar su almacenamiento.
- Selección automática del tipo y el tamaño del recurso en función de las métricas: utilice las métricas de la carga de trabajo para seleccionar los recursos de la carga de trabajo de forma manual o automática. En el caso de los recursos de computación, configurar AWS Auto Scaling o implementar el código en su aplicación puede reducir el esfuerzo necesario si deben hacerse cambios frecuentes, y así podrá implementar cambios potenciales antes que con el proceso manual. Puede lanzar y escalar automáticamente una flota de instancias en diferido e instancias de spot en un solo grupo de Auto Scaling. Además de los descuentos relacionados con las instancias de spot, puede utilizar instancias reservadas o un Savings Plan para conseguir mejores precios de los habituales en las instancias en diferido. Todos estos factores combinados le ayudarán a optimizar el ahorro de costos de las instancias de Amazon EC2 y a determinar la escala y el rendimiento que desea para su aplicación. También puede utilizar una estrategia de selección de tipo de instancia basada en atributos (ABS) en los grupos de escalado automático (ASG), lo que le permite expresar sus requisitos de instancia como un conjunto de atributos, como, por ejemplo, vCPU, memoria y almacenamiento. Puede utilizar automáticamente los tipos de instancia de nueva generación cuando se lancen y acceder a una gama más amplia de capacidad con las instancias de spot de Amazon EC2. Flota de Amazon EC2 y Amazon EC2 Auto Scaling seleccionan y lanzan instancias que se ajusten a los atributos especificados, por lo que no es necesario elegir manualmente los tipos de instancia. En cuanto a los recursos de almacenamiento, puede utilizar las características Amazon S3 Intelligent-Tiering y Amazon EFS Infrequent Access,

que le permiten seleccionar automáticamente las clases de almacenamiento que ofrecen ahorros automáticos en los costos de almacenamiento cuando cambian los patrones de acceso a los datos, sin que ello afecte al rendimiento ni a la sobrecarga operativa.

Recursos

Documentos relacionados:

- AWS Auto Scaling
- Tamaño correcto de AWS
- AWS Compute Optimizer
- Características de Amazon CloudWatch
- Configuración inicial de CloudWatch
- Publicar métricas personalizadas de CloudWatch
- Getting Started with Amazon EC2 Auto Scaling
- Amazon S3 Storage Lens
- Amazon S3 Intelligent-Tiering
- Amazon EFS Infrequent Access
- Launch an Amazon EC2 Instance Using the SDK

Videos relacionados:

- Right Size Your Services

Ejemplos relacionados:

- Attribute based Instance Type Selection for Auto Scaling for Amazon EC2 Fleet
- Optimizing Amazon Elastic Container Service for cost using scheduled scaling
- Predictive scaling with Amazon EC2 Auto Scaling
- Optimice los costos y obtenga visibilidad sobre el uso gracias a la Lente de almacenamiento de Amazon S3

COST06-BP04 Consideración del uso de los recursos compartidos

En el caso de los servicios ya implementados en la organización para varias unidades de negocio, plantéese la posibilidad de usar los recursos compartidos para aumentar su uso y reducir el costo total de propiedad (TCO). El uso de recursos compartidos puede ser una opción rentable para centralizar la administración y los costos mediante el uso de las soluciones existentes, el uso compartido de componentes o ambas opciones. Administre funciones comunes, como la supervisión, las copias de seguridad y la conectividad, ya sea dentro de los límites de una cuenta o en una cuenta dedicada. También puede reducir los costos mediante la implementación de la estandarización y la reducción de la duplicación y la complejidad.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

Cuando varias cargas de trabajo provoquen la misma función, utilice las soluciones existentes y los componentes compartidos para mejorar la administración y optimizar los costos. Considere la posibilidad de utilizar los recursos existentes (especialmente los compartidos), como, por ejemplo, los servidores de bases de datos o los servicios de directorio que no sean de producción, para reducir los costos de la nube. Para ello, siga las directrices indicadas en las prácticas recomendadas de seguridad y las normativas de la organización. Para generar valor y lograr un nivel de eficiencia de forma óptima, es fundamental volver a asignar los costos (mediante el análisis y los reintegros) a las áreas pertinentes de la empresa que impulsan el uso de recursos.

El análisis se refiere a los informes que desglosan los costos de la nube en categorías atribuibles, como consumidores, unidades de negocio, cuentas de contabilidad general u otras entidades responsables. El objetivo del análisis es mostrar a los equipos, las unidades de negocio o las personas el costo de los recursos de nube que se utilizan.

El reintegro consiste en asignar los gastos del servicio central a las unidades de costo en función de una estrategia adecuada para un proceso de gestión financiera específico. En el caso de los clientes, el reintegro cobra el costo generado desde una cuenta de servicios compartidos a diferentes categorías de costos financieros adecuadas para un proceso de generación de informes de clientes. Al establecer mecanismos de reintegro, puede informar de los costos generados por diferentes unidades de negocio, productos y equipos.

Las cargas de trabajo se pueden clasificar en esenciales y no esenciales. Según esta clasificación, utilice recursos compartidos con configuraciones generales para cargas de trabajo menos

esenciales. Para optimizar aún más los costos, reserve servidores dedicados exclusivamente para cargas de trabajo esenciales. Comparta recursos o aprovisiónelos en varias cuentas para administrarlos de manera eficaz. Incluso con distintos entornos de desarrollo, pruebas y producción, es posible compartir de forma segura sin que se vea afectada la estructura organizativa.

Para mejorar sus conocimientos y optimizar el costo y el uso de las aplicaciones en contenedores, utilice datos de asignación de costos divididos que le ayudan a asignar los costos a entidades empresariales individuales en función de cómo la aplicación consume los recursos de computación y memoria compartidos. Los datos de asignación de costos divididos le ayudan a lograr análisis y reintegros por tareas en cargas de trabajo del contenedor que se ejecutan en Amazon Elastic Container Service (Amazon ECS) o Amazon Elastic Kubernetes Service (Amazon EKS).

En el caso de arquitecturas distribuidas, cree una VPC de servicios compartidos, que ofrezca acceso centralizado a los servicios compartidos que requieren las cargas de trabajo en cada una de las VPC. En estos servicios compartidos se pueden incluir recursos como servicios de directorio o puntos de conexión de VPC. Para reducir los costos generales y administrativos, comparta los recursos desde una ubicación central en lugar de crearlos en cada VPC.

Al utilizar recursos compartidos, puede ahorrar en costos operativos, sacar el máximo partido de la utilización de los recursos y mejorar la coherencia. En un diseño de varias cuentas, puede alojar algunos servicios de AWS de forma centralizada y acceder a ellos mediante varias aplicaciones y cuentas en un centro para ahorrar costos. Puede usar AWS Resource Access Manager (AWS RAM) para compartir otros recursos comunes, como subredes de VPC y adjuntos de AWS Transit Gateway, AWS Network Firewall o canalizaciones de Amazon SageMaker AI. En un entorno de varias cuentas, utilice AWS RAM para crear un recurso una vez y compartirlo con otras cuentas.

Las organizaciones deben etiquetar los costos compartidos de forma eficaz y verificar que no tengan una parte considerable de sus costos sin etiquetar o sin asignar. Si no se asignan los costos compartidos de forma eficaz y nadie se hace responsable de la administración de estos costos, los costos compartidos de la nube pueden aumentar vertiginosamente. Debe saber dónde se han producido costos en el nivel de recursos, carga de trabajo, equipo u organización, ya que podrá comprender mejor el valor entregado en el nivel aplicable en comparación con los resultados empresariales logrados. En definitiva, las organizaciones sacan partido de los ahorros de costos como resultado del uso compartido de la infraestructura en la nube. Fomente la asignación de costos en los recursos compartidos de la nube para optimizar el gasto en la nube.

Pasos para la implementación

- Evaluación de los recursos existentes: revise las cargas de trabajo existentes que utilizan servicios similares para su carga de trabajo. En función de los componentes de la carga de trabajo, tenga en cuenta las plataformas existentes si la lógica empresarial o los requisitos técnicos lo permiten.
- Uso compartido de recursos de AWS RAM y restricción en consecuencia: utilice AWS RAM para compartir recursos con otras cuentas de AWS de su organización. Al compartir recursos, no necesita duplicarlos en varias cuentas, lo que reduce al mínimo la carga operativa del mantenimiento de los recursos. Este proceso también le ayuda a compartir de forma segura los recursos que ha creado con roles y usuarios de su cuenta y con otras Cuentas de AWS.
- Etiquetado de recursos: etiquete los recursos que sean candidatos para la elaboración de informes de costos y clasifíquelos dentro de las categorías de costos. Active estas etiquetas de recursos relacionados con los costos para la asignación de costos con el fin de proporcionar visibilidad del uso de los recursos de AWS. Céntrese en crear un nivel adecuado de especificidad con respecto a la visibilidad de los costos y el uso, e influya en los comportamientos de consumo de la nube mediante los informes de asignación de costos y el seguimiento de los KPI.

Recursos

Prácticas recomendadas relacionadas:

- SEC03-BP08 Uso compartido de recursos de forma segura en su organización

Documentos relacionados:

- What is AWS Resource Access Manager?
- Servicios de AWS que se pueden utilizar con AWS Organizations
- Recursos de AWS que se pueden compartir
- AWS Cost and Usage (CUR) Queries

Videos relacionados:

- AWS Resource Access Manager - granular access control with managed permissions
- How to design your AWS cost allocation strategy
- AWS Cost Categories

Ejemplos relacionados:

- How-to chargeback shared services: An AWS Transit Gateway example
- How to build a chargeback/showback model for Savings Plans using the CUR
- Using VPC Sharing for a Cost-Effective Multi-Account Microservice Architecture
- Improve cost visibility of Amazon EKS with AWS Split Cost Allocation Data
- Improve cost visibility of Amazon ECS and AWS Batch with AWS Split Cost Allocation Data

Selección del mejor modelo de precios

Modelado de costos de la carga de trabajo: considere los requisitos de los componentes de la carga de trabajo y comprenda los posibles modelos de precios. Defina el requisito de disponibilidad del componente. Determine si hay varios recursos independientes que ejecuten la función en la carga de trabajo y cuáles son los requisitos de la carga de trabajo a lo largo del tiempo. Compare el costo de los recursos con el modelo de precios bajo demanda predeterminado y otros modelos aplicables. Tenga en cuenta cualquier cambio potencial en los recursos o en los componentes de la carga de trabajo.

Análisis periódicos por cuenta: hacer modelados de costos periódicos garantiza que se puedan implementar oportunidades de optimización en múltiples cargas de trabajo. Por ejemplo, si varias cargas de trabajo usan instancias bajo demanda, en un nivel agregado el riesgo de cambio es menor, e implementar un descuento por compromiso tendrá un costo general inferior. Se recomienda hacer análisis en ciclos regulares de dos semanas a un mes. Este análisis le permite hacer compras de ajustes pequeños para que sus modelos de precios puedan seguir evolucionando a medida que cambien sus cargas de trabajo y sus componentes.

Utilice la herramienta de recomendaciones del AWS Cost Explorer para buscar oportunidades de descuentos por compromiso.

Para buscar oportunidades para cargas de trabajo de spot, use una vista por hora del uso general y busque periodos regulares de uso cambiante o de elasticidad.

Modelos de precios: AWS tiene varios modelos de precios que le permiten pagar por sus recursos de la manera más rentable que se adapte a las necesidades de su organización. En la siguiente sección se describe cada modelo de compra:

- instancias bajo demanda
- Spot Instances

- Descuentos por compromiso: Savings Plans
- Descuentos por compromiso: instancias o capacidad reservadas
- Selección geográfica
- Precios y acuerdos de terceros

Instancias bajo demanda: este es el modelo predeterminado de precios de pago por uso. Cuando utiliza recursos (por ejemplo, instancias de EC2 o servicios como DynamoDB bajo demanda), paga una tarifa fija sin compromiso a largo plazo. Puede aumentar o disminuir la capacidad de sus recursos o servicios según las demandas de su aplicación. Las instancias bajo demanda tienen una tarifa por hora, pero, según el servicio, se pueden facturar en incrementos de un segundo (por ejemplo, instancias de EC2 de Linux o Amazon RDS). Se recomienda la opción bajo demanda para aplicaciones con cargas de trabajo a corto plazo (por ejemplo, un proyecto de cuatro meses) con picos periódicos o cargas de trabajo impredecibles que no se puedan interrumpir. La opción bajo demanda también es adecuada para cargas de trabajo, como los entornos de preproducción, que requieren tiempos de ejecución ininterrumpidos, pero que no duran lo suficiente como para obtener un descuento por compromiso (Savings Plans o instancias reservadas).

Instancias de spot: una instancia de spot es la capacidad de computación sobrante de Amazon EC2 que está disponible con descuentos de hasta un 90 % sobre los precios bajo demanda, sin necesidad de un compromiso a largo plazo. Con las instancias de spot, puede reducir significativamente el costo de ejecución de sus aplicaciones o escalar la capacidad de computación de sus aplicaciones con el mismo presupuesto. A diferencia de las instancias bajo demanda, las de spot se pueden interrumpir con una advertencia de 2 minutos si Amazon EC2 necesita recuperar la capacidad o el precio de la instancia de spot supera el configurado. De media, las instancias de spot se interrumpen menos del 5 % de las veces.

Las instancias de spot son ideales cuando hay una cola o un búfer, o cuando hay varios recursos que trabajan de forma independiente para procesar las solicitudes (por ejemplo, el procesamiento de datos de Hadoop). Por lo general, estas cargas de trabajo son tolerantes a errores, sin estado y flexibles, como el procesamiento por lotes, los macrodatos y el análisis, los entornos en contenedores y la computación de alto rendimiento (HPC). Las cargas de trabajo no críticas, como los entornos de pruebas y desarrollo, también son aptas para spot.

Las instancias de spot también están integradas en varios servicios de AWS, como los grupos de Amazon EC2 Auto Scaling, Amazon EMR, Amazon Elastic Container Service (Amazon ECS) y AWS Batch.

Cuando es necesario recuperar una instancia de spot, Amazon EC2 envía una advertencia de dos minutos mediante un aviso de interrupción de la instancia de spot que se envía a través de Eventos de CloudWatch, así como en los metadatos de la instancia. Durante ese periodo de dos minutos, la aplicación puede utilizar ese tiempo para guardar su estado, vaciar los contenedores en ejecución, cargar los archivos de registro finales o eliminarse automáticamente de un equilibrador de carga. Al cabo de los dos minutos, tiene la opción de hibernar, detener o terminar la instancia de spot.

Tenga en cuenta las siguientes prácticas recomendadas al adoptar las instancias de spot en las cargas de trabajo:

- Sea flexible con tantos tipos de instancias como sea posible: sea flexible tanto en la familia como en el tamaño del tipo de instancia para aumentar la probabilidad de cumplir con sus requisitos de capacidad objetivo, obtener el menor costo posible y minimizar el impacto de las interrupciones.
- Sea flexible en cuanto a dónde se ejecutará su carga de trabajo: la capacidad disponible puede variar según la zona de disponibilidad. Esto aumenta la probabilidad de cumplir con la capacidad objetivo al aprovechar varios grupos de capacidad sobrante y proporciona el menor costo posible.
- Diseñe para garantizar la continuidad: diseñe sus cargas de trabajo teniendo en cuenta la ausencia de estado y la tolerancia a errores, de modo que, si parte de la capacidad de EC2 se interrumpe, esto no repercuta en la disponibilidad ni el rendimiento de la carga de trabajo.
- Recomendamos usar instancias de spot en combinación con instancias bajo demanda y Savings Plans o instancias reservadas para maximizar la optimización de costos de la carga de trabajo con el rendimiento.

Descuentos por compromiso (Savings Plans): AWS le ofrece varias formas de reducir los costos al reservar o comprometerse a usar una cantidad determinada de recursos y recibir una tarifa con descuento para sus recursos. Un plan de Savings Plans le permite comprometerse a un gasto por hora durante uno o tres años y recibir descuentos en los precios en todos sus recursos. Los Savings Plans ofrecen descuentos en servicios de computación de AWS como Amazon EC2, AWS Fargate y AWS Lambda. Cuando se compromete, paga ese importe cada hora y se deduce del uso bajo demanda a la tasa de descuento. Por ejemplo, se compromete a pagar 50 USD por hora y tiene 150 USD por hora de uso bajo demanda. Teniendo en cuenta los precios de Savings Plans, su uso específico tiene una tasa de descuento del 50 %. Por lo tanto, el compromiso de 50 USD cubre 100 USD de uso bajo demanda. Pagará 50 USD (compromiso) y 50 USD del resto del uso bajo demanda.

Los Savings Plans para computación son los más flexibles y ofrecen descuentos de hasta el 66 %. Se aplican automáticamente en todas las zonas de disponibilidad, tamaños de instancias, familias de instancias, sistemas operativos, tenencias, regiones y servicios de computación.

Los Savings Plans para instancias tienen menos flexibilidad, pero ofrecen una tasa de descuento más alta (hasta un 72 %). Se aplican automáticamente en todas las zonas de disponibilidad, tamaños de instancias, sistemas operativos y tenencias.

Hay tres opciones de pago:

- Sin pago inicial: no se exige ningún pago inicial; pague una tarifa por hora reducida cada mes según el total de horas del mes más adelante.
- Pago inicial parcial: ofrece una tasa de descuento más alta que la de Sin pago inicial. Parte del uso se paga al inicio; pague una tarifa por hora reducida más baja cada mes según el total de horas del mes más adelante.
- Pago inicial total: el uso de todo el periodo se paga por adelantado y no se incurre en ningún otro costo durante el resto del periodo por el uso cubierto por el compromiso.

Puede aplicar cualquier combinación de estas tres opciones de compra a las cargas de trabajo.

Los Savings Plans se aplican primero al uso de la cuenta en la que se han comprado, desde el porcentaje de descuento más alto al más bajo, y luego se aplican al uso consolidado de todas las demás cuentas, desde el porcentaje de descuento más alto hasta el más bajo.

Se recomienda comprar todos los Savings Plans en una cuenta sin uso ni recursos, como la cuenta de administración. Esto garantiza que el plan de Savings Plans se aplique a las tasas de descuento más altas de todo su uso, lo que maximiza el importe del descuento.

Las cargas de trabajo y el uso suelen cambiar con el tiempo. Se recomienda adquirir continuamente pequeñas cantidades de compromiso de Savings Plans a lo largo del tiempo. Esto garantiza que mantenga altos niveles de cobertura para maximizar sus descuentos y que sus planes se ajusten perfectamente a su carga de trabajo y a los requisitos de la organización en todo momento.

No establezca un objetivo de cobertura en sus cuentas, debido a la posible variabilidad del descuento. Una cobertura baja no necesariamente indica un alto potencial de ahorro. Es posible que la cobertura de su cuenta sea baja, pero si su uso se compone de instancias pequeñas con un sistema operativo con licencia, el ahorro potencial podría reducirse a un pequeño porcentaje. En su lugar, haga un seguimiento de los posibles ahorros disponibles en la herramienta de recomendación

de Savings Plans y supervíselos. Revise con frecuencia las recomendaciones de Savings Plans en el Explorador de costos (haga análisis periódicos) y continúe adquiriendo compromisos hasta que los ahorros estimados estén por debajo del descuento requerido para la organización. Por ejemplo, haga un seguimiento y supervise que los posibles descuentos se mantengan por debajo del 20 %; si se supera ese porcentaje, se debe hacer una compra.

Supervise la utilización y la cobertura, pero solo para detectar cambios. No busque un porcentaje de utilización o un porcentaje de cobertura específicos, ya que esto no necesariamente se escala con los ahorros. Asegúrese de que la compra de Savings Plans se traduzca en un aumento de la cobertura y, si hay disminuciones en la cobertura o la utilización, asegúrese de cuantificarlas y conocerlas. Por ejemplo, se migra un recurso de carga de trabajo a un tipo de instancia más reciente, lo que reduce la utilización de un plan existente, pero el beneficio de rendimiento supera la reducción del ahorro.

Descuentos por compromiso (instancias reservadas o compromiso): al igual que los Savings Plans, las instancias reservadas ofrecen descuentos de hasta el 72 % si se compromete a ejecutar una cantidad mínima de recursos. Las instancias reservadas están disponibles para Amazon RDS, Amazon OpenSearch Service, Amazon ElastiCache, Amazon Redshift y DynamoDB. Amazon CloudFront y AWS Elemental MediaConvert también ofrecen descuentos cuando contrae compromisos de uso mínimo. Las instancias reservadas están disponibles actualmente para Amazon EC2, pero los Savings Plans ofrecen los mismos niveles de descuento con mayor flexibilidad y sin gastos de administración.

Las instancias reservadas ofrecen las mismas opciones de precios (sin pago inicial, pago inicial parcial y pago inicial total) y los mismos plazos de uno o tres años.

Las instancias reservadas se pueden adquirir en una región o en una zona de disponibilidad específica. Proporcionan una reserva de capacidad cuando se compran en una zona de disponibilidad.

Aunque Amazon EC2 incluye instancias reservadas convertibles, los Savings Plans deben utilizarse en todas las instancias de EC2 debido a su mayor flexibilidad y a la reducción de los costos operativos.

Se deben utilizar el mismo proceso y las mismas métricas para hacer el seguimiento de las instancias reservadas y comprarlas. Se recomienda no hacer un seguimiento de la cobertura de las instancias reservadas en sus cuentas. También se recomienda no hacer un seguimiento del porcentaje de uso ni supervisarlo, sino que se aconseja consultar el informe de utilización en el Explorador de costos y utilizar la columna de ahorro neto de la tabla. Si el ahorro neto es un importe

negativo significativamente elevado, debe tomar medidas para corregir la instancia reservada no utilizada.

Flota de EC2: Flota de EC2 es una característica que le permite definir una capacidad de computación objetivo y especificar los tipos de instancias y el saldo de las instancias de la flota bajo demanda y de spot. A continuación, Flota de EC2 lanzará automáticamente la combinación de recursos con el precio más bajo a fin de satisfacer la capacidad definida.

Selección geográfica: al diseñar soluciones, una práctica recomendada es intentar colocar los recursos de computación más cerca de los usuarios a fin de brindar una latencia más baja y una soberanía de datos sólida. En el caso de audiencias globales, debe usar varias ubicaciones para satisfacer estas necesidades. Debe seleccionar la ubicación geográfica que minimice sus costos.

La infraestructura de la Nube de AWS se compone de regiones y zonas de disponibilidad. Una región es una ubicación física en el mundo donde disponemos de varias zonas de disponibilidad. Las zonas de disponibilidad constan de uno o varios centros de datos separados, cada uno de ellos con alimentación, redes y conectividad redundantes, que se alojan en instalaciones independientes.

Cada Región de AWS opera dentro de las condiciones del mercado local y el precio de los recursos varía en cada región. Elija una región específica en la que desee aplicar un componente de su solución o la solución completa a fin de poder ejecutar al precio más bajo posible a nivel mundial. Puede utilizar la sencilla Calculadora mensual de AWS para estimar los costos de su carga de trabajo en varias regiones.

Acuerdos y precios de terceros: cuando utilice soluciones o servicios de terceros en la nube, es importante que las estructuras de precios estén alineadas con los resultados de la optimización de costos. Los precios deben ir a la par de los resultados y el valor que aportan. Un ejemplo de ello es el software que se lleva una parte del ahorro que proporciona: cuanto más ahorra (resultado), más cobra. Los acuerdos que se escalan con su factura no suelen estar alineados con la optimización de costos, a menos que proporcionen resultados para cada parte de su factura específica. Por ejemplo, una solución que proporciona recomendaciones para Amazon EC2 y cobra un porcentaje de toda la factura será más cara si usa otros servicios para los que no genera ningún beneficio. Otro ejemplo es un servicio administrado que se cobra a un porcentaje del costo de los recursos que se administran. Un mayor tamaño de la instancia no tiene por qué requerir un mayor esfuerzo de administración, aunque sí se cobrará más. A fin de impulsar la eficiencia, asegúrese de que, en estos acuerdos de precios del servicio se incluya un programa o características de optimización de costos en su servicio.

Prácticas recomendadas

- COST07-BP01 Análisis de los modelos de precios
- COST07-BP02 Elección de regiones según el costo
- COST07-BP03 Selección de acuerdos de terceros con condiciones rentables
- COST07-BP04 Implementación de modelos de precios para todos los componentes de la carga de trabajo
- COST07-BP05 Análisis de modelos de precios en el nivel de la cuenta de administración

COST07-BP01 Análisis de los modelos de precios

Analice cada componente de la carga de trabajo. Determine si el componente y los recursos se ejecutarán durante periodos extensos (por descuentos por compromiso) o periodos dinámicos y de corta ejecución (para spot o bajo demanda). Haga un análisis de la carga de trabajo mediante las recomendaciones de las herramientas de administración de costos y aplique las reglas empresariales a dichas recomendaciones para conseguir un alto rendimiento.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

AWS tiene varios modelos de precios que le permiten pagar sus recursos de la manera más rentable que se adapte a las necesidades de su organización y en función del producto. Determine con sus equipos el modelo de precios más apropiado. Su modelo de precios suele constar de una combinación de varias opciones, según lo determine su disponibilidad

Las instancias bajo demanda le permiten pagar la capacidad de computación o de la base de datos por hora o por segundo (60 segundos como mínimo) en función de las instancias que ejecute, sin necesidad de asumir compromisos a largo plazo ni de hacer pagos iniciales.

Savings Plans es un modelo de precios flexible que ofrece precios económicos por el uso de Amazon EC2, Lambda y AWS Fargate a cambio de comprometerse a una cantidad constante de uso (medida en USD/hora) durante un plazo de uno o tres años.

Las instancias de spot son un mecanismo de precios de Amazon EC2 que le permite solicitar capacidad de computación sobrante a una tarifa horaria reducida (hasta un 90 % menos sobre el precio bajo demanda) sin compromiso previo.

Las instancias reservadas le ofrecen hasta un 75 % de descuento si paga por adelantado la capacidad. Para obtener más información, consulte Optimización de costos con reservas.

Puede incluir Savings Plans para los recursos asociados a los entornos de producción, calidad y desarrollo. Como alternativa, dado que los recursos de entorno aislado solo se activan cuando es necesario, puede elegir un modelo bajo demanda para los recursos de ese entorno. Utilice las instancias de spot de Amazon para reducir los costos de Amazon EC2 o utilice Savings Plans para computación para reducir los costos de Amazon EC2, Fargate y Lambda. La herramienta de recomendaciones de AWS Cost Explorer ofrece oportunidades de descuentos por compromiso con los Savings Plans.

Si ha adquirido instancias reservadas para Amazon EC2 en el pasado o ha establecido prácticas de asignación de costos en su organización, puede seguir utilizando las instancias reservadas de Amazon EC2 por el momento. Sin embargo, recomendamos elaborar una estrategia para usar Savings Plans en el futuro como mecanismo más flexible de ahorro de costos. Puede actualizar las recomendaciones de Savings Plans (SP) en AWS Cost Management para generar nuevas recomendaciones de Savings Plans en cualquier momento. Use las instancias reservadas para reducir los costos de Amazon RDS, Amazon Redshift, Amazon ElastiCache y Amazon OpenSearch Service. Los Savings Plans y las instancias reservadas están disponibles en tres modalidades de pago: puede abonarse el total por adelantado, abonarse parte por adelantado y no abonarse nada por adelantado. Utilice las recomendaciones de compra de instancias reservadas y Savings Plans de AWS Cost Explorer.

Para buscar oportunidades para cargas de trabajo de spot, use una vista por hora del uso general y busque periodos regulares de uso cambiante o de elasticidad. Puede utilizar instancias de spot para diversas aplicaciones flexibles y tolerantes a errores. Algunos ejemplos son los servidores web sin estado, los puntos de conexión de API, las aplicaciones de macrodatos y análisis, las cargas de trabajo en contenedores, CI/CD y otras cargas de trabajo flexibles.

Analice sus instancias de Amazon EC2 y Amazon RDS si pueden desactivarse cuando no las utilice (fuera de horario laboral y en fines de semana). Este enfoque le permitirá reducir costos en un 70 % o más con respecto a su uso ininterrumpido. Si tiene clústeres de Amazon Redshift que solo deben estar disponibles en momentos concretos, puede pausar el clúster y reanudarlo más tarde. Cuando se detiene el clúster de Amazon Redshift o la instancia de Amazon EC2 y Amazon RDS, la facturación de computación se detiene y solo se aplica el cargo por almacenamiento.

Tenga en cuenta que las reservas de capacidad bajo demanda (ODCR) no son un descuento en el precio. Las reservas de capacidad se cobrarán según la tarifa bajo demanda equivalente, independientemente de que ejecute instancias en la capacidad reservada o no. Deben tenerse en cuenta cuando necesite proporcionar suficiente capacidad para los recursos que tiene previsto ejecutar. Las ODCR no tienen por qué estar vinculadas a compromisos a largo plazo, ya que pueden

cancelarse cuando ya no las necesite, pero también pueden beneficiarse de los descuentos que ofrecen los Savings Plans o las instancias reservadas.

Pasos para la implementación

- Análisis de la elasticidad de la carga de trabajo: mediante el grado de detalle por horas del Explorador de costos o un panel personalizado, analice la elasticidad de la carga de trabajo. Busque cambios regulares en el número de instancias que se están ejecutando. Las instancias de corta duración son candidatas para las instancias o la flota de spot.
  - Well-Architected Lab: Cost Explorer
  - Well-Architected Lab: Cost Visualization
- Revisión de los contratos de precios existentes: revise los contratos o compromisos actuales para determinar las necesidades a largo plazo. Analice lo que tiene actualmente y en qué medida se utilizan esos compromisos. Aproveche los descuentos contractuales o los acuerdos empresariales preexistentes. Los contratos Enterprise ofrecen a los clientes la opción de personalizar los acuerdos que mejor se adapten a sus necesidades. En el caso de compromisos a largo plazo, considere los descuentos por precios reservados, las instancias reservadas o Savings Plans para el tipo de instancia específico, la familia de instancias, la Región de AWS y las zonas de disponibilidad.
- Análisis de descuento por compromiso: utilice el Explorador de costos en su cuenta y revise las recomendaciones de Savings Plans e instancias reservadas. Para comprobar que está implementando las recomendaciones correctas con los descuentos y el riesgo necesarios, siga los laboratorios de Well-Architected.

Recursos

Documentos relacionados:

- Accessing Reserved Instance recommendations
- Opciones de compra de instancias
- AWS Enterprise

Videos relacionados:

- Save up to 90% and run production workloads on Spot

Ejemplos relacionados:

- Well-Architected Lab: Cost Explorer
- Well-Architected Lab: Cost Visualization
- Well-Architected Lab: Pricing Models

COST07-BP02 Elección de regiones según el costo

Los precios de los recursos pueden variar según la región. Identifique las diferencias regionales de costos y efectúe la implementación solo en las regiones con costos más elevados para cumplir los requisitos de latencia, residencia de datos y soberanía de los datos. Si tiene en cuenta el costo de la región, podrá pagar el precio global más bajo por esta carga de trabajo.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

La infraestructura de la Nube de AWS es global, está alojada en múltiples ubicaciones en todo el mundo y se basa en Regiones de AWS, zonas de disponibilidad, zonas locales, AWS Outposts y zonas de Wavelength. Una región es una ubicación física en el mundo y cada región es un área geográfica independiente en la que AWS tiene varias zonas de disponibilidad. Las zonas de disponibilidad, que son varias ubicaciones aisladas en cada región, constan de uno o varios centros de datos separados, cada uno de ellos con alimentación, redes y conectividad redundantes.

Cada Región de AWS opera según las condiciones del mercado local y el precio de los recursos es distinto en cada región debido a las diferencias del costo del terreno, la fibra, la electricidad y los impuestos, por ejemplo. Elija una región específica en la que desee aplicar un componente de su solución o la solución completa a fin de poder ejecutar al precio más bajo posible a nivel mundial. Utilice la Calculadora de AWS para calcular los costos de su carga de trabajo en varias regiones mediante la búsqueda de servicios por tipo de ubicación (región, zona de Wavelength y zona local) y región.

Al diseñar soluciones, una práctica recomendada es intentar colocar los recursos de computación más cerca de los usuarios a fin de brindar una latencia más baja y una soberanía de datos sólida. Seleccione la ubicación geográfica en función de los requisitos de su empresa, privacidad de datos, rendimiento y seguridad. En el caso de aplicaciones con usuarios finales en todo el mundo, utilice varias ubicaciones.

Recurra a las regiones que ofrecen precios más bajos por los servicios de AWS para implementar sus cargas de trabajo si no tiene obligaciones en materia de privacidad de datos, seguridad y requisitos de empresa. Por ejemplo, si su región predeterminada es Asia-Pacífico (Sídney) (`ap-southwest-2`) y si no existen restricciones (privacidad de los datos o seguridad, por ejemplo) para utilizar otras regiones, implementar instancias de Amazon EC2 no críticas (desarrollo y pruebas) en la región Este de EE. UU. (Norte de Virginia) (`us-east-1`) tendrá menos costos.

	Cumplimiento	*Latencia*	*Coste*	*Servicios/características*
Región 1	✓	15 ms	$$	✓
Región 2	✓	20 ms	$$$	X
Región 3	✓	80 ms	$	✓
Región 4	✓	15 ms	$$	✓
Región 5	✓	20 ms	$$$	X
Región 6	✓	15 ms	$	✓
Región 7	✓	80 ms	$	✓
Región 8	✓	15 ms	$	X

Tabla matricial de características de las regiones

En la tabla matricial anterior se nos muestra que la Región 6 es la mejor opción para este escenario específico, porque la latencia es baja en comparación con otras regiones, el servicio está disponible y es la región menos cara.

Pasos para la implementación

- Revisión de los precios de la Región de AWS: analice los costos de la carga de trabajo de la región actual. A partir de los costos más elevados por servicio y tipo de uso, calcule los costos en otras regiones que estén disponibles. Si el ahorro previsto supera el costo de trasladar el componente o la carga de trabajo, migre a la nueva región.
- Revisión de los requisitos para las implementaciones de varias regiones: analice los requisitos y las obligaciones de su empresa (privacidad de los datos, seguridad o rendimiento) para averiguar

si existe alguna restricción que le impida utilizar varias regiones. Si no hay obligaciones que restrinjan el uso de una sola región, utilice varias.

- Revisión de la transferencia de datos requerida: tenga en cuenta los costos de transferencia de datos al seleccionar las regiones. Mantenga sus datos cerca de su cliente y de los recursos. Seleccione Regiones de AWS menos costosas donde fluyan los datos y donde la transferencia de datos sea mínima. En función de los requisitos empresariales en materia de transferencia de datos, puede utilizar Amazon CloudFront, AWS PrivateLink, AWS Direct Connect y AWS Virtual Private Network para reducir los costos de red, y mejorar el rendimiento y la seguridad.

Recursos

Documentos relacionados:

- Accessing Reserved Instance recommendations
- Precios de Amazon EC2
- Opciones de compra de instancias
- Tabla de regiones

Videos relacionados:

- Save up to 90% and run production workloads on Spot

Ejemplos relacionados:

- Overview of Data Transfer Costs for Common Architectures
- Cost Considerations for Global Deployments
- What to Consider when Selecting a Region for your Workloads

COST07-BP03 Selección de acuerdos de terceros con condiciones rentables

Los acuerdos y condiciones rentables garantizan que el costo de estos servicios vaya a la par de los beneficios que proporcionan. Seleccione acuerdos y precios que se escalen cuando proporcionen beneficios adicionales a la organización.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

Hay varios productos en el mercado que pueden ayudarlo a administrar los costos en sus entornos de nube. Puede que haya algunas diferencias en lo que se refiere a las características que dependen de los requisitos del cliente. Por ejemplo, puede que algunos se centren en la gobernanza de costos o la visibilidad de costos y otros en la optimización de costos. Un factor clave para que la optimización de los costos y la gobernanza sean efectivas es utilizar la herramienta adecuada con las características necesarias y el modelo de precios correcto. Estos productos tienen diferentes modelos de precios. Algunos cobran un porcentaje determinado de la factura mensual, mientras que otros cobran un porcentaje del ahorro que se consigue. Lo ideal es que pague solo lo que necesita.

Al utilizar soluciones o servicios de terceros en la nube, es importante que las estructuras de precios se ajusten a los resultados deseados. Los precios deben ir a la par de los resultados y el valor que aportan. Un ejemplo de ello es el software que se lleva una parte del ahorro que proporciona: cuanto más ahorra (resultado), más cobra. Los acuerdos de licencias en los que paga más a medida que aumentan sus gastos no siempre le convienen para optimizar costos. Sin embargo, si el proveedor ofrece ventajas claras en todas las partes de su factura, este aumento de tarifa podría estar justificado.

Por ejemplo, una solución que proporciona recomendaciones para Amazon EC2 y cobra un porcentaje de toda la factura podría ser más cara si usa otros servicios que no generan ningún beneficio. Otro ejemplo es un servicio administrado que se cobra a un porcentaje del costo de los recursos que se administran. Un mayor tamaño de la instancia no tiene por qué requerir un mayor esfuerzo de administración, aunque sí se podría cobrar más. A fin de impulsar la eficiencia, asegúrese de que, en estos acuerdos de precios del servicio, se incluya un programa o características de optimización de costos en su servicio.

Los clientes podrían encontrar en el mercado estos productos más avanzados o fáciles de usar. Debe considerar el costo de estos productos y pensar en los posibles resultados de optimización de costos a largo plazo.

Pasos para la implementación

- Análisis de los acuerdos y condiciones de terceros: revise los precios de los acuerdos de terceros. Haga modelados de los diferentes niveles de uso y tenga en cuenta nuevos costos, como el uso de nuevos servicios o incrementos en los servicios actuales debido al crecimiento de la carga de trabajo. Decida si los costos adicionales proporcionan los beneficios necesarios para su empresa.

Recursos

Documentos relacionados:

- Accessing Reserved Instance recommendations
- Opciones de compra de instancias

Videos relacionados:

- Save up to 90% and run production workloads on Spot

COST07-BP04 Implementación de modelos de precios para todos los componentes de la carga de trabajo

Al ejecutar recursos de forma permanente se debe utilizar la capacidad reservada como los Savings Plans o las instancias reservadas. La capacidad a corto plazo se configura con instancias o una flota de spot. Las instancias bajo demanda solo se usan para cargas de trabajo a corto plazo que no se pueden interrumpir y que no se ejecutan lo suficiente como para tener capacidad reservada, es decir, de un 25 a un 75 % del periodo, según el tipo de recurso.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

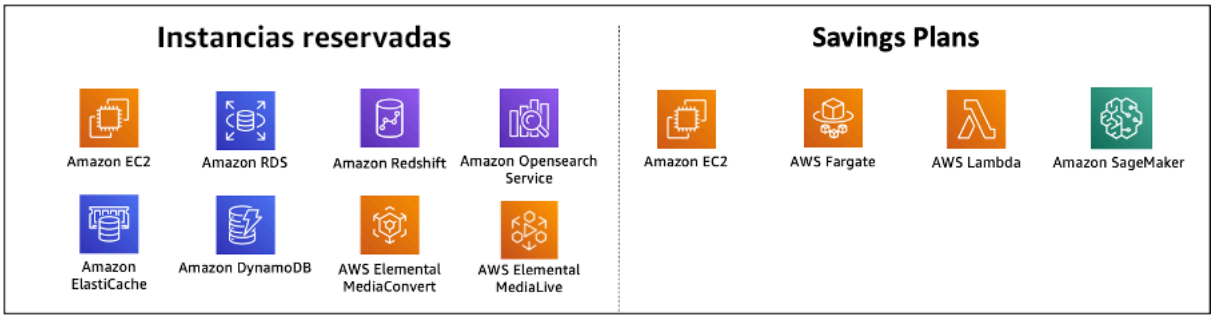
Guía para la implementación

Para mejorar la rentabilidad, AWS proporciona varias recomendaciones de compromiso basadas en el uso pasado. Estas recomendaciones pueden servirle para saber lo que puede ahorrar y cómo se utilizará el compromiso. Puede utilizar estos servicios como instancias bajo demanda o spot, o comprometerse por un periodo de tiempo determinado y reducir sus costos bajo demanda con instancias reservadas y Savings Plans (SP). Necesita conocer, no solo los componentes de cada carga de trabajo y los múltiples servicios de AWS, sino también los descuentos por compromiso, las opciones de compra y las instancias de spot de estos servicios para optimizar su carga de trabajo.

Tenga en cuenta los requisitos de los componentes de su carga de trabajo e infórmese de los diferentes modelos de precios de estos servicios. Defina el requisito de disponibilidad de estos componentes. Determine si hay varios recursos independientes que ejecuten la función en la carga de trabajo y cuáles son los requisitos de la carga de trabajo a lo largo del tiempo. Compare el costo de los recursos con el modelo de precios bajo demanda predeterminado y otros modelos aplicables.

Tenga en cuenta cualquier cambio potencial en los recursos o en los componentes de la carga de trabajo.

Veamos, por ejemplo, esta arquitectura de aplicaciones web en AWS. Este ejemplo de carga de trabajo consta de varios servicios de AWS, como Amazon Route 53, AWS WAF, Amazon CloudFront, instancias de Amazon EC2, instancias de Amazon RDS, equilibradores de carga, almacenamiento de Amazon S3 y Amazon Elastic File System (Amazon EFS). Debe revisar cada uno de estos servicios e identificar las posibles oportunidades de ahorro de costos con diferentes modelos de precios. Algunos de ellos podrían ser elegibles para RI o SP, mientras que otros podrían estar disponibles solo bajo demanda. Como se muestra en la siguiente imagen, algunos de los servicios de AWS pueden comprometerse mediante instancias reservadas o Savings Plans.



Servicios de AWS comprometidos mediante instancias reservadas y Savings Plans

Pasos para la implementación

- Implementación de modelos de precios: con los resultados de sus análisis, compre Savings Plans o instancias reservadas o implemente instancias de spot. Si se trata de su primera compra de compromiso, elija las cinco o diez mejores recomendaciones de la lista y, a continuación, supervise y analice los resultados durante uno o dos meses. La AWS Cost Management Console le guiará a lo largo del proceso. Revise las recomendaciones de instancias reservadas o Savings Plans desde la consola, personalice las recomendaciones (tipo, pago y plazo), revise el compromiso por hora (por ejemplo, 20 USD por hora) y, a continuación, agréguelas a la cesta. Los descuentos se aplican automáticamente al uso elegible. Compre una pequeña cantidad de descuentos por compromiso en ciclos regulares (por ejemplo, cada 2 semanas o mensualmente). Implemente instancias de spot para las cargas de trabajo que se puedan interrumpir o no tengan estado. Por último, seleccione instancias de Amazon EC2 bajo demanda y asigne recursos para los requisitos restantes.
- Ciclo de revisión de la carga de trabajo: implemente un ciclo de revisión de la carga de trabajo que analice específicamente la cobertura del modelo de precios. Cuando la carga de trabajo tenga la

cobertura requerida, compre descuentos por compromiso adicionales parcialmente (cada pocos meses) o a medida que cambie el uso en la organización.

Recursos

Documentos relacionados:

- Understanding your Savings Plans recommendations
- Accessing Reserved Instance recommendations
- Cómo adquirir instancias reservadas
- Opciones de compra de instancias
- Spot Instances
- Reservation models for other AWS services
- Savings Plans Supported Services

Videos relacionados:

- Save up to 90% and run production workloads on Spot

Ejemplos relacionados:

- ¿Qué debo tener en cuenta antes de comprar un Savings Plan?
- ¿Cómo puedo usar el Explorador de costos para analizar mis gastos y mi consumo?

COST07-BP05 Análisis de modelos de precios en el nivel de la cuenta de administración

Consulte las herramientas de facturación y administración de costos y vea los descuentos recomendados con compromisos y reservas para hacer análisis periódicos en el nivel de la cuenta de administración.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

La creación periódica de modelos de costos lo ayuda a implementar oportunidades de optimización en múltiples cargas de trabajo. Por ejemplo, si varias cargas de trabajo usan instancias bajo demanda en un nivel agregado, el riesgo de cambio es menor e implementar un descuento basado en el compromiso puede tener un costo general inferior. Se recomienda hacer análisis en ciclos regulares de dos semanas a un mes. De este modo, podrá hacer compras de ajustes pequeños para que sus modelos de precios puedan seguir evolucionando a medida que cambien sus cargas de trabajo y sus componentes.

Utilice el AWS Cost Explorer para buscar descuentos por compromiso en su cuenta de administración. Las recomendaciones en el nivel de cuenta de administración se calculan teniendo en cuenta el uso en todas las cuentas de su organización de AWS que tengan activadas las instancias reservadas o Savings Plans (SP). También se calculan cuando se activa la opción de compartir descuentos para recomendar un compromiso que maximice los ahorros en todas las cuentas.

Aunque la compra en el nivel de cuenta de administración permite conseguir el máximo ahorro en muchos casos, puede haber situaciones en las que podría plantearse la posibilidad de comprar SP en el nivel de la cuenta vinculada cuando, por ejemplo, desee que los descuentos se apliquen primero al uso de esa cuenta vinculada en particular. Las recomendaciones para las cuentas de miembros se calculan en el nivel de cuenta individual para maximizar el ahorro de cada cuenta. Si su cuenta es propietaria de compromisos de instancias reservadas y Savings Plans, se aplicarán en este orden:

1. Instancia reservada de zona
2. Instancia reservada estándar
3. Instancia reservada convertible
4. Savings Plans para instancias
5. Savings Plans para computación

Si compra un plan de Savings Plans en el nivel de cuenta de administración, el ahorro se aplicará en función del porcentaje de descuento más alto al más bajo. Los Savings Plans de nivel de cuenta de administración examinan todas las cuentas vinculadas y aplican el ahorro allí donde el descuento sea más alto. Si desea restringir dónde se aplica el ahorro, puede comprar un plan de Savings Plans en el nivel de cuenta vinculada para que, cada vez que esa cuenta utilice servicios de computación que cumplan los requisitos, el descuento se aplique primero allí. Cuando la cuenta no utilice servicios

de computación que cumplan los requisitos, el descuento se compartirá entre las demás cuentas vinculadas de la misma cuenta de administración. La opción de compartir descuentos está activada de forma predeterminada, pero se puede desactivar si es necesario.

En una familia de facturación consolidada, los Savings Plans se aplican primero al uso de la cuenta del propietario y, después, al uso de otras cuentas. Esto ocurre solo si tiene activado el uso compartido. Sus Savings Plans se aplican primero a su porcentaje de ahorro más alto. Si hay varios usos con porcentajes de ahorro iguales, los Savings Plans se aplican al primer uso con la tarifa más baja de Savings Plans. Los Savings Plans seguirán aplicándose hasta que no queden más usos o hasta que se agote su compromiso. El uso restante se cobra según la tarifa bajo demanda. Puede actualizar las recomendaciones de Savings Plans (SP) en Administración de costos de AWS para generar nuevas recomendaciones de Savings Plans en cualquier momento.

Tras analizar la flexibilidad de las instancias, puede comprometerse siguiendo las recomendaciones. Cree modelos de costos al analizar los costos de la carga de trabajo a corto plazo con posibles opciones de recursos diferentes, además de analizar los modelos de precios de AWS y su alineación con los requisitos empresariales para averiguar el costo total de propiedad y las oportunidades de optimización de costos.

Pasos para la implementación

Análisis de descuento por compromiso: utilice el Explorador de costos en su cuenta y consulte las recomendaciones de Savings Plans e instancias reservadas. Asegúrese de que entiende las recomendaciones de Savings Plans y calcule el gasto y el ahorro mensual. Revise las recomendaciones en el nivel de cuenta de administración que se calculan teniendo en cuenta el uso en todas las cuentas de miembro de su organización de AWS que tengan activadas las instancias reservadas o el reparto de descuentos de Savings Plans para obtener el máximo ahorro en todas las cuentas. Puede verificar que ha implementado las recomendaciones correctas con los descuentos y riesgos necesarios si sigue los laboratorios de Well-Architected.

Recursos

Documentos relacionados:

- ¿Cómo funcionan los precios de AWS?
- Opciones de compra de instancias
- Saving Plan Overview
- Saving Plan recommendations

- Accessing Reserved Instance recommendations
- Understanding your Saving Plans recommendation
- How Savings Plans apply to your AWS usage
- Savings Plans con facturación unificada
- Turning on shared reserved instances and Savings Plans discounts

Videos relacionados:

- Save up to 90% and run production workloads on Spot

Ejemplos relacionados:

- Qué debo tener en cuenta antes de comprar un Savings Plan?
- How can I use rolling Savings Plans to reduce commitment risk?
- When to Use Spot Instances

Planificación para la transferencia de datos

Una ventaja de la nube es que es un servicio de red administrado. Ya no es necesario administrar y operar una flota de conmutadores, enrutadores y otros equipos de red asociados. Los recursos de red de la nube se consumen y se pagan del mismo modo que se paga por la CPU y el almacenamiento: solo se paga por lo que se utiliza. Se requiere un uso eficiente de los recursos de red para optimizar los costos en la nube.

Prácticas recomendadas

- COST08-BP01 Modelado de transferencia de datos
- COST08-BP02 Selección de componentes para optimizar el costo de la transferencia de datos
- COST08-BP03 Implementación de servicios para reducir los costos de transferencia de datos

COST08-BP01 Modelado de transferencia de datos

Reúna los requisitos de la organización y haga un modelado de transferencia de datos de la carga de trabajo y de cada uno de sus componentes. Se identifica el punto de costo más bajo para los requisitos de transferencia de datos actuales.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Cuando se diseña una solución en la nube, las tarifas de transferencia de datos suelen olvidarse. Esto se debe a la costumbre de diseñar la arquitectura utilizando centros de datos en las instalaciones o a la falta de conocimientos. Los cargos por transferencia de datos en AWS vienen determinados por el origen, el destino y el volumen del tráfico. Si se tienen en cuenta estas tarifas durante la fase de diseño, es posible ahorrar costos. Saber dónde se produce la transferencia de datos en su carga de trabajo, el costo de la transferencia y su beneficio asociado es muy importante para calcular con precisión el costo total de propiedad (TCO). De este modo, podrá tomar una decisión informada para modificar o aceptar la decisión arquitectónica. Por ejemplo, podría tener una configuración de la zona de disponibilidad múltiple donde replicar los datos entre las zonas de disponibilidad.

Después, modele los componentes de los servicios que transfieren los datos en su carga de trabajo y determine que se trata de un costo aceptable (similar al de pagar por la computación y el almacenamiento en la zona de disponibilidad) para lograr la fiabilidad y resiliencia requeridas. Debe modelar los costos según los distintos niveles de uso. El uso de la carga de trabajo puede cambiar con el tiempo y algunos servicios podrían ser más rentables en diferentes niveles.

Al modelar la transferencia de datos, piense en la cantidad de datos que se ingieren y de dónde provienen esos datos. Además, considere cuántos datos se procesan y cuánta capacidad de almacenamiento o computación se necesita. Durante el modelado, siga las prácticas recomendadas de redes para la arquitectura de la carga de trabajo con el fin de optimizar los costos potenciales de la transferencia de datos.

Calculadora de precios de AWS puede ayudarlo a conocer los costos estimados de servicios de AWS específicos y la transferencia de datos esperada. Si ya tiene una carga de trabajo en ejecución (con fines de prueba o en un entorno de preproducción), utilice el AWS Cost Explorer o el AWS Cost and Usage Report (CUR) para comprender y modelar sus costos de transferencia de datos. Configure una prueba de concepto (POC) o pruebe su carga de trabajo y ejecute una prueba con una carga simulada realista. Puede modelar sus costos con distintas demandas de carga de trabajo.

Pasos para la implementación

- Identificación de los requisitos: ¿Cuál es el objetivo principal y los requisitos empresariales para la transferencia de datos planificada entre el origen y el destino? ¿Cuál es el resultado empresarial que se espera al final? Recopile los requisitos empresariales y defina los resultados esperados.

- Identificación del origen y el destino: ¿Cuál es el origen y el destino de los datos para la transferencia de datos, por ejemplo, dentro de Regiones de AWS a los servicios de AWS o hacia Internet?
  - Data transfer within an Región de AWS
  - Data transfer between Regiones de AWS
  - Data transfer out to the internet
- Identificación de las clasificaciones de datos: ¿Cuál es la clasificación de datos para esta transferencia de datos? ¿Qué tipo de datos son? ¿Qué tamaño tienen los datos? ¿Con qué frecuencia se deben transferir los datos? ¿Los datos son confidenciales?
- Identificación de los servicios o herramientas de AWS que se van a utilizar: ¿Qué servicios de AWS se utilizan para esta transferencia de datos? ¿Es posible utilizar un servicio ya aprovisionado para otra carga de trabajo?
- Cálculo de los costos de transferencia de datos: utilice Precios de AWS, el modelo de transferencia de datos que creó anteriormente, para calcular los costos de transferencia de datos para la carga de trabajo. Calcule los costos de transferencia de datos en distintos niveles de uso para los incrementos y las reducciones del uso de la carga de trabajo. Si hay múltiples opciones para la arquitectura de la carga de trabajo, calcule el costo de cada opción para compararlas.
- Enlace de los costos con los resultados: especifique el resultado obtenido por la carga de trabajo para cada costo de transferencia de datos incurrido. Si es una transferencia entre componentes, puede deberse a un desacoplamiento y, si es entre zonas de disponibilidad, puede deberse a la redundancia.
- Creación de un modelado de transferencia de datos: después de recopilar toda la información, cree un modelo de transferencia de datos base conceptual para múltiples casos de uso y diferentes cargas de trabajo.

Recursos

Documentos relacionados:

- AWS caching solutions
- AWS Precios de
- Precios de Amazon EC
- Precios de Amazon VPC
- Understanding data transfer charges

Videos relacionados:

- Monitoring and Optimizing Your Data Transfer Costs
- S3 Transfer Acceleration

Ejemplos relacionados:

- Overview of Data Transfer Costs for Common Architectures
- AWS Prescriptive Guidance for Networking

COST08-BP02 Selección de componentes para optimizar el costo de la transferencia de datos

Se seleccionan todos los componentes y se diseña la arquitectura para reducir los costos de transferencia de datos. Incluye el uso de componentes como la optimización de la red de área extendida (WAN) y las configuraciones de varias zonas de disponibilidad (AZ).

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

La arquitectura para la transferencia de datos minimiza los costos de transferencia de datos. Para ello, es posible que deba usar redes de entrega de contenido para colocar los datos más cerca de los usuarios, o bien enlaces de red dedicados entre sus instalaciones y AWS. También puede usar la optimización de WAN y de las aplicaciones para reducir la cantidad de datos que se transfieren entre los componentes.

Al transferir datos a la Nube de AWS, o dentro de ella, es muy importante conocer el destino en función de los diversos casos de uso, la naturaleza de los datos y los recursos de red disponibles a fin de seleccionar los servicios de AWS correctos para optimizar la transferencia de datos. AWS ofrece una amplia gama de servicios de transferencia de datos diseñados para satisfacer diversos requisitos de migración de datos. Seleccione las opciones de almacenamiento de datos y transferencia de datos adecuadas en función de las necesidades empresariales de su organización.

Al planificar o revisar la arquitectura de la carga de trabajo, tenga en cuenta lo siguiente:

- Uso de los puntos de conexión de VPC en AWS: los puntos de conexión de VPC habilitan conexiones privadas entre la VPC y los servicios de AWS admitidos. Esto le permite evitar el uso de la red pública de Internet, que puede dar lugar a costos de transferencia de datos.
- Uso de una puerta de enlace de NAT: utilice una puerta de enlace de NAT para que las instancias de una subred privada puedan conectarse a Internet o a servicios fuera de la VPC. Compruebe si los recursos que hay detrás de la puerta de enlace de NAT que envían la mayor cantidad de tráfico se encuentran en la misma zona de disponibilidad que la puerta de enlace de NAT. Si no lo están, cree nuevas puertas de enlace de NAT en la misma zona de disponibilidad que el recurso para reducir los cargos por transferencia de datos entre AZ.
- Uso de AWS Direct Connect: Direct Connect evita la red pública de Internet y establece una conexión privada y directa entre la red en las instalaciones y AWS. Esto puede resultar más rentable y coherente que la transferencia de grandes volúmenes de datos a través de Internet.
- Imposibilidad de transferir datos a través de las fronteras regionales: las transferencias de datos entre Regiones de AWS (de una región a otra) suelen conllevar gastos. La vía multirregional debería ser una decisión muy meditada. Para obtener más información, consulte Multi-Region scenarios.
- Supervisión de la transferencia de datos: utilice los registros de flujo de Amazon CloudWatch y VPC para recopilar detalles sobre la transferencia de datos y el uso de la red. Analice la información de tráfico de red de sus VPC, como la dirección IP o el rango, que va y viene de las interfaces de red.
- Análisis del uso de la red: utilice herramientas de medición e informes, como el AWS Cost Explorer, CUDOS Dashboards o CloudWatch, para comprender el costo de transferencia de datos de su carga de trabajo.

Pasos para la implementación

- Selección de componentes para la transferencia de datos: use el modelo de transferencia de datos que se explica en COST08-BP01 Modelado de transferencia de datos para centrarse en dónde se encuentran los mayores costos de transferencia de datos o dónde estarían si cambia el uso de la carga de trabajo. Busque arquitecturas alternativas o componentes adicionales que eliminen o reduzcan la necesidad de transferir datos (o que reduzcan su costo).

Recursos

Prácticas recomendadas relacionadas:

- COST08-BP01 Modelado de transferencia de datos
- COST08-BP03 Implementación de servicios para reducir los costos de transferencia de datos

Documentos relacionados:

- Migración de datos a la nube
- AWS caching solutions
- Deliver content faster with Amazon CloudFront

Ejemplos relacionados:

- Overview of Data Transfer Costs for Common Architectures
- AWS Network Optimization Tips
- Optimize performance and reduce costs for network analytics with VPC Flow Logs in Apache Parquet format

COST08-BP03 Implementación de servicios para reducir los costos de transferencia de datos

Implemente servicios para reducir la transferencia de datos. Por ejemplo, utilice ubicaciones periféricas o redes de entrega de contenido (CDN) para ofrecer contenido a los usuarios finales, cree capas de almacenamiento en caché delante de sus servidores de aplicaciones o bases de datos y utilice conexiones de red dedicadas en lugar de VPN para conectarse a la nube.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

Existen varios servicios de AWS que pueden ayudarlo a optimizar el uso de la transferencia de datos de la red. Según los componentes de la carga de trabajo, el tipo y la arquitectura de la nube, estos servicios pueden ayudar a comprimir, almacenar en caché y compartir y distribuir su tráfico en la nube.

- Amazon CloudFront es una red global de entrega de contenido que brinda datos con baja latencia y altas velocidades de transferencia. Almacena en caché los datos en ubicaciones periféricas en todo el mundo, lo que reduce la carga de sus recursos. Mediante CloudFront, puede reducir

el esfuerzo administrativo que supone entregar contenido a una cantidad grande de usuarios a nivel mundial y hacerlo con una latencia mínima. El Paquete de promociones en seguridad puede ayudarlo a ahorrar hasta un 30 % en el uso de CloudFront si planea un aumento de este con el tiempo.

- AWS Direct Connect permite establecer una conexión de red dedicada con AWS. Esto puede reducir los costos de red, aumentar el ancho de banda y brindar una experiencia de red más coherente que las conexiones basadas en Internet.
- Site-to-Site VPN le permite establecer una conexión segura y privada entre su red privada y la red global de AWS. Es ideal para oficinas pequeñas o socios empresariales porque proporciona una conectividad simplificada, además de ser un servicio elástico y completamente administrado.
- Los puntos de conexión de VPC permiten la conexión entre los servicios de AWS a través de la red privada y se pueden usar para reducir los costos de la transferencia de datos pública y los costos de la puerta de enlace de NAT. Los puntos de conexión de VPC de puerta de enlace no tienen cargos por hora y son compatibles con Amazon S3 y Amazon DynamoDB. Los puntos de conexión de VPC de interfaz están proporcionados por AWS PrivateLink y tienen una tarifa por hora y un costo de uso por GB.
- Las puertas de enlace de NAT ofrecen escalado y administración integrados para reducir los costos, a diferencia de una instancia de NAT independiente. Coloque las puertas de enlace de NAT en las mismas zonas de disponibilidad que las instancias de alto tráfico y plantéese usar puntos de conexión de VPC para las instancias que necesiten acceder a Amazon DynamoDB o Amazon S3 a fin de reducir los costos de transferencia y procesamiento de datos.
- Utilice dispositivos de AWS Snow Family con recursos informáticos para recopilar y procesar datos en la periferia. Los dispositivos de AWS Snow Family (Snowball Edge, Snowball Edge y Snowmobile) le permiten mover petabytes de datos a la Nube de AWS, una opción rentable y sin conexión a Internet.

Pasos para la implementación

- Implementación de los servicios: seleccione los servicios de red de AWS aplicables en función del tipo de carga de trabajo del servicio mediante el modelado de transferencia de datos y la revisión de registros de flujo de VPC. Observe dónde están los mayores costos y los mayores flujos de volumen. Revise los servicios de AWS y evalúe si existe un servicio que reduzca o elimine la transferencia, especialmente en relación con la entrega de contenido y las redes. Busque también servicios de almacenamiento en caché donde haya acceso repetido a los datos o grandes cantidades de datos.

Recursos

Documentos relacionados:

- AWS Direct Connect
- Explore nuestros productos de AWS
- AWS caching solutions
- Amazon CloudFront
- AWS Snow Family
- Paquete de promociones en seguridad de Amazon CloudFront

Videos relacionados:

- Monitoring and Optimizing Your Data Transfer Costs
- AWS Cost Optimization Series: CloudFront
- How can I reduce data transfer charges for my NAT gateway?

Ejemplos relacionados:

- How-to chargeback shared services: An AWS Transit Gateway example
- Understand AWS data transfer details in depth from cost and usage report using Athena query and QuickSight
- Overview of Data Transfer Costs for Common Architectures
- Using AWS Cost Explorer to analyze data transfer costs
- Cost-Optimizing your AWS architectures by utilizing Amazon CloudFront features
- How can I reduce data transfer charges for my NAT gateway?

Administración de la demanda y suministro de recursos

Al migrar a la nube, paga solo por lo que necesita. Puede proporcionar recursos para que se adapten a la demanda de carga de trabajo en el momento que se requieran, eliminando así la necesidad de sobreaprovisionamiento, que es algo caro e inútil. También puede modificar la demanda con un límite, un búfer o una cola para suavizar la demanda y usar menos recursos.

Los beneficios económicos del suministro justo a tiempo deben equilibrarse con respecto a la necesidad de aprovisionamiento a fin de tener en cuenta los errores de recursos, la alta disponibilidad y el tiempo de aprovisionamiento. En función de si su demanda es fija o variable, planifique la creación de métricas y automatización que garanticen que la administración de su entorno sea mínima, incluso cuando escale. Al modificar la demanda, debe conocer el retraso aceptable y máximo que puede permitir la carga de trabajo.

En AWS, puede usar distintos enfoques para administrar la demanda y suministrar recursos. En las siguientes prácticas recomendadas se describe cómo se usan estos enfoques.

Prácticas recomendadas

- COST09-BP01 Análisis de la demanda de la carga de trabajo
- COST09-BP02 Implementación de un búfer o una limitación para administrar la demanda
- COST09-BP03 Suministro dinámico de recursos

COST09-BP01 Análisis de la demanda de la carga de trabajo

Analice la demanda de la carga de trabajo a lo largo del tiempo. Compruebe que el análisis cubra las tendencias estacionales y represente con precisión las condiciones de servicio durante toda la vida útil de la carga de trabajo. El análisis debe reflejar los posibles beneficios; por ejemplo, el tiempo empleado es proporcional al costo de la carga de trabajo.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

El análisis de la demanda de la carga de trabajo para la computación en la nube implica comprender los patrones y las características de las tareas de computación que se inician en el entorno de la nube. Este análisis ayuda a los usuarios a optimizar la asignación de recursos, administrar los costos y verificar que el rendimiento cumpla con los niveles requeridos.

Debe conocer los requisitos de la carga de trabajo. Los requisitos de su organización deben indicar los tiempos de respuesta de la carga de trabajo frente a las solicitudes. El tiempo de respuesta se puede usar para determinar si la demanda está administrada o si el suministro de recursos debe cambiar para adaptarse a la demanda.

El análisis debe incluir la previsibilidad y la repetibilidad de la demanda, el ritmo y la cantidad de cambio en la demanda. Haga el análisis a lo largo de un periodo suficiente de tiempo para que incorpore variantes estacionales, como un procesamiento de final de mes o los picos de las vacaciones.

La actividad de análisis debe reflejar los posibles beneficios de la implementación del escalado. Consulte el costo total previsto del componente y cualquier incremento o descenso del uso, así como el costo durante el periodo de vida de la carga de trabajo.

Estos son algunos aspectos clave que se deben tener en cuenta al hacer un análisis de la demanda de la carga de trabajo para la computación en la nube:

1. Métricas de uso y rendimiento de los recursos: analice cómo se utilizan los recursos de AWS a lo largo del tiempo. Determine los patrones de uso en las horas punta y fuera de las horas punta para optimizar la asignación de recursos y las estrategias de escalado. Supervise las métricas de rendimiento, como los tiempos de respuesta, la latencia, el rendimiento y las tasas de error. Estas métricas ayudan a evaluar el estado general y la eficiencia de la infraestructura de la nube.
2. Comportamiento del escalado de usuarios y aplicaciones: comprenda el comportamiento de los usuarios y cómo afecta a la demanda de carga de trabajo. El análisis de los patrones del tráfico de usuarios ayuda a mejorar la entrega de contenido y la capacidad de respuesta de las aplicaciones. Analice cómo se escalan las cargas de trabajo a medida que aumenta la demanda. Determine si los parámetros de escalado automático están configurados de forma correcta y eficaz para gestionar las fluctuaciones de carga.
3. Tipos de cargas de trabajo: identifique los diferentes tipos de cargas de trabajo que se ejecutan en la nube, como el procesamiento por lotes, el procesamiento de datos en tiempo real, las aplicaciones web, las bases de datos o el machine learning. Cada tipo de carga de trabajo puede tener requisitos de recursos y perfiles de rendimiento diferentes.
4. Acuerdos de nivel de servicio (SLA): compare el rendimiento real con los SLA para garantizar el cumplimiento e identificar las áreas que necesitan mejoras.

Puede utilizar Amazon CloudWatch para recopilar métricas y hacer un seguimiento de ellas, supervisar archivos de registros, establecer alarmas y reaccionar automáticamente a los cambios

en sus recursos AWS. También puede utilizar Amazon CloudWatch para obtener visibilidad de todo el sistema sobre la utilización de recursos, el rendimiento de las aplicaciones y el estado de funcionamiento.

Con AWS Trusted Advisor, puede aprovisionar sus recursos conforme a las prácticas recomendadas para mejorar el rendimiento y la fiabilidad del sistema, aumentar la seguridad y buscar oportunidades para ahorrar dinero. También puede desactivar las instancias que no son de producción y utilizar Amazon CloudWatch y Auto Scaling para adaptarlas a los aumentos o reducciones de la demanda.

Por último, puede utilizar AWS Cost Explorer o Quick con el archivo de AWS Cost and Usage Report (CUR) o los registros de la aplicación para hacer un análisis avanzado de la demanda de la carga de trabajo.

En general, un análisis integral de la demanda de la carga de trabajo permite a las organizaciones tomar decisiones informadas sobre el aprovisionamiento, el escalado y la optimización de los recursos, lo que se traduce en un mejor rendimiento, rentabilidad y satisfacción de los usuarios.

Pasos para la implementación

- Análisis de los datos de la carga de trabajo existente: analice los datos de la carga de trabajo existente, las versiones anteriores de la carga de trabajo o los patrones de uso previstos. Utilice Amazon CloudWatch, los archivos de registro y los datos de supervisión para obtener información sobre cómo se utilizó la carga de trabajo. Analice un ciclo completo de la carga de trabajo y recopile datos de los cambios estacionales, como los eventos de final de mes o de final de año. El esfuerzo reflejado en este análisis debe mostrar las características de la carga de trabajo. Debe ponerse mayor empeño en las cargas de trabajo de mayor valor con mayores cambios en la demanda. Debe ponerse menor empeño en las cargas de trabajo de menor valor con menores cambios en la demanda.
- Previsión de los factores externos: reúnase con miembros de equipos de toda la organización que puedan influir en la demanda de la carga de trabajo o cambiarla. Estos equipos suelen ser los de Ventas, Marketing o Desarrollo empresarial. Colabore con ellos para conocer los ciclos en los que operan y si hay eventos especiales que puedan cambiar la demanda de la carga de trabajo. Haga una previsión de la demanda de la carga de trabajo con estos datos.

Recursos

Documentos relacionados:

- Amazon CloudWatch
- AWS Trusted Advisor
- AWS X-Ray
- AWS Auto Scaling
- AWS Instance Scheduler
- Getting started with Amazon SQS
- AWS Cost Explorer
- Quick

Ejemplos relacionados:

- Monitor, Track and Analyze for cost optimization
- Searching and analyzing logs in CloudWatch

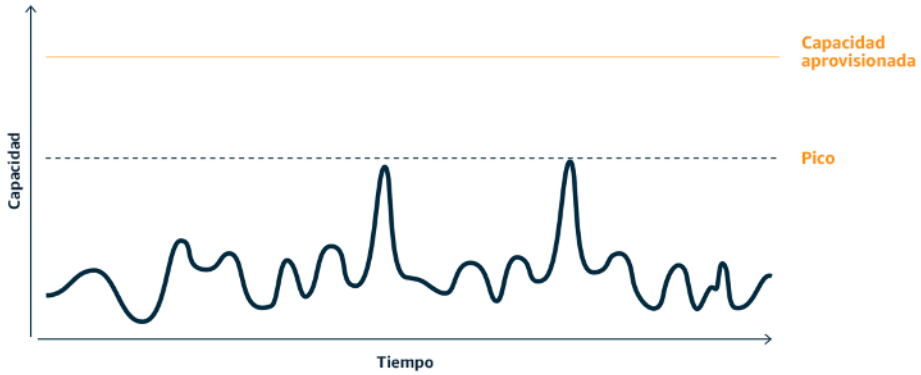
COST09-BP02 Implementación de un búfer o una limitación para administrar la demanda

El almacenamiento en búfer y la limitación modifican la demanda de la carga de trabajo y suavizan los picos. Implemente limitaciones cuando sus clientes lleven a cabo reintentos. Implemente el almacenamiento en búfer para almacenar la solicitud y aplazar el procesamiento para más adelante. Verifique que las limitaciones y los búferes se hayan diseñado de tal manera que los clientes reciban una respuesta en el tiempo requerido.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

La implementación de un búfer o limitación es crucial en la computación en la nube para administrar la demanda y reducir la capacidad aprovisionada necesaria para la carga de trabajo. Para conseguir un rendimiento óptimo, es esencial evaluar la demanda total, incluidos los picos, el ritmo de cambio en las solicitudes y el tiempo de respuesta necesario. Cuando los clientes tienen la posibilidad de reenviar sus solicitudes, es práctico aplicar la limitación. Por el contrario, para los clientes que carecen de funcionalidades de reintento, lo ideal es implementar una solución de búfer. Estos búferes agilizan la afluencia de solicitudes y optimizan la interacción de las aplicaciones con diferentes velocidades operativas.



Curva de demanda con dos picos distintos que requieren una elevada capacidad aprovisionada

Supongamos que tenemos una carga de trabajo con la curva de demanda que se muestra en la imagen anterior. Esta carga de trabajo tiene dos picos y, para gestionarlos, se aprovisiona la capacidad de recursos que muestra la línea naranja. Los recursos y la energía utilizados para esta carga de trabajo no están indicados en el área situada debajo de la curva de demanda, sino en el área situada debajo de la línea de capacidad aprovisionada, ya que esta capacidad es la que se necesita para gestionar esos dos picos. El aplanamiento de la curva de demanda de la carga de trabajo puede ayudarle a reducir la capacidad aprovisionada para una carga de trabajo y a reducir su impacto medioambiental. Para suavizar el pico, considere la posibilidad de implementar una solución de limitación o almacenamiento en búfer.

Vamos a profundizar en las limitaciones y el almacenamiento en búfer para entenderlos mejor.

Limitación: si la fuente de la demanda tiene capacidad de reintento, puede implementar la limitación. La limitación le dice al origen que, si no puede atender la solicitud en ese momento, debe intentarlo más tarde. El origen espera un tiempo y vuelve a intentar la solicitud. Implementar una limitación tiene la ventaja de que se limita la cantidad máxima de recursos y costos de la carga de trabajo. En AWS, puede utilizar Amazon API Gateway para implementar la limitación.

Basado en búfer: un enfoque basado en búfer utiliza productores (componentes que envían mensajes a la cola), consumidores (componentes que reciben los mensajes de la cola) y una cola (que contiene los mensajes) para almacenar los mensajes. De este modo, los consumidores pueden leer y procesar los mensajes, lo que permite que dichos mensajes se ejecuten a la velocidad que cumpla con los requisitos empresariales de los consumidores. Al utilizar una metodología centrada en los búferes, los mensajes de los productores se alojan en colas o secuencias, listos para que los consumidores accedan a ellos a un ritmo que se ajuste a sus demandas operativas.

En AWS, puede elegir entre varios servicios para implementar un enfoque basado en almacenamiento en búfer. Amazon Simple Queue Service (Amazon SQS) es un servicio administrado que incorpora colas que permiten a un solo consumidor leer mensajes individuales. Amazon Kinesis ofrece una secuencia que permite que muchos consumidores lean los mismos mensajes.

El almacenamiento en búfer y las limitaciones pueden suavizar cualquier pico al modificar la demanda de la carga de trabajo. Utilice limitaciones cuando los clientes vuelvan a intentar efectuar acciones y utilice el almacenamiento en búfer para retener la solicitud y procesarla más adelante. Al trabajar con un enfoque basado en búfer, diseñe su carga de trabajo para atender la solicitud en el tiempo requerido y verifique que pueda gestionar las solicitudes duplicadas. Analice la demanda general, la tasa de cambio y el tiempo de respuesta requerido para dimensionar correctamente la limitación o el búfer requeridos.

Pasos para la implementación

- Analice los requisitos del cliente: analice las solicitudes del cliente para determinar si puede llevar a cabo reintentos. Para los clientes que no puedan llevarlos a cabo, deberán implementarse búferes. Analice la demanda general, el ritmo de cambio y el tiempo de respuesta requerido para determinar el tamaño de la limitación o del búfer requeridos.
- Implemente un búfer o una limitación: implemente un búfer o una limitación en la carga de trabajo. Una cola, como Amazon Simple Queue Service (Amazon SQS), puede proporcionar un búfer a los componentes de la carga de trabajo. Amazon API Gateway puede proporcionar limitación a los componentes de la carga de trabajo.

Recursos

Prácticas recomendadas relacionadas:

- SUS02-BP06 Implementación del almacenamiento en búfer o la limitación para aplanar la curva de demanda
- REL05-BP02 Limitación de las solicitudes

Documentos relacionados:

- AWS Auto Scaling
- AWS Instance Scheduler

- Amazon API Gateway
- Amazon Simple Queue Service
- Getting started with Amazon SQS
- Amazon Kinesis

Videos relacionados:

- Choosing the Right Messaging Service for Your Distributed App

Ejemplos relacionados:

- Managing and monitoring API throttling in your workloads
- Throttling a tiered, multi-tenant REST API at scale using API Gateway
- Enabling Tiering and Throttling in a Multi-Tenant Amazon EKS SaaS Solution Using Amazon API Gateway
- Application integration Using Queues and Messages

COST09-BP03 Suministro dinámico de recursos

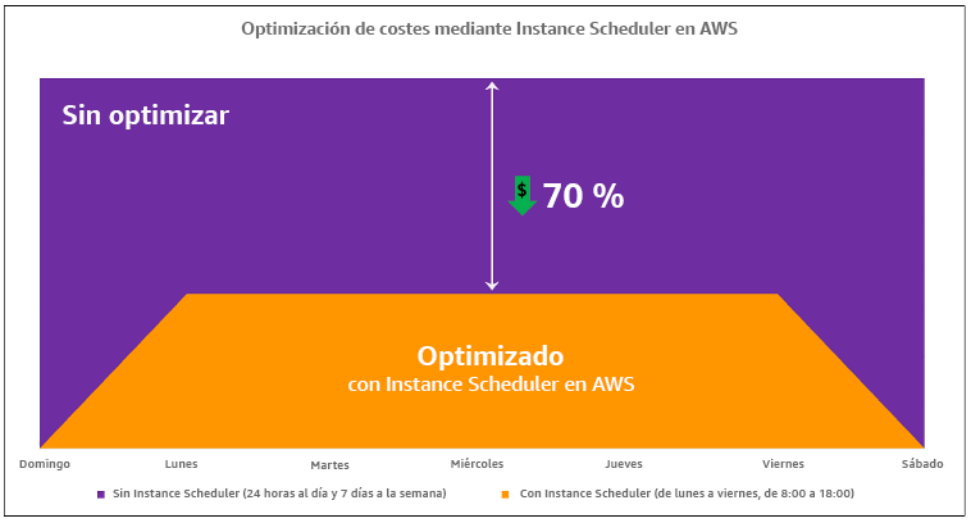
Los recursos se aprovisionan de manera planificada. Esto puede basarse en la demanda (por ejemplo, mediante el escalado automático) o en el tiempo, donde la demanda es predecible y los recursos se proporcionan en función del tiempo. Estos métodos conllevan la menor cantidad de aprovisionamiento excesivo o insuficiente.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

Hay varias formas en que los clientes de AWS pueden aumentar los recursos disponibles para sus aplicaciones y suministrar recursos para satisfacer la demanda. Una de estas opciones consiste en utilizar AWS Instance Scheduler, que automatiza el inicio y la detención de instancias de Amazon Elastic Compute Cloud (Amazon EC2) y de Amazon Relational Database Service (Amazon RDS). La otra opción es usar AWS Auto Scaling, que le permite escalar automáticamente los recursos de computación en función de la demanda de la aplicación o servicio. Suministrar recursos en función de la demanda le permitirá pagar únicamente por los recursos que utilice y reducir los costos, ya que solo lanza los recursos cuando se necesitan y los cancela cuando no.

AWS Instance Scheduler le permite configurar la detención y el inicio de sus instancias de Amazon EC2 y Amazon RDS en momentos definidos para que pueda satisfacer la demanda de los mismos recursos según un patrón temporal coherente; por ejemplo, que todos los días los usuarios accedan a las ocho de la mañana a instancias de Amazon EC2 que no se necesitan después de las seis de la tarde. Esta solución contribuye a reducir los costos operativos, ya que se detienen los recursos que no están en uso y se ponen en marcha cuando se necesitan.



Optimización de costos con AWS Instance Scheduler.

También puede configurar fácilmente los horarios de sus instancias de Amazon EC2 en todas sus cuentas y regiones con una interfaz de usuario (IU) sencilla mediante la configuración rápida de AWS Systems Manager. Puede programar instancias de Amazon EC2 o Amazon RDS con AWS Instance Scheduler y detener e iniciar las instancias existentes. Sin embargo, no puede detener ni iniciar instancias que formen parte de su grupo de escalado automático (ASG) ni que administren servicios como Amazon Redshift o Amazon OpenSearch Service. Los grupos de escalado automático tienen su propia programación para las instancias del grupo y estas instancias se crean.

AWS Auto Scaling lo ayuda a ajustar la capacidad para mantener un rendimiento predecible y estable al menor costo posible para satisfacer los cambios en la demanda. Se trata de un servicio gratuito y totalmente administrado para escalar la capacidad de su aplicación que se integra con las instancias de Amazon EC2 y flotas de spot, Amazon ECS, Amazon DynamoDB y Amazon Aurora. El escalado automático detecta los recursos automáticamente, lo que lo ayuda a buscar recursos en su carga de trabajo que se pueden configurar. Además, tiene estrategias de escalado integradas para optimizar el rendimiento y los costos, o un equilibrio entre ambos, y proporciona escalado predictivo para ayudar en los picos que se producen periódicamente.

Hay varias opciones de escalado disponibles para escalar su grupo de escalado automático:

- Mantener los niveles de instancia actuales en todo momento
- Escalar manualmente
- Escalado según una programación
- Escalado basado en la demanda
- Usar el escalado predictivo

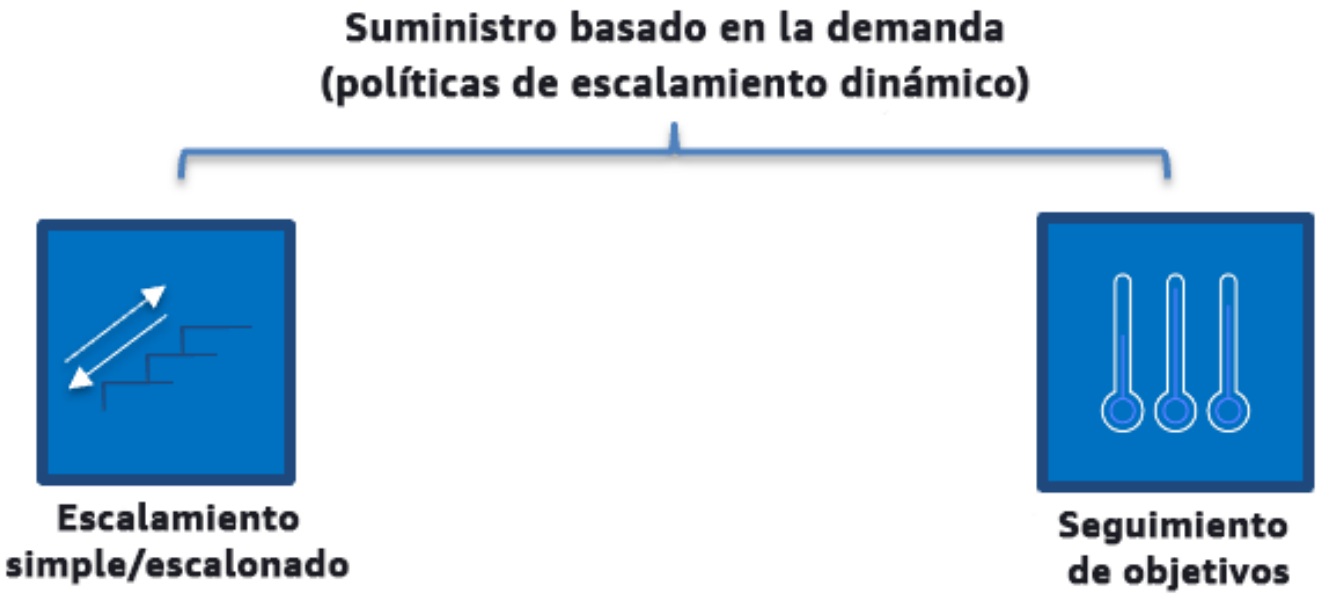
Existen diferentes políticas de escalado automático. Se pueden clasificar en políticas de escalado dinámicas y programadas. Las políticas dinámicas son escalados manuales o dinámicos, que pueden ser programados o predictivos. Puede utilizar políticas de escalado para un escalado dinámico, programado y predictivo. También puede usar métricas y alarmas de Amazon CloudWatch para activar eventos de escalado de su carga de trabajo. Le recomendamos que utilice plantillas de lanzamiento que le permiten acceder a las últimas características y mejoras. No todas las funciones de escalado automático están disponibles cuando se utilizan configuraciones de lanzamiento. Por ejemplo, no puede crear un grupo de Auto Scaling que lance instancias de spot y bajo demanda o que especifique varios tipos de instancia. Debe utilizar una plantilla de lanzamiento para configurar estas características. Cuando utilice plantillas de lanzamiento, le recomendamos que haga un control de versiones en cada una de ellas. Con el control de versiones de plantillas de lanzamiento, puede crear un subconjunto del conjunto completo de parámetros. A continuación, puede reutilizarlo para crear otras versiones de la misma plantilla de lanzamiento.

Puede usar AWS Auto Scaling o incorporar el escalado en su código con las API o SDK de AWS. Esto reduce los costos generales de la carga de trabajo al eliminar el costo operativo de aplicar los cambios manualmente en su entorno. Además, los cambios se pueden aplicar mucho más rápido. De este modo, también se adapta la dotación de recursos de la carga de trabajo a su demanda en cualquier momento. Para seguir esta práctica recomendada y suministrar recursos de forma dinámica a su organización, debe comprender el escalado horizontal y vertical en la Nube de AWS, así como la naturaleza de las aplicaciones que se ejecutan en las instancias de Amazon EC2. Es mejor que su equipo de administración financiera en la nube colabore con los equipos técnicos para seguir esta práctica recomendada.

Elastic Load Balancing (Elastic Load Balancing) lo ayuda a escalar mediante la distribución de la demanda entre varios recursos. Mediante ASG y Elastic Load Balancing, puede administrar las solicitudes entrantes mediante el enrutamiento óptimo del tráfico para que ninguna instancia se sobrecargue en un grupo de escalado automático. Las solicitudes se distribuirían entre todos los destinatarios de un grupo objetivo por turnos, sin tener en cuenta la capacidad ni la utilización.

Las métricas habituales pueden ser métricas de Amazon EC2 estándar, como el uso de la CPU, el rendimiento de la red y la latencia de solicitud y respuesta observada de Elastic Load Balancing. Si es posible, debe usar una métrica indicativa de la experiencia del cliente. Suele ser una métrica personalizada que se puede originar en el código de la aplicación en la carga de trabajo. Para explicar cómo satisfacer la demanda de forma dinámica, vamos a agrupar el escalado automático en dos categorías (modelos de suministro basados en la demanda y modelos de suministro basados en el tiempo) y analizaremos en profundidad cada una de ellas.

Suministro basado en la demanda: aproveche la elasticidad de la nube para suministrar recursos que satisfagan los cambios en la demanda utilizando el estado de la demanda casi en tiempo real. Para el suministro basado en la demanda, utilice las API o las características del servicio para cambiar mediante programación la cantidad de recursos en la nube de su arquitectura. De esta forma, puede escalar componentes en su arquitectura y aumentar la cantidad de recursos durante los picos de demanda para mantener el rendimiento, así como disminuir la capacidad cuando la demanda disminuya para reducir los costos.

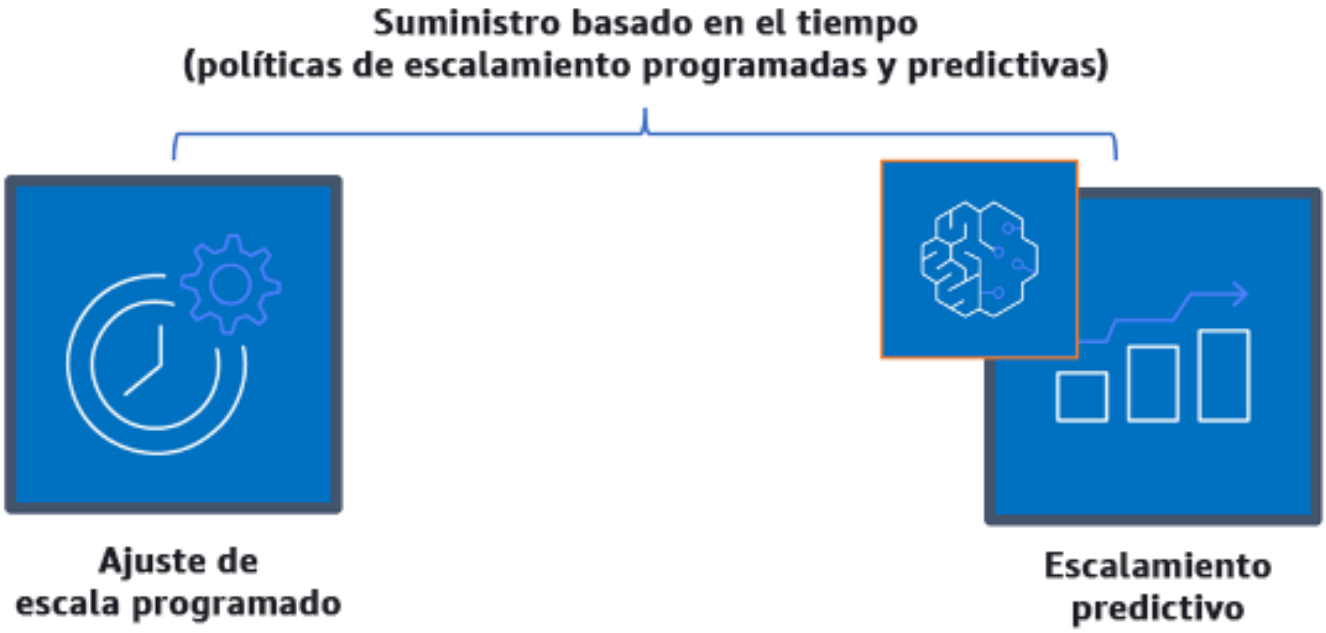


Políticas de escalado dinámico basadas en la demanda

- Escalado sencillo o por pasos: supervisa las métricas y agrega o elimina instancias de acuerdo con los pasos definidos manualmente por los clientes.
- Seguimiento de objetivos: mecanismo de control similar a un termostato que agrega o elimina instancias automáticamente para mantener las métricas en un objetivo definido por el cliente.

Al diseñar con un enfoque basado en la demanda, tenga en cuenta dos consideraciones clave. La primera: debe conocer la rapidez con la que necesita aprovisionar recursos nuevos. La segunda: tenga en cuenta que el tamaño del margen entre la oferta y la demanda cambiará. Debe estar preparado para poder hacer frente a la velocidad del cambio en la demanda y también a los errores de recursos.

Suministro basado en el tiempo: el enfoque basado en el tiempo adapta la capacidad de los recursos a una demanda que es predecible o que está bien definida por el tiempo. Normalmente, este enfoque no depende de los niveles de utilización de los recursos. El enfoque basado en el tiempo garantiza que los recursos estén disponibles en el momento específico en que se necesiten y que se puedan proporcionar sin retrasos debidos a los procedimientos de inicio y comprobaciones del sistema o de coherencia. Con el enfoque basado en el tiempo, puede brindar recursos adicionales o aumentar la capacidad durante los periodos de mayor actividad.



Políticas de escalado basadas en el tiempo

Puede utilizar el escalado automático programado o predictivo para implementar un enfoque basado en el tiempo. Las cargas de trabajo se pueden programar para escalarse o reducirse horizontalmente en momentos definidos (como el inicio del horario laboral). De este modo, los recursos están disponibles cuando lleguen los usuarios o aumente la demanda. El escalado predictivo utiliza patrones para escalar horizontalmente, mientras que el escalado programado utiliza tiempos

predefinidos para escalar horizontalmente. También puede utilizar una estrategia de selección de tipo de instancia basada en atributos (ABS) en los grupos de escalado automático (ASG), lo que le permite expresar sus requisitos de instancia como un conjunto de atributos, por ejemplo, vCPU, memoria y almacenamiento. De este modo, también puede utilizar automáticamente los tipos de instancia de nueva generación cuando se lancen y acceder a una gama más amplia de capacidad con las instancias de spot de Amazon EC2. Flota de Amazon EC2 y Amazon EC2 Auto Scaling seleccionan y lanzan instancias que se ajusten a los atributos especificados, por lo que no es necesario elegir manualmente los tipos de instancia.

También puede utilizar las API y los SDK de AWS y AWS CloudFormation para aprovisionar y retirar entornos completos de manera automática según los necesite. Este enfoque es ideal para los entornos de desarrollo o pruebas que se ejecutan únicamente en horarios laborales o periodos definidos. Puede usar las API para escalar el tamaño de los recursos dentro de un entorno (escalado vertical). Por ejemplo, puede escalar verticalmente una carga de trabajo de producción mediante el cambio del tamaño o la clase de instancia. Para ello, hay que detener o iniciar la instancia y seleccionar el tamaño o la clase de instancia diferente. Esta técnica también se puede aplicar a otros recursos tales como los volúmenes elásticos de Amazon EBS, los cuales se pueden modificar para aumentar el tamaño, ajustar el rendimiento (IOPS) o cambiar el tipo de volumen mientras están en uso.

Al diseñar con un enfoque basado en el tiempo, tenga en cuenta dos consideraciones clave. La primera: ¿qué grado de consistencia presenta el patrón? La segunda: ¿en qué afectaría el patrón si cambiara? Puede aumentar la precisión de las predicciones mediante la supervisión de sus cargas de trabajo y el uso de la inteligencia empresarial. Si observa cambios considerables en el patrón de uso, puede ajustar los tiempos para asegurarse de que se proporcione cobertura.

Pasos para la implementación

- Configuración del escalado programado: en caso de cambios predecibles en la demanda, el escalado basado en el tiempo puede proporcionar el número correcto de recursos de manera oportuna. También es útil si la creación y configuración de recursos no son suficientemente rápidas a la hora de responder a los cambios en la demanda. Use el análisis de las cargas de trabajo para configurar el escalado programado con AWS Auto Scaling. Para configurar la programación en función del tiempo, puede utilizar el escalado predictivo del escalado programado para aumentar por adelantado el número de instancias de Amazon EC2 de sus grupos de escalado automático en función de los cambios de carga previstos o predecibles.
- Configuración del escalado predictivo: el escalado predictivo le permite aumentar el número de instancias de Amazon EC2 en su grupo de escalado automático antes de los patrones diarios

y semanales de flujos de tráfico. Si tiene picos de tráfico regulares y aplicaciones que tardan mucho en iniciarse, debería plantearse el uso del escalado predictivo. El escalado predictivo puede ayudarlo a escalar más rápidamente mediante la inicialización de la capacidad antes de la carga prevista si se compara con el escalado dinámico únicamente, que es de naturaleza reactiva. Por ejemplo, si los usuarios empiezan a utilizar su carga de trabajo con el inicio del horario laboral y no la utilizan fuera de dicho horario, el escalado predictivo puede agregar capacidad antes del horario laboral, lo que elimina el retraso del escalado dinámico para reaccionar ante los cambios en el tráfico.

- Configuración del escalado automático dinámico: use el escalado automático para configurar el escalado según las métricas de carga de trabajo activas. Use los análisis y configure el escalado automático para que se lance en los niveles de recursos correctos y verifique que la carga de trabajo se escala en el tiempo requerido. Puede lanzar y escalar automáticamente una flota de instancias en diferido e instancias de spot en un solo grupo de Auto Scaling. Además de los descuentos relacionados con las instancias de spot, puede utilizar instancias reservadas o un Savings Plan para conseguir mejores precios de los habituales en las instancias en diferido. La combinación de todos estos factores le permite optimizar el ahorro de costos en las instancias de Amazon EC2, a la vez que se asegura de obtener la escala y el rendimiento deseados para su aplicación.

Recursos

Documentos relacionados:

- AWS Auto Scaling
- AWS Instance Scheduler
- Escalar el tamaño de un grupo de escalado automático
- Getting Started with Amazon EC2 Auto Scaling
- Getting started with Amazon SQS
- Scheduled Scaling for Amazon EC2 Auto Scaling
- Predictive scaling for Amazon EC2 Auto Scaling

Videos relacionados:

- Target Tracking Scaling Policies for Auto Scaling
- AWS Instance Scheduler

Ejemplos relacionados:

- Attribute based Instance Type Selection for Auto Scaling for Amazon EC2 Fleet
- Optimizing Amazon Elastic Container Service for cost using scheduled scaling
- Predictive Scaling with Amazon EC2 Auto Scaling
- ¿Cómo utilizo Instance Scheduler con CloudFormation para programar instancias de Amazon EC2?

Optimización a lo largo del tiempo

En AWS, puede optimizar a lo largo del tiempo al revisar nuevos servicios e implementarlos en su carga de trabajo.

A medida que AWS lanza nuevos servicios y características, recomendamos que revise sus decisiones arquitectónicas para asegurarse de que continúen siendo rentables. A medida que cambian sus requisitos, retire de forma agresiva recursos, componentes y cargas de trabajo que ya no necesite. Tenga en cuenta las siguientes prácticas recomendadas para ayudarlo en la optimización a lo largo del tiempo.

Mientras optimiza sus cargas de trabajo a lo largo del tiempo y mejora su cultura de CFM en la organización, evalúe el costo del esfuerzo de las operaciones en la nube, revise las operaciones en la nube que consumen mucho tiempo e implemente automatizaciones para reducir los esfuerzos humanos y los costos mediante la adopción de servicios de AWS relacionados, productos de terceros o herramientas personalizadas (como la AWS CLI o los SDK de AWS).

Temas

- Definición de un proceso de revisión y análisis periódico de la carga de trabajo
- Automatización de operaciones

Definición de un proceso de revisión y análisis periódico de la carga de trabajo

Prácticas recomendadas

- COST10-BP01 Desarrollo de un proceso de revisión de la carga de trabajo
- COST10-BP02 Revisión y análisis regulares de esta carga de trabajo

COST10-BP01 Desarrollo de un proceso de revisión de la carga de trabajo

Desarrolle un proceso que defina los criterios y el proceso para la revisión de las cargas de trabajo. El esfuerzo de revisión debe reflejar la ventaja potencial. Por ejemplo, las cargas de trabajo principales o las cargas de trabajo con un valor por encima del diez por ciento de la factura se revisan trimestral o semestralmente, mientras que las cargas de trabajo por debajo del diez por ciento se revisan anualmente.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: alto

Guía para la implementación

Para tener siempre la carga de trabajo más rentable, debe revisar periódicamente la carga de trabajo para saber si hay oportunidades de implementar nuevos servicios, características y componentes. Para conseguir reducir los costos totales, el proceso debe ser proporcional al volumen potencial de ahorro. Por ejemplo, las cargas de trabajo que suponen el 50 % de sus gastos totales se deben revisar con mayor regularidad y más a fondo que las cargas de trabajo que constituyen el cinco por ciento de sus gastos totales. Tenga en cuenta los factores externos o la volatilidad. Si la carga de trabajo da servicio a un segmento geográfico o de mercado específico y se prevén cambios en ese ámbito, unas revisiones más frecuentes podrían suponer un ahorro de costos. Otro factor de revisión es el esfuerzo para implementar los cambios. Si las pruebas y la validación de los cambios suponen un costo importante, las revisiones deberían ser menos frecuentes.

Hay que tener en cuenta el costo a largo plazo del mantenimiento de componentes y recursos obsoletos y heredados, y la imposibilidad de implementar nuevas características en ellos. El costo actual de las pruebas y la validación puede superar el beneficio propuesto. Sin embargo, con el tiempo, el costo de aplicar el cambio puede aumentar significativamente a medida que se incrementa la brecha entre la carga de trabajo y las tecnologías actuales, lo que se traduce en costo aún mayores. Por ejemplo, el costo de pasar a un nuevo lenguaje de programación puede no ser rentable en la actualidad. No obstante, dentro de cinco años, el costo de las personas con competencias en ese lenguaje puede aumentar y, debido al crecimiento de la carga de trabajo, estaría trasladando un sistema aún mayor al nuevo lenguaje, lo que requeriría un esfuerzo aún mayor que el anterior.

Divida la carga de trabajo en componentes, asigne el costo del componente (basta con una estimación) y, a continuación, enumere los factores (por ejemplo, el esfuerzo y los mercados externos) junto a cada componente. Utilice estos indicadores para determinar la frecuencia de revisión de cada carga de trabajo. Por ejemplo, es posible que los servidores web tengan un costo elevado, un esfuerzo de cambio bajo y unos factores externos elevados, lo que da lugar a una frecuencia de revisión alta. Una base de datos central puede tener un costo medio, un esfuerzo de cambio alto y unos factores externos bajos, lo que da lugar a una frecuencia de revisión media.

Defina un proceso para evaluar nuevos servicios, patrones de diseño, tipos de recursos y configuraciones para optimizar el costo de la carga de trabajo a medida que estén disponibles. Al igual que en los procesos de revisión del pilar de rendimiento y revisión del pilar de fiabilidad, identifique, valide y priorice las actividades de optimización y mejora y la solución de problemas e incorpórelas a su cartera de trabajo.

Pasos para la implementación

- Defina la frecuencia de revisión: defina con qué frecuencia se deben revisar la carga de trabajo y sus componentes. Asigne tiempo y recursos a la mejora continua y revise la frecuencia para mejorar la eficacia y la optimización de su carga de trabajo. Se trata de una combinación de factores y puede diferir de una carga de trabajo a otra en su organización y entre los componentes de la carga de trabajo. Entre los factores más comunes se encuentran la importancia para la organización medida en cuanto a los ingresos o la marca, el costo total de la ejecución de la carga de trabajo (incluidos los costos de funcionamiento y de recursos), la complejidad de la carga de trabajo, la facilidad para implementar un cambio, cualquier acuerdo de licencia de software y si un cambio supusiera un aumento significativo de los costos de licencia debido a las licencias punitivas. Los componentes pueden definirse funcional o técnicamente, como servidores web y bases de datos, o recursos de computación y almacenamiento. Equilibre los factores de la forma correspondiente y desarrolle un periodo para la carga de trabajo y sus componentes. Puede decidir revisar toda la carga de trabajo cada 18 meses, revisar los servidores web cada seis meses, la base de datos cada 12 meses, la computación y el almacenamiento a corto plazo cada seis meses y el almacenamiento a largo plazo cada 12 meses.
- Definición de la exhaustividad de la revisión: defina cuánto esfuerzo se dedica a la revisión de la carga de trabajo o de los componentes de la carga de trabajo. Al igual que sucede con la frecuencia de revisión, se trata de equilibrar múltiples factores. Evalúe y priorice las oportunidades de mejora para centrar los esfuerzos donde aporten los mayores beneficios, al tiempo que estima el esfuerzo necesario para estas actividades. Si los resultados previstos no alcanzan los objetivos y el esfuerzo necesario cuesta más, repita el proceso con acciones alternativas. Sus procesos de revisión deben incluir tiempo y recursos de sus procesos para hacer posibles las mejoras incrementales continuas. Por ejemplo, puede decidir dedicar una semana de análisis al componente de base de datos, una semana de análisis a los recursos de computación y cuatro horas a las revisiones de almacenamiento.

Recursos

Documentos relacionados:

- Blog de noticias de AWS
- Tipos de computación en la nube
- Novedades de AWS

Ejemplos relacionados:

- AWS Support Proactive Services
- Revisiones regulares de las cargas de trabajo de SAP

COST10-BP02 Revisión y análisis regulares de esta carga de trabajo

Las cargas de trabajo existentes se revisan periódicamente en función de cada proceso definido para averiguar si se pueden adoptar nuevos servicios, reemplazar los existentes o rediseñar las cargas de trabajo.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: medio

Guía para la implementación

AWS agrega constantemente nuevas características para que pueda experimentar e innovar más rápidamente con la tecnología más reciente. En Novedades de AWS se detalla cómo le está yendo a AWS y se proporciona una descripción general rápida de los servicios, las características y los anuncios de expansión regional de AWS a medida que se lanzan. Puede profundizar en los lanzamientos que se han anunciado y utilizarlos para revisar y analizar sus cargas de trabajo existentes. Para obtener las ventajas de los nuevos servicios y características de AWS, debe revisar sus cargas de trabajo e implementar los nuevos servicios y características según sea necesario. Esto significa que es posible que tenga que reemplazar los servicios existentes que utiliza para la carga de trabajo o modernizarla para adoptar estos nuevos servicios de AWS. Por ejemplo, podría revisar sus cargas de trabajo y reemplazar el componente de mensajería por Amazon Simple Email Service. Esto elimina el costo de utilizar y mantener una flota de instancias, a la vez que proporciona toda la funcionalidad a un costo reducido.

Para analizar su carga de trabajo y destacar las posibles oportunidades, debe tener en cuenta no solo nuevos servicios, sino también nuevas formas de crear soluciones. Consulte los videos This is My Architecture en AWS para obtener información sobre los diseños arquitectónicos de otros clientes, sus desafíos y sus soluciones. Consulte la serie All-In para conocer las aplicaciones de los servicios de AWS en el mundo real y las historias de los clientes. También puede ver la serie de videos Back to Basics, en la que se explican, examinan y desglosan las prácticas recomendadas básicas sobre los patrones de arquitectura de nube. Otra fuente son los videos How to Build This, que están diseñados para ayudar a las personas con grandes ideas sobre cómo hacer realidad su producto mínimo viable (MVP) mediante los servicios de AWS. Es una forma de que los creadores de todo el mundo que tengan una idea sólida reciban orientación sobre arquitectura de arquitectos

de soluciones de AWS experimentados. Por último, puede revisar los materiales de recursos de Introducción, que incluyen tutoriales paso a paso.

Antes de empezar el proceso de revisión, cumpla los requisitos de su empresa en cuanto a carga de trabajo, seguridad y privacidad de los datos para poder utilizar un servicio específico o los requisitos de la región y rendimiento mientras sigue el proceso de revisión acordado.

Pasos para la implementación

- Revisión periódica de la carga de trabajo: lleve a cabo las revisiones con la frecuencia especificada aplicando el proceso que haya definido. Compruebe que dedica el esfuerzo adecuado a cada componente. Este proceso sería similar al del diseño inicial, en el que seleccionó los servicios para la optimización de costos. Analice los servicios y las ventajas que aportarían. Esta vez, tenga en cuenta el costo de aplicar el cambio, no solo las ventajas a largo plazo.
- Implementación de nuevos servicios: si la conclusión del análisis es implementar cambios, establezca primero una base de referencia de la carga de trabajo para conocer el costo actual de cada resultado. Implemente los cambios y, a continuación, haga un análisis para confirmar el nuevo costo de cada resultado.

Recursos

Documentos relacionados:

- Blog de noticias de AWS
- Novedades de AWS
- Documentación de AWS
- Introducción a AWS
- Recursos generales de AWS

Videos relacionados:

- AWS - This is My Architecture
- AWS - Back to Basics
- AWS - Serie All-In
- How to Build This

Automatización de operaciones

Prácticas recomendadas

- COST11-BP01 Automatización de las operaciones

COST11-BP01 Automatización de las operaciones

Evalúe los costos operativos en la nube, centrándose en cuantificar el ahorro de tiempo y esfuerzo en tareas administrativas, implementaciones, mitigación del riesgo de errores humanos, cumplimiento y otras operaciones mediante la automatización. Evalúe el tiempo y los costos asociados que son necesarios para los esfuerzos operativos e implemente la automatización de las tareas administrativas para minimizar el esfuerzo manual siempre que sea factible.

Nivel de riesgo expuesto si no se establece esta práctica recomendada: bajo

Guía para la implementación

La automatización de las operaciones reduce la frecuencia de las tareas manuales, mejora la eficacia y beneficia a los clientes, ya que ofrece una experiencia coherente y fiable a la hora de implementar, administrar u operar cargas de trabajo. Puede liberar recursos de la infraestructura de las tareas operativas manuales y utilizarlos para tareas de mayor valor e innovaciones, lo que mejora el valor empresarial. Las empresas necesitan una forma probada y contrastada de administrar sus cargas de trabajo en la nube. Esta solución debe ser segura, rápida y rentable, con un riesgo mínimo y máxima fiabilidad.

Comience por priorizar sus actividades operativas en función del esfuerzo necesario examinando el costo global de las operaciones. Por ejemplo, ¿cuánto tiempo se tarda en implementar nuevos recursos en la nube, aplicar cambios de optimización en los existentes o implementar las configuraciones necesarias? Analice el costo total de las acciones humanas teniendo en cuenta el costo de las operaciones y de la administración. Dé prioridad a las automatizaciones de las tareas administrativas para reducir el esfuerzo manual.

El esfuerzo de revisión debe reflejar la ventaja potencial. Por ejemplo, fíjese en el tiempo que se dedica a llevar a cabo tareas manuales comparado con el que se tarda en efectuarlas de forma automática. Priorice la automatización de actividades repetitivas, de alto valor, lentas y complejas. Las actividades de alto valor o que entrañan un mayor riesgo de errores humanos suelen ser las mejores con las que empezar la automatización, ya que el riesgo suele plantear un costo operativo adicional no deseado (por ejemplo, que el equipo de operaciones trabaje horas extra).

Utilice herramientas de automatización como AWS Systems Manager o AWS Config para optimizar las operaciones, el cumplimiento, la supervisión, el ciclo de vida y los procesos de terminación. Con servicios de AWS, herramientas y productos de terceros, puede personalizar las automatizaciones que implemente para satisfacer sus requisitos específicos. En la tabla siguiente, se muestran algunas de las funciones y capacidades de funcionamiento básicas que puede conseguir con los servicios de AWS para automatizar la administración y el funcionamiento:

- AWS Audit Manager: audite continuamente su uso de AWS para simplificar la evaluación de riesgos y cumplimiento.
- AWS Backup: administre y automatice la protección de datos de forma centralizada.
- AWS Config: configure los recursos de computación, evalúe y audite las configuraciones y el inventario de recursos.
- AWS CloudFormation: lance recursos de alta disponibilidad con infraestructura como código.
- AWS CloudTrail: administración, cumplimiento y control de los cambios de TI.
- Amazon EventBridge: programe eventos y active AWS Lambda para que actúen.
- AWS Lambda: automatice los procesos repetitivos activándolos con eventos o ejecutándolos según una programación fija con AWS EventBridge.
- AWS Systems Manager: inicie y detenga las cargas de trabajo, aplique revisiones a los sistemas operativos y automatice la configuración y la administración continua.
- AWS Step Functions: programe trabajos y automatice los flujos de trabajo.
- AWS Service Catalog: consumo de plantillas, infraestructura como código con cumplimiento y control.

Si desea adoptar las automatizaciones de forma inmediata con el uso de productos y servicios de AWS y no tiene destrezas en la organización, contacte con AWS Managed Services (AMS), AWS Professional Services o los socios de AWS para aumentar la adopción de la automatización y mejorar su excelencia operativa en la nube.

AWS Managed Services (AMS) es un servicio que utiliza la infraestructura de AWS en nombre de los socios y clientes de la empresa. Proporciona un entorno seguro y conforme a las normativas en el que puede implementar sus cargas de trabajo. AMS utiliza modelos operativos de nube empresarial con automatización para permitirle satisfacer los requisitos de su organización, trasladarse a la nube más rápidamente y reducir los costos de administración continua.

AWS Professional Services también puede ayudarlo a conseguir los resultados empresariales deseados y a automatizar las operaciones con AWS. Ayudan a los clientes a implementar

operaciones de TI automatizadas, sólidas y ágiles, así como capacidades de gobernanza optimizadas para la nube. Para ver ejemplos detallados de supervisión y las prácticas recomendadas, consulte el documento técnico sobre el pilar de excelencia operativa.

Pasos para la implementación

- Creación única e implementación múltiple: utilice infraestructura como código, como CloudFormation, el SDK de AWS o la AWS CLI, para implementar una vez y utilice la creación varias veces en entornos similares o en escenarios de recuperación de desastres. Etiquete mientras implementa para hacer un seguimiento de su consumo, tal y como se define en otras prácticas recomendadas. Utilice AWS Launch Wizard para reducir el tiempo de implementación de muchas cargas de trabajo empresariales populares. AWS Launch Wizard le guía en el proceso de dimensionamiento, configuración e implementación de cargas de trabajo empresariales siguiendo las prácticas recomendadas de AWS. También puede usar Service Catalog, que lo ayuda a crear y administrar plantillas aprobadas de infraestructura como código para su uso en AWS, de modo que cualquiera pueda descubrir recursos de nube aprobados y de autoservicio.
- Automatice el cumplimiento continuo: considere la posibilidad de automatizar la evaluación y la corrección de las configuraciones registradas en función de los estándares predefinidos. Si combina AWS Organizations con las capacidades de AWS Config y AWS CloudFormation, puede administrar y automatizar de manera eficiente el cumplimiento de la configuración a escala para cientos de cuentas de miembro. Puede revisar los cambios en las configuraciones y las relaciones entre los recursos de AWS y profundizar en el historial de una configuración de recursos.
- Automatice las tareas de supervisión: AWS proporciona varias herramientas que puede utilizar para supervisar servicios. Puede configurar estas herramientas para automatizar las tareas de supervisión. Cree e implemente un plan de supervisión con el que se recopilen datos de supervisión de todas las partes de su carga de trabajo para que pueda depurar más fácilmente un error multipunto en el caso de que se produzca. Por ejemplo, puede usar las herramientas de supervisión automatizadas para observar a Amazon EC2 y que le informe cuando algo va mal en las comprobaciones del estado del sistema, las comprobaciones del estado de las instancias y las alarmas de Amazon CloudWatch.
- Automatice el mantenimiento y las operaciones: ejecute las operaciones de rutina automáticamente sin intervención humana. Usando los servicios y las herramientas de AWS, puede elegir qué automatizaciones de AWS implementar y personalizar según sus requisitos específicos. Por ejemplo, utilice el Generador de imágenes de EC2 para crear, probar e implementar imágenes de máquinas virtuales y contenedores para su uso en AWS, en las instalaciones o para aplicar revisiones a las instancias de EC2 con AWS SSM. Si la acción deseada no se puede ejecutar con los servicios de AWS o si necesita acciones más complejas

con los recursos de filtrado, automatice sus operaciones mediante AWS Command Line Interface (AWS CLI) o las herramientas de AWS SDK. AWS CLI ofrece la posibilidad de automatizar todo el proceso de control y administración de los servicios de AWS mediante scripts sin utilizar la Consola de administración de AWS. Seleccione sus SDK de AWS preferidos para interactuar con los servicios de AWS. Para ver otros ejemplos de código, consulte el repositorio de ejemplos de código del SDK de AWS.

- Creación de un ciclo de vida continuo con las automatizaciones: es importante que establezca y conserve políticas de ciclo de vida consolidadas, no solo en lo que respecta a las normativas o la redundancia, sino también a fin de optimizar los costos. Puede usar AWS Backup para administrar y automatizar de manera centralizada la protección de datos de los almacenes de datos, como buckets, volúmenes, bases de datos y sistemas de archivos. Puede utilizar Amazon Data Lifecycle Manager para automatizar la creación, retención y eliminación de instantáneas de EBS y las AMI respaldadas por EBS.
- Eliminación de recursos innecesarios: es bastante común acumular recursos no utilizados de Cuentas de AWS en un entorno de pruebas o de desarrollo. Los desarrolladores crean y experimentan con diversos servicios y recursos como parte del ciclo de desarrollo normal y, después, no eliminan esos recursos cuando ya no los necesitan. Los recursos no utilizados pueden dar lugar a costos innecesarios para la organización que a veces son elevados. La eliminación de estos recursos puede reducir los costos de operación de estos entornos. Asegúrese de que sus datos no son necesarios o haga una copia de seguridad si no está claro. Puede utilizar AWS CloudFormation para limpiar las pilas implementadas, con lo que se elimina automáticamente la mayoría de los recursos definidos en la plantilla. Como alternativa, puede crear una automatización para la eliminación de recursos de AWS con herramientas como aws-nuke.

Recursos

Documentos relacionados:

- Modernizing operations in the Nube de AWS
- AWS Services for Automation
- Infraestructura y automatización
- AWS Systems Manager Automation
- Supervisión automatizada y manual
- AWS automations for SAP administration and operations

- AWS Managed Services
- AWS Servicios profesionales de

Videos relacionados:

- Automate Continuous Compliance at Scale in AWS
- AWS Backup Demo: Cross-Account & Cross-Region Backup
- Patching for your Amazon EC2 Instances

Ejemplos relacionados:

- Reinventing automated operations (Part I)
- Reinventing automated operations (Part II)
- Automate deletion of AWS resources by using aws-nuke
- Delete unused Amazon EBS volumes by using AWS Config and AWS SSM
- Automate continuous compliance at scale in AWS
- IT Automations with AWS Lambda

Conclusión

La optimización de costos y la administración financiera en la nube son un esfuerzo continuo. Debe trabajar periódicamente con sus equipos de Finanzas y Tecnología, revisar el enfoque arquitectónico y actualizar su selección de componentes.

El objetivo de AWS es ayudarlo a minimizar los costos a la vez que crea implementaciones con gran resiliencia, capacidad de respuesta y capacidad de adaptación. Para optimizar realmente el costo de su implementación, utilice las herramientas, técnicas y prácticas recomendadas que se detallan en el presente documento.

Colaboradores

Los colaboradores de este documento son:

- Fatih (Ben) Mergen, Cost Optimization Pillar Lead, Well-Architected, Amazon Web Services
- Keith Jarrett, Business Development Lead – Cost Optimization, Amazon Web Services
- Arthur Basbaum, Business Developer Manager, Amazon Web Services
- Jarman Hauser, Commercial Architect, Amazon Web Services

Documentación adicional

Para obtener información adicional, consulte:

- Marco de AWS Well-Architected
- Centro de arquitectura de AWS

Revisiones del documento

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

Cambio	Descripción	Fecha
Actualización de las directrices de prácticas recomendadas	Se hicieron varias actualizaciones en las prácticas recomendadas. Se agregó una nueva práctica recomendada: COST06-BP04.	27 de junio de 2024
Actualización de las directrices de prácticas recomendadas	Se hicieron actualizaciones menores en todas las prácticas recomendadas.	6 de diciembre de 2023
Actualización de las directrices de prácticas recomendadas	Se actualizaron las prácticas recomendadas con nuevas directrices en todo el pilar.	3 de octubre de 2023
Actualización de las directrices de prácticas recomendadas	Se actualizaron las prácticas recomendadas con nuevas directrices en las siguientes áreas: Gobernanza, Supervisión del costo y el uso, Selección del mejor modelo de precios y Administración de la demanda y suministro de recursos.	13 de julio de 2023
Actualizaciones del nuevo marco	Se actualizaron las prácticas recomendadas con una guía prescriptiva y se agregaron nuevas prácticas recomendadas. Se agregó la pregunta COST 11 con la nueva	10 de abril de 2023

	práctica recomendada: COST11-BP01.	
Actualización menor	Se restablecieron las directric es que faltaban en la sección del modelo de precios.	13 de enero de 2023
Documento técnico actualiza do	Se actualizaron las prácticas recomendadas con una nueva guía de implementación.	15 de diciembre de 2022
Documento técnico actualiza do	Se ampliaron las prácticas recomendadas y se agregaron planes de mejora.	20 de octubre de 2022
Actualización menor	Se agregó el pilar de sostenibi lidad a la introducción.	2 de diciembre de 2021
Actualización menor	Se actualizaron los enlaces.	25 de abril de 2021
Actualización menor	Se actualizaron los enlaces.	10 de marzo de 2021
Actualizaciones del nuevo marco	Se actualizó para incorpora r CFM, nuevos servicios y también la integración con Well-Architected.	8 de julio de 2020
Documento técnico actualiza do	Se actualizó para reflejar los cambios en AWS e incorpora r los aprendizajes de las revisiones con los clientes.	1 de julio de 2018
Documento técnico actualiza do	Se actualizó para reflejar los cambios en AWS e incorpora r los aprendizajes de las revisiones con los clientes.	1 de noviembre de 2017

Publicación inicial	Se publicó Pilar de optimización de costos: Marco de AWS Well-Architected.	1 de noviembre de 2016

Avisos

Es responsabilidad de los clientes realizar su propia evaluación independiente de la información que contiene este documento. El presente documento: (a) tiene sólo fines informativos, (b) representa las ofertas y prácticas actuales de los productos de AWS, que están sujetas a cambios sin previo aviso, y (c) no supone ningún compromiso ni garantía por parte de AWS y sus filiales, proveedores o licenciantes. Los productos o servicios de AWS se proporcionan “tal cual”, sin garantías, afirmaciones ni condiciones de ningún tipo, ya sean expresas o implícitas. Las responsabilidades y obligaciones de AWS con respecto a sus clientes se controlan mediante los acuerdos de AWS y este documento no forma parte ni modifica ningún acuerdo entre AWS y sus clientes.

© 2023 Amazon Web Services, Inc. o sus filiales. Todos los derechos reservados.

Glosario de AWS

Para ver la terminología más reciente de AWS, consulte el Glosario de AWS en la Referencia de Glosario de AWS.