Optimice los costos de las cargas de trabajo de Microsoft en AWS

AWS Guía prescriptiva

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

AWS Guía prescriptiva: Optimice los costos de las cargas de trabajo de Microsoft en AWS

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

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

Table of Contents

Introducción ... 1
Descripción general ... 1
Público ... 1
Cómo usar esta guía ... 2
Resultados empresariales específicos ... 3
Proceso de optimización de costes ... 4
Principales recomendaciones para optimizar los costes ... 7
Descripción general ... 7
Recomendaciones principales ... 7
AWS Evaluación de optimización y licencias ... 9
Descripción general ... 9
Opciones de evaluación ... 10
Evaluación completa ... 10
Alcance las cargas de trabajo ... 11
Recopile datos ... 11
Análisis de datos ... 12
Planifique los próximos pasos ... 14
Impacto de la evaluación ... 15
Pasos a seguir a continuación ... 16
Recursos adicionales ... 16
Windows en Amazon EC2 ... 17
Automatice los horarios de parada e inicio ... 18
Descripción general ... 18
Casos prácticos ... 18
Escenario de optimización de costos ... 19
Recomendaciones de optimización de costos ... 21
Recursos adicionales ... 34
Cargas de trabajo de Windows del tamaño correcto ... 35
Descripción general ... 35
Escenario de optimización de costes ... 35
Recomendaciones de optimización de costos ... 36
Recomendaciones ... 46
Recursos adicionales ... 46
Seleccione el tipo de instancia adecuado para las cargas de trabajo de Windows ... 47

Descripción general ... 47
Recomendaciones de optimización de costos ... 48
Pasos a seguir a continuación ... 57
Recursos adicionales ... 59
Traiga licencias para cargas de trabajo de Windows y SQL Server ... 59
Descripción general ... 59
Hosts EC2 dedicados de Amazon ... 60
AWS opciones de licencia ... 64
Incorpora licencias de Windows Server ... 65
Escenarios de optimización de costes ... 66
Recomendaciones de optimización de costos ... 73
Recursos adicionales ... 73
Optimice el gasto para Windows en Amazon EC2 ... 74
Descripción general ... 74
Descripción de Savings Plans ... 74
Escenarios de optimización de costes ... 81
Recomendaciones de optimización de costos ... 85
Recursos adicionales ... 87
Supervise los costes mediante herramientas AWS ... 87
Descripción general ... 87
Recomendaciones de optimización de costos ... 88
Recursos adicionales ... 91
SQL Server ... 92
Elija una solución de alta disponibilidad y recuperación ante desastres ... 93
Descripción general ... 93
Grupos de disponibilidad de SQL Server Always On ... 94
Instancias de clúster de conmutación por error de SQL Server Always On ... 96
SIOS DataKeeper ... 98
Grupos de disponibilidad AlwaysOn ... 100
Grupos de disponibilidad distribuida ... 101
Envío de registros ... 102
AWS Database Migration Service ... 104
AWS Elastic Disaster Recovery ... 105
Comparación de costos ... 106
Recomendaciones de optimización de costos ... 111
Recursos adicionales ... 112

Comprenda las licencias de SQL Server ... 113
Descripción general ... 113
AWS opciones de licenciamiento ... 113
El impacto en los costos de traer licencias ... 114
Optimización de licencias ... 115
Recomendaciones de optimización de costos ... 115
Recursos adicionales ... 46
Seleccione la EC2 instancia adecuada para las cargas de trabajo de SQL Server ... 121
Descripción general ... 121
Comparación de costos ... 122
Escenario de optimización de costes ... 123
Recomendaciones de optimización de costos ... 125
Recursos adicionales ... 129
Consolide las instancias ... 129
Descripción general ... 129
Escenario de optimización de costes ... 130
Recomendaciones de optimización de costos ... 131
Recursos adicionales ... 133
Compare las ediciones de SQL Server ... 133
Descripción general ... 133
Impacto del costo ... 134
Recomendaciones de optimización de costos ... 136
Recursos adicionales ... 142
Evalúe la edición SQL Server Developer ... 143
Descripción general ... 143
Impacto del costo ... 143
Recursos adicionales ... 46
Evalúe SQL Server en Linux ... 147
Descripción general ... 147
Impacto del costo ... 148
Recomendaciones de optimización de costos ... 149
Recursos adicionales ... 150
Optimice las estrategias de respaldo de SQL Server ... 150
Descripción general ... 150
Respaldo a nivel de servidor mediante instantáneas habilitadas para VSS ... 151
Realice una copia de seguridad de SQL Server mediante AWS Backup ... 153

Copia de seguridad a nivel de la base de datos 155
Recomendaciones de optimización de costos 165
Recursos adicionales 168
Modernice las bases de datos de SQL Server 169
Descripción general 169
Ofertas de bases de datos 169
Comparación entre Amazon RDS y Aurora 170
Recomendaciones de optimización de costos 172
Recursos adicionales 177
Optimice el almacenamiento para SQL Server 178
Descripción general 178
Tipos de almacenamiento SSD, rendimiento y costo para Amazon EBS 178
Optimización general de costos de SSD para Amazon EBS 180
Recursos adicionales 182
Optimice las licencias de SQL Server mediante Compute Optimizer 182
Descripción general 183
Recomendaciones de optimización de costos 183
Configurar Compute Optimizer 184
Recursos adicionales 185
Optimice el tamaño de SQL Server mediante Compute Optimizer 186
Descripción general 186
Configurar Compute Optimizer 186
Recursos adicionales 187
Revise Trusted Advisor las recomendaciones para las cargas de trabajo de SQL Server 187
Descripción general 187
Recomendaciones de optimización de costos 188
Configurar Trusted Advisor 189
Recursos adicionales 189
Contenedores 190
Mueva las aplicaciones de Windows a contenedores 191
Descripción general 191
Beneficios de costos 191
Recomendaciones de optimización de costos 193
Pasos a seguir a continuación 197
Recursos adicionales 197
Optimice los costos de AWS Fargate las tareas en Amazon ECS 197

Descripción general ... 197
Beneficios de costos ... 198
Recomendaciones de optimización de costos ... 198
Pasos a seguir a continuación ... 205
Recursos adicionales ... 205
Obtenga visibilidad de sus costos de Amazon EKS ... 205
Descripción general ... 205
Costos y beneficios ... 206
Recomendaciones de optimización de costos ... 206
Pasos a seguir a continuación ... 210
Recursos adicionales ... 210
Cambie la plataforma de las aplicaciones de Windows con App2Container ... 211
Descripción general ... 211
Costos y beneficios ... 212
Recomendaciones de optimización de costos ... 213
Pasos a seguir a continuación ... 213
Recursos adicionales ... 213
Almacenamiento ... 214
Amazon EBS ... 214
Migre los volúmenes de Amazon EBS de gp2 a gp3 ... 215
Modifique las instantáneas de Amazon EBS ... 219
Eliminar volúmenes de Amazon EBS no adjuntos ... 222
Amazon FSx ... 226
Elija el almacenamiento de archivos SMB adecuado ... 227
Habilite la deduplicación de datos en Amazon FSx ... 232
Conozca la fragmentación de datos en Windows FSx File Server ... 234
Comprenda el uso del volumen del disco duro en Amazon FSx ... 239
Utilice una única zona de disponibilidad ... 242
AWS Storage Gateway ... 244
Puerta de enlace de archivos Amazon S3 ... 245
Amazon FSx File Gateway ... 245
Impacto del costo ... 245
Recomendaciones de optimización de costos ... 248
Recursos adicionales ... 250
Active Directory ... 251
Active Directory autogestionado en Amazon EC2 ... 251

Descripción general ... 251
Impacto del costo ... 251
Recomendaciones de optimización de costos ... 252
Recursos adicionales ... 257
AWS Managed Microsoft AD ... 257
Descripción general ... 257
Impacto del costo ... 257
Recomendaciones de optimización de costos ... 257
Recursos adicionales ... 259
Conector de AD ... 259
Descripción general ... 259
Impacto del costo ... 260
Recomendaciones de optimización de costos ... 260
Recursos adicionales ... 261
.NET ... 262
Refactoriza la versión moderna de.NET y cámbiate a Linux ... 263
Descripción general ... 263
Impacto del costo ... 263
Recomendaciones de optimización de costos ... 264
Consideraciones y recursos adicionales ... 265
Contenedore las aplicaciones.NET ... 266
Descripción general ... 266
Impacto del costo ... 266
Recomendaciones de optimización de costos ... 268
Recursos adicionales ... 271
Utilice instancias y contenedores de Graviton ... 271
Descripción general ... 271
Impacto del costo ... 271
Recomendaciones de optimización de costos ... 273
Recursos adicionales ... 274
Support dinámico para aplicaciones estáticas de.NET Framework ... 275
Descripción general ... 275
Impacto del costo ... 280
Recomendaciones de optimización de costos ... 281
Recursos adicionales ... 283
Utilice el almacenamiento en caché para reducir la demanda de bases de datos ... 283

Descripción general ... 283
Impacto del costo ... 283
Recomendaciones de optimización de costos ... 284
Recursos adicionales ... 291
Considere la posibilidad de utilizar .NET sin servidor ... 291
Descripción general ... 291
Impacto del costo ... 292
Recomendaciones de optimización de costos ... 292
Recursos adicionales ... 297
Considere la posibilidad de crear bases de datos especialmente diseñadas ... 297
Descripción general ... 297
Impacto del costo ... 301
Recomendaciones de optimización de costos ... 304
Recursos adicionales ... 305
Pasos a seguir a continuación ... 307
Historial de documentos ... 308
Glosario ... 309
... 309
A ... 310
B ... 313
C ... 315
D ... 318
E ... 322
F ... 325
G ... 327
H ... 328
I ... 329
L ... 332
M ... 333
O ... 337
P ... 340
Q ... 343
R ... 343
S ... 346
T ... 350
U ... 352

V 353
W 353
Z 354
..... ccclvi

Optimice los costos de las cargas de trabajo de Microsoft en AWS

Bill Pfeiffer, Chase Lindeman y Kevin Sookhan, Amazon Web Services ()AWS

Octubre de 2024 (historia del documento)

Descripción general

Esta guía proporciona recomendaciones, prácticas recomendadas y estrategias para ayudar a optimizar los costes de las cargas de trabajo de Microsoft en AWS. La guía también incluye AWS conocimientos básicos, técnicas de optimización de costes y arquitecturas de referencia para ayudarle a crear y automatizar cargas de trabajo rentables y de alto rendimiento que cumplan sus objetivos empresariales. En conjunto, esta guía se denomina Microsoft on AWS Cost Optimization (MACO). La guía MACO fue desarrollada por expertos del sector y se basa en escenarios del mundo real.

Esta guía cubre las siguientes cargas de trabajo de Microsoft:

- Windows en Amazon Elastic Compute Cloud (Amazon EC2)
- SQL Server
- Contenedores
- Almacenamiento
- Active Directory
- .NET

Público

Esta guía está destinada a arquitectos, ingenieros, administradores, directores CTOs, responsables de la toma de decisiones técnicas y AWS socios. Es útil, pero no necesario, tener experiencia previa y conocimientos básicos sobre AWS facturación, tecnologías de Microsoft y administración de AWS sistemas.

Cómo usar esta guía

Puede utilizar esta guía para planificar e implementar su transición de MACO a la nube. Le recomendamos que lea esta guía de principio a fin para obtener una comprensión completa de las opciones y los enfoques para optimizar los costos de sus cargas de trabajo de Microsoft. AWS Puede revisar las siguientes secciones de carga de trabajo en función de las necesidades de su organización:

- Windows en Amazon EC2
- SQL Server
- Contenedores
- Almacenamiento
- Active Directory
- .NET

> **Important**
>
> Los ejemplos de código que se proporcionan en esta guía son únicamente para fines de demostración. Se recomienda probar todo el código en un entorno de desarrollo antes de usarlo en un entorno de producción. Antes de implementar cualquier código, le recomendamos que lo pruebe en lotes pequeños y, a continuación, revise los cambios en los costos que se derivan del código mediante AWS Cost Explorersu uso. Esto puede ayudarte a solucionar casos extremos y otros problemas que pueden convertirse en problemas más adelante.

> **Important**
>
> Los ejemplos de precios de esta guía se basan en los precios vigentes en el momento de la publicación. Los precios están sujetos a cambios. Además, sus costos pueden variar en función de sus Región de AWS Servicio de AWS cuotas y otros factores relacionados con su entorno de nube.

Resultados empresariales específicos

Esta guía puede ayudarle a usted y a su organización a lograr los siguientes resultados empresariales:

- Aprenda a utilizar una evaluación de AWS optimización y licencias (AWS OLA) para evaluar y optimizar sus entornos locales y en la nube actuales, en función del uso de los recursos, las licencias de terceros y las dependencias de las aplicaciones.
- Desarrolle un modelo de negocio para la optimización de costos mediante la Calculadora de AWS modernización para cargas de trabajo de Microsoft.
- Optimice los costos de sus cargas de trabajo específicas de Microsoft, incluidas las cargas de trabajo para Windows en Amazon Elastic Compute Cloud EC2 (Amazon), SQL Server, contenedores, almacenamiento, Active Directory y.NET.

Proceso de optimización de costes

El alcance, el calendario y la ruta específica de su migración a la nube dependen de los objetivos empresariales, los requisitos técnicos y otros factores. En esta sección se ofrece un ejemplo de un proceso de migración a la nube que se centra en la gestión financiera en la nube AWS y sigue las recomendaciones y prácticas recomendadas de MACO. Puede usar este ejemplo para comprender cómo diseñar un proceso de migración a la nube para las cargas de trabajo de Microsoft.

Las siguientes tareas de alto nivel ilustran el enfoque que una organización podría adoptar para implementar las recomendaciones y mejores prácticas de MACO:

- Establezca una estrategia de etiquetado y habilite las etiquetas de asignación de costes definidas por el usuario. Para obtener más información, consulte el AWS documento técnico Mejores prácticas para etiquetar recursos. AWS
- Defina los presupuestos en función de las aplicaciones, los equipos o los departamentos. Para obtener más información, consulte Gestión de los costes AWS Budgets en la Guía del usuario AWS de Billing and Cost Management.
- Realice una evaluación AWS de optimización y licencias (AWS OLA) para acelerar los ahorros. Para obtener más información, consulte la evaluación de la AWS optimización y las licencias en la AWS documentación.
- Traiga su propia licencia (BYOL) para cargas de trabajo de Windows y SQL Server mediante el uso de. Amazon Elastic Compute Cloud Dedicated Hosts Para obtener más información, consulte la sección Traiga licencias para cargas de trabajo de Windows y SQL Server de esta guía.
- Optimice sus licencias de SQL Server en. AWS Para obtener más información, consulte la sección Comprenda las licencias de SQL Server de esta guía.
- Seleccione el tipo de instancia adecuado para las cargas de trabajo de Windows. Para obtener más información, consulte la sección Seleccione el tipo de instancia adecuado para las cargas de trabajo de Windows de esta guía.
- Seleccione el tipo de instancia correcto para las cargas de trabajo de SQL. Para obtener más información, consulte la sección Seleccione la EC2 instancia adecuada para las cargas de trabajo de SQL Server de esta guía.
- Migre Amazon Elastic Block Store (Amazon EBS) de gp2 a gp3. Para obtener más información, consulte la sección Migre los volúmenes de Amazon EBS de gp2 a gp3 de esta guía.
- Controle las cargas de trabajo con EC2 el programador de instancias activado. AWS Para obtener más información, consulte la sección Automatice los horarios de parada e inicio de esta guía.

- Elimine los costos de SQL Server para las cargas de trabajo que no son de producción mediante SQL Server Developer Edition. Para obtener más información, consulte la sección Evalúe la edición SQL Server Developer de esta guía.
- Utilice una única zona de disponibilidad para Amazon FSx for Windows File Server para las cargas de trabajo de desarrollo y prueba. Para obtener más información, consulte la sección Utilice una única zona de disponibilidad de esta guía.
- Ajuste el tamaño adecuado de sus cargas de trabajo de Windows mediante. AWS Compute Optimizer Para obtener más información, consulte la sección Cargas de trabajo de Windows del tamaño correcto de esta guía.
- Optimice sus gastos en Windows en Amazon EC2 con Savings Plans. Para obtener más información, consulte la sección Optimice el gasto para Windows en Amazon EC2 de esta guía.
- Active la deduplicación de datos FSx para Windows File Server. Para obtener más información, consulte la sección Habilite la deduplicación de datos en Amazon FSx de esta guía.
- Utilice la fragmentación de datos para los sistemas de archivos y FSx para el servidor de archivos de Windows. Para obtener más información, consulte la sección Conozca la fragmentación de datos en Windows FSx File Server de esta guía.
- Optimice sus estrategias de respaldo de SQL Server. Para obtener más información, consulte la sección Optimice las estrategias de respaldo de SQL Server de esta guía.
- Haga que las aplicaciones estáticas de.NET Framework admitan el escalado dinámico. Para obtener más información, consulte esta guía. Support dinámico para aplicaciones estáticas de.NET Framework
- Utilice microservicios.NET sin servidor. Para obtener más información, consulte la sección Considere la posibilidad de utilizar .NET sin servidor de esta guía.
- Mueva sus aplicaciones de Windows a contenedores. Para obtener más información, consulte la sección Contenedore las aplicaciones.NET de esta guía.
- Úselo AWS Compute Optimizerpara ajustar el tamaño de los contenedores de Windows que se estén ejecutando AWS Fargate para Amazon Elastic Container Service (Amazon ECS). Para obtener más información, consulte la sección Activar Compute Optimizer de esta guía.
- Refactoriza la versión moderna de.NET y cámbiate a Linux. Para obtener más información, consulte la sección Refactoriza la versión moderna de.NET y cámbiate a Linux de esta guía.
- Aproveche las instancias y los contenedores de Graviton. Para obtener más información, consulte la sección Utilice instancias y contenedores de Graviton de esta guía.
- Modernice las bases de datos de SQL Server. Para obtener más información, consulte la sección Modernice las bases de datos de SQL Server de esta guía.

- Diseñe la infraestructura de Active Directory. Para obtener más información, consulte la sección Active Directory de esta guía.

Para obtener más información sobre la experiencia de un cliente centrada en la gestión financiera en la nube AWS, consulte el AWS documento técnico sobre la capacidad de gestión financiera en la nube.

Principales recomendaciones para optimizar los costes

Descripción general

La optimización de costes es uno de los pilares del AWS Well-Architected Framework y desempeña un papel fundamental en sus planes de migración a la nube. Encontrará recomendaciones para la optimización de costos en esta guía, pero en esta sección se describen las recomendaciones de mayor impacto. Puede implementar estas recomendaciones rápidamente y tendrán un impacto significativo en su organización. Estas recomendaciones pueden ayudar a sentar las bases para todo su esfuerzo de optimización de costos.

Recomendaciones principales

En la siguiente tabla, se enumeran las principales recomendaciones para las optimizaciones de costes de mayor impacto. La columna «Dificultad de implementación» clasifica cada optimización en función de una escala que va desde lo que es más fácil de implementar (1) hasta lo que es más difícil de implementar (5). La columna «Ahorros estimados» muestra una estimación basada en porcentajes de cuánto puede ahorrar su organización por cada optimización recomendada.

Optimizaciones	Dificultad de implementación	Ahorro estimado
Cargas de trabajo de Windows del tamaño correcto	3	25%
Traiga licencias para cargas de trabajo de Windows y SQL Server	3	30%
Evalúe la edición SQL Server Developer	2	20%
Comprenda las licencias de SQL Server	2	Hasta un 50%
Automatice los horarios de parada e inicio	3	Hasta un 40%

Optimizaciones	Dificultad de implementación	Ahorro estimado
Seleccione el tipo de instancia adecuado para las cargas de trabajo de Windows	1	10— 30%
Refactoriza la versión moderna de.NET y cámbiate a Linux	5	10— 20%
Optimice el gasto para Windows en Amazon EC2	3	Hasta un 20— 40%
Migre los volúmenes de Amazon EBS de gp2 a gp3	4	Hasta un 20%

> **Important**
>
> Los ahorros estimados de la tabla anterior se aplican a cada ámbito técnico individual, no al AWS gasto total de una cuenta. Por ejemplo, puede implementar el programador de instancias en diversos tipos y tamaños de entornos, lo que puede alterar los posibles ahorros. Las estimaciones se aplican específicamente a los costos de las EC2 instancias de Amazon y no implican ningún ahorro general para otras Servicios de AWS. Estas estimaciones se proporcionan como un indicador, no como una garantía.

Los expertos de MACO están disponibles para hablar más a fondo sobre las optimizaciones de costos. Para programar una reunión para analizar en profundidad su caso de uso, póngase en contacto con el equipo de su cuenta o envíe un correo electrónico a optimize-microsoft@amazon.com.

AWS Evaluación de optimización y licencias

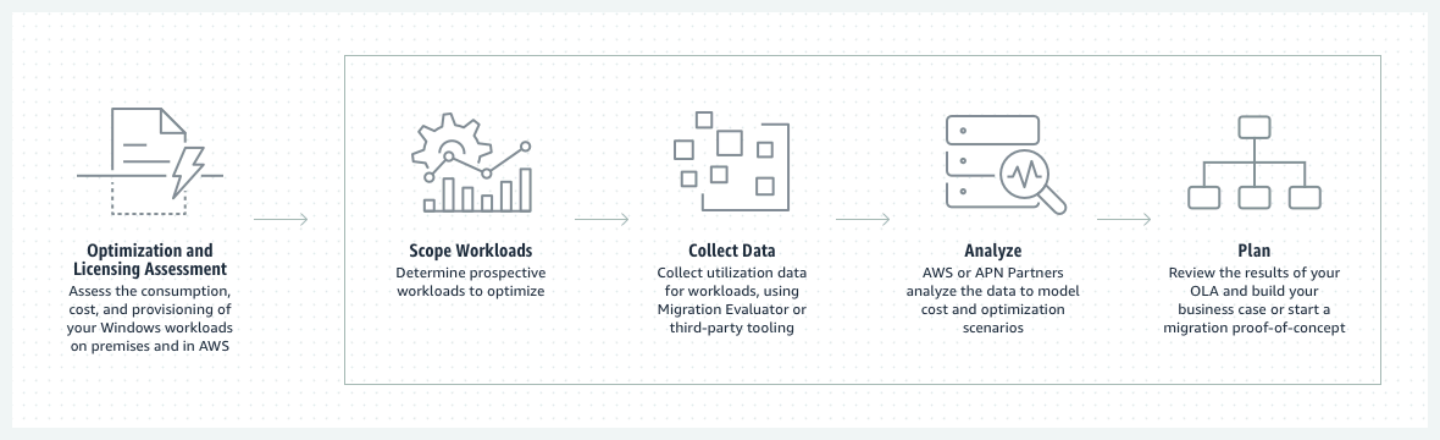
Descripción general

Una evaluación de AWS optimización y licencias (AWS OLA) puede ayudarlo a evaluar y optimizar sus entornos locales y de nube actuales, en función del uso de los recursos, las licencias de terceros y las dependencias de las aplicaciones. Puede usar AWS OLA para ayudar a su organización a desarrollar una estrategia de migración y licencias que permita ahorrar costos a medida que migra AWS o evalúa las cargas de trabajo de Microsoft existentes. AWS Un AWS OLA también puede ayudarle a lograr lo siguiente:

- Comprenda las implementaciones, el rendimiento de las aplicaciones y los contratos existentes.
- Asigne el tamaño adecuado a sus recursos.
- Desarrolle una hoja de ruta para. Nube de AWS
- Reduzca o elimine los costos utilizando las inversiones existentes y pagando solo por lo que utilice.

Le recomendamos que haga de la AWS OLA el primer paso en su proceso de optimización de costes. Puede trabajar con el AWS Partner Network para completar un AWS OLA. Le ayudarán a recopilar datos de evaluación y le ofrecerán recomendaciones para optimizar los costes de licencias e instancias.

El siguiente diagrama proporciona una visión general del proceso de evaluación.



Opciones de evaluación

Puede elegir entre dos opciones de AWS OLA para sus cargas de trabajo de Microsoft en AWS:

- Versión Lite: en este caso de uso, todas tus cargas de trabajo están activadas. VMware Puede proporcionar AWS un resultado de RVTools. Luego, AWS puede ofrecer un tiempo de entrega de 1 a 5 días. Este enfoque utiliza point-in-time información extraída directamente de VMware vCenter para desarrollar recomendaciones de tamaño y ofrecer opciones de precios a pedido.
- Versión completa: en este caso de uso, tiene un entorno mixto que se ejecuta en diferentes proveedores de nube, servidores físicos y servidores virtuales. AWS utiliza agentes del sistema operativo para recopilar datos de uso durante un período de 14 a 30 días. Esto le permite AWS tomar decisiones informadas sobre el tamaño de las instancias en función de los patrones de uso de las aplicaciones. AWS utiliza varias herramientas de terceros, como Cloudamize, para completar el análisis. AWS trabaja con él AWS Partner Network para ayudar a realizar la evaluación final del costo total de propiedad (TCO) con múltiples opciones de precios que tienen en cuenta los modelos de precios y las diferentes arquitecturas.

Evaluación completa

La evaluación completa de la AWS OLA comienza con una llamada telefónica de una hora. Durante esta llamada, le AWS ayuda a determinar la AWS infraestructura más óptima para respaldar su migración, a elegir un método de recopilación de datos y a establecer un cronograma para su finalización. La implementación de herramientas de detección en su organización depende del método de recopilación de datos, del tamaño de la organización y de las herramientas que utilice la organización para administrar su flota de servidores. La recopilación de datos de uso suele tardar dos semanas.

El proceso completo de la AWS OLA tarda entre 30 y 45 días y consta de las siguientes fases:

- Alcance las cargas de trabajo
- Recopile datos
- Análisis de datos
- Planifique los próximos pasos

Alcance las cargas de trabajo

En primer lugar, AWS trabaja con usted y su equipo para determinar el alcance de la evaluación. Por lo general, se desglosa por tipo de entorno (por ejemplo, no productivo y de producción). El alcance incluye la ubicación de las cargas de trabajo. Pueden ser cargas de trabajo a las que está migrando AWS, cargas de trabajo que ya se están ejecutando AWS (por ejemplo, OLA AWS para Amazon EC2) o cargas de trabajo que se ejecutan en otros proveedores de nube.

Recopile datos

A continuación, AWS implementa herramientas para ayudar a descubrir recursos y recopilar datos de rendimiento de sus servidores. Estas herramientas vienen en cuatro opciones de implementación:

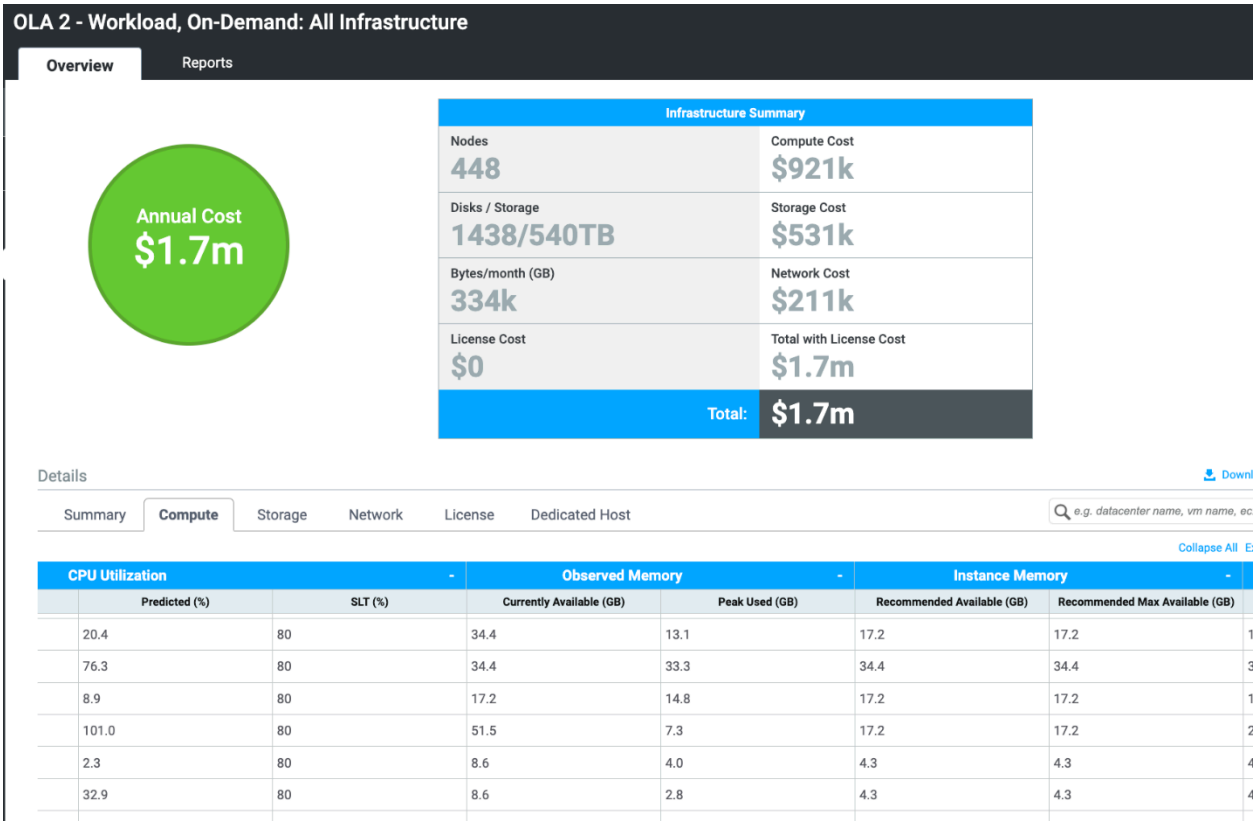
- Herramientas que pueden consultar el hipervisor (solo se requieren credenciales de VMware vCenter o Hyper-V)
- Agentes que se pueden implementar en máquinas físicas o virtuales
- Descubrimiento sin agente mediante SSH, Administración remota de Windows (WinRM) o Instrumentación de administración de Windows (WMI) según el entorno y el sistema operativo
- Recopilación y análisis de datos en archivos planos

Para el despliegue de sus herramientas, puede combinar cada opción y consolidar los resultados. Es fundamental asegurarse de que la opción que elija no suponga una carga excesiva para sus recursos de TI. AWS se esfuerza por hacer que el proceso de evaluación sea lo más completo posible. Más allá de una breve llamada telefónica para ayudar con la configuración, el equipo de AWS OLA y los arquitectos de soluciones especializados de Microsoft prepararán el análisis del costo total de propiedad (TCO) y las recomendaciones para su revisión.

La recopilación de datos suele tardar de dos a tres semanas cuando se analizan la utilización de la CPU, la utilización de la RAM, el rendimiento del almacenamiento, las IOPS y el rendimiento de la red. Lo ideal es que esta recopilación se lleve a cabo durante las horas de mayor actividad del mes laboral (por ejemplo, durante la presentación de informes end-of-month financieros). AWS quiere registrar los picos de uso porque esto proporciona buenos ejemplos estadísticos de cuál debería ser la AWS instancia del tamaño correcto y, al mismo tiempo, garantiza que el rendimiento pueda superar lo que está disponible en las instalaciones. AWS combina las métricas de uso con las heurísticas de rendimiento de varias generaciones de procesadores para determinar exactamente la cantidad de CPU y RAM que requiere una carga de trabajo determinada. Estos objetivos suelen

ser inferiores a los que se asignan en las instalaciones. Esto no solo reduce el costo de cómputo en función del tamaño de la instancia, sino que también optimiza los costos de licencia.

La siguiente vista del panel muestra un ejemplo de los costos de infraestructura que se pueden capturar mediante una evaluación.

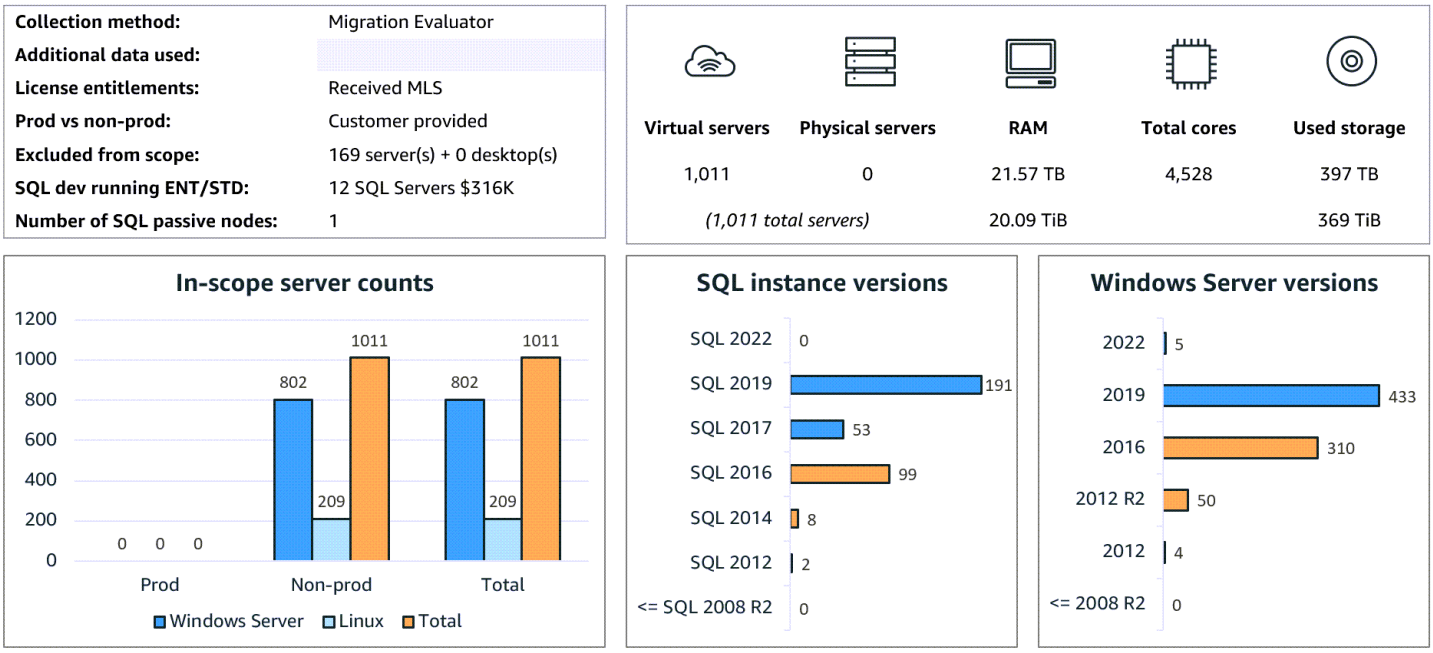


Análisis de datos

AWS ofrece una presentación resumida una vez finalizada la recopilación de datos. AWS revisa los datos, resume las conclusiones y, a continuación, hace recomendaciones para el uso local y la migración a la nube. Para reducir los costes informáticos y de licencias, examine las oportunidades de consolidación, el aumento de la elasticidad (ya que las cargas de trabajo pueden desactivarse o ajustarse según la temporada) y las oportunidades con los SKU adecuados (por ejemplo, se está utilizando la edición SQL Server Enterprise, pero los requisitos de recursos y el uso de funciones sugieren que la edición SQL Server Standard es adecuada). En el caso de productos como SQL Server, cuya licencia depende del núcleo, suele tener sentido desde el punto de vista económico colocar las cargas de trabajo en una instancia informática más cara. Es decir, si el perfil de la CPU y la relación entre la RAM y la vCPU tienen el efecto neto de reducir la cantidad de núcleos con

licencia tanto en los casos de uso con licencia incluida como en los de Bring Your Own License (BYOL).

A continuación se muestra un ejemplo de análisis basado en los datos recopilados por la evaluación.



Los escenarios de optimización más comunes incluyen la identificación tanto de las oportunidades AWS de optimización de los recursos como de los ahorros en licencias de terceros.

Ejemplos de oportunidades de optimización de AWS recursos:

- Evite el sobreaprovisionamiento para los picos de uso.
- Evite especificar en exceso o infrautilizar los recursos.
- Dimensione sus instancias de forma adecuada y migre a las generaciones más recientes de EC2 instancias.
- Ahorre en costos de operaciones al migrar a bases de datos administradas.

Ejemplos de ahorros en licencias de terceros:

- Reduzca los núcleos necesarios para ejecutar la misma carga de trabajo.
- Elimine la edición SQL Server Enterprise y los paquetes de complementos innecesarios.
- Elimine los servidores zombis y sustituya el hardware obsoleto.
- Utilice BYOL y las opciones con licencia incluida para reducir los futuros acuerdos comerciales.
- Modernícese con soluciones de código abierto y nativas de la nube.

Planifique los próximos pasos

Por último, AWS utiliza los datos de rendimiento recopilados para estimar el tamaño y el costo específicos de la carga de trabajo. AWS también puede analizar en conjunto su entorno específico y proporcionar un análisis cuantitativo. Esto puede ayudarle a determinar si la mejor opción es una actualización local o una migración a. AWS Para crear un modelo de negocio económico basado en la nube, utilice el resumen del análisis del TCO (como se muestra en el siguiente ejemplo) que se proporciona al final de una AWS OLA.

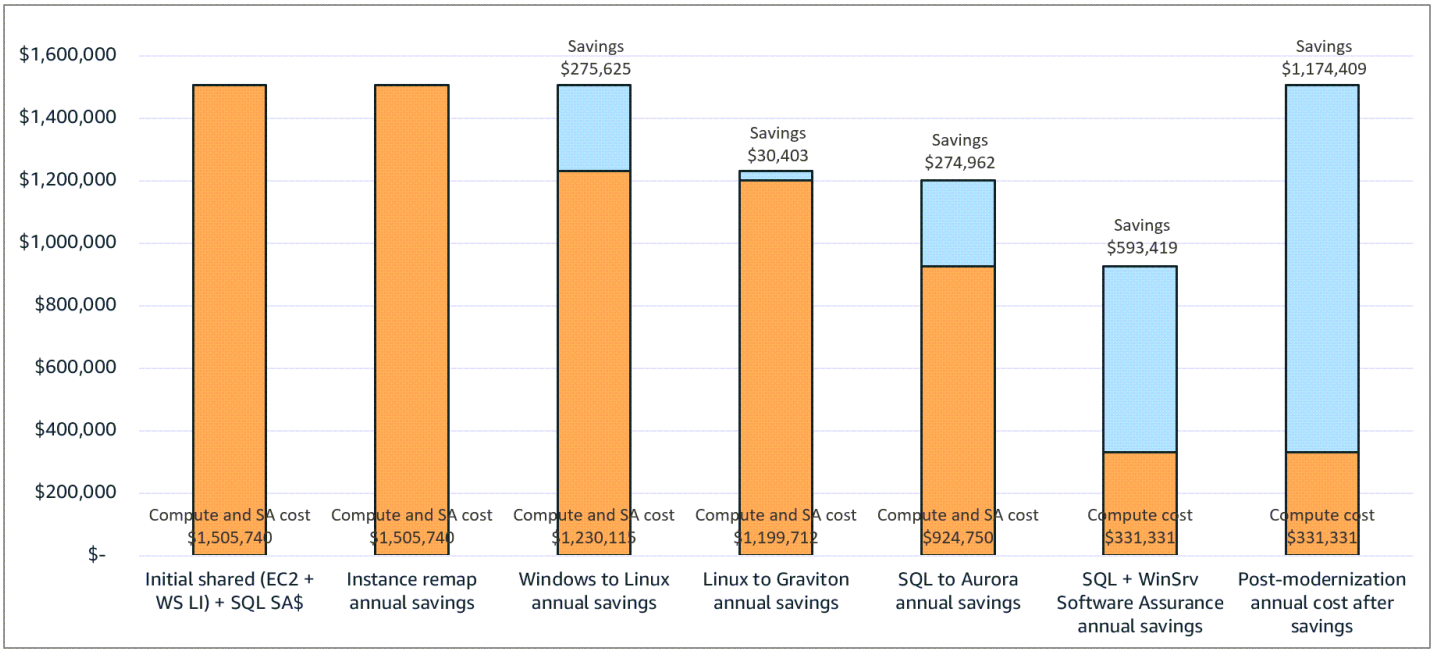
	Option 1: Amazon EC2 shared	**Option 1a: Amazon EC2 shared + power management**	**Option 2: Amazon EC2 mixed**	**Option 2a: Amazon EC2 mixed + power management**
Option details: compute	*100% Reserved Instances (RIs)*	*RIs + on-demand power management*	*100% Ris*	*RIs + on-demand power management*
Option details: Microsoft licenses	*WS LI and SQL BYOL*	*WS LI and SQL BYOL*	*WS BYOL or LI+SQL BYOL*	*WS BYOL or LI+SQL BYOL*
Compute costs[1]				
Year 1 compute cost	$414,546	$482,623	$504,019	$513,941
Year 1 vendor license included cost	$392,858	$244,415	$9,804	$4,783
	$807,404	**$727,038**	**$513,823**	**$518,724**
Total compute savings in year 1, compared to Option 1	—	*10% ($80,366)*	*36% ($293,581)*	*36% ($288,680)*
Storage and networking costs[2]				
Annual estimated storage cost	$336,494	$336,494	$336,494	$336,494
Annual estimated networking cost	$41,455	$41,455	$41,455	$41,455
	$377,949	**$377,949**	**$377,949**	**$377,949**
Microsoft license costs**				
WS/CIS annual Software Assurance (SA) + current SPLA/Subs cost	$0	$0	$0	$0
WS/CIS license + SA + SPLA/Subs true-up cost	$0	$0	$0	$0
SQL annual SA + current SPLA/Subs cost	$0	$0	$0	$0
SQL license SA + current SPLA/Subs true-up cost	$0	$0	$0	$0
	$0	**$0**	**$0**	**$0**
Total estimated costs	**$1,185,353**	**$1,104,987**	**$891,772**	**$896,673**
Annual TCO savings in year 1, compared to Option 1	—	*7% ($80,366)*	*25% ($293,581)*	*24% ($288,680)*

[1] Pricing model used: 3-year, no upfront RI
[2] Software Assurance and true-up costs provided by Microsoft

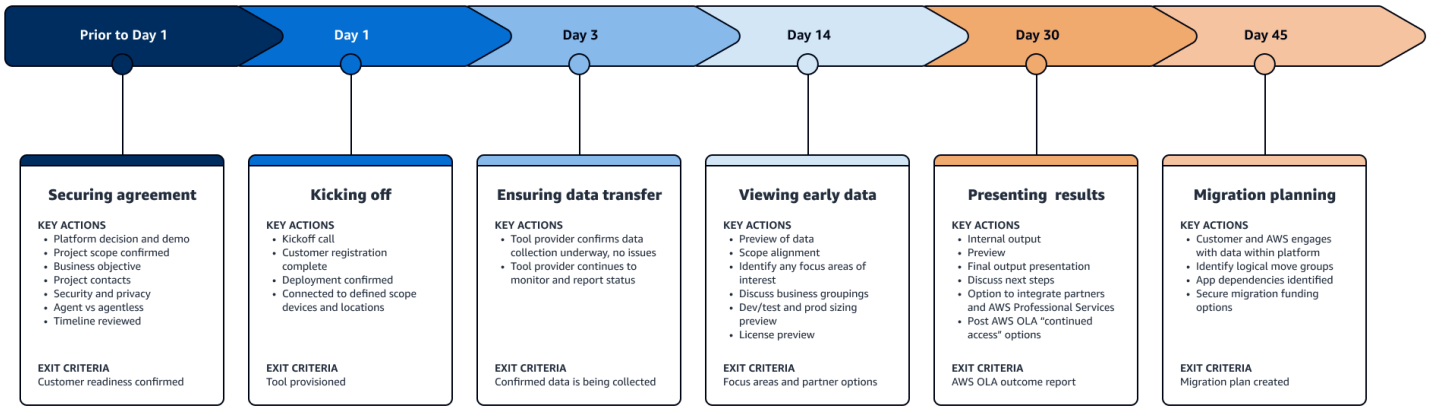
Un AWS OLA también proporciona información sobre el impacto que la modernización puede tener en sus cargas de trabajo actuales mediante sugerencias como las siguientes:

- Cambie a un sistema operativo Linux.
- Añada compatibilidad con aplicaciones para procesadores ARM (AWS Graviton).
- Mueva las cargas de trabajo de SQL Server a Amazon Aurora.
- Elimine la seguridad del software al migrar las cargas de trabajo de Windows y SQL Server a tecnologías de código abierto.

El siguiente diagrama muestra los ahorros de costos que se pueden lograr mediante técnicas de modernización, como pasar de Windows a Linux o de SQL Server a Aurora.



El proceso completo de AWS OLA tarda aproximadamente 45 días de principio a fin. El siguiente diagrama muestra un ejemplo de cronograma.



Si tiene un VMware entorno puro y desde el que puede proporcionar resultados RVTools, puede reducir este plazo a una semana laboral. Además, AWS puede analizar un archivo plano que incluye datos de activos y uso, como el promedio de la CPU, el pico de la CPU, el promedio de RAM y el pico de RAM.

Impacto de la evaluación

El cliente medio suele experimentar una reducción de costes del 20 al 30 por ciento si se intenta dimensionar correctamente. El tamaño correcto hace coincidir la carga de trabajo de origen con las AWS instancias de mejor tamaño en función de los datos de uso. Estos ajustes de tamaño correctos

no solo reducen el costo mensual del AWS entorno, sino que con frecuencia se traducen en ahorros en otros sectores de la organización. Por ejemplo, un aumento del 20 al 30 por ciento en las licencias de Windows o SQL Server puede reducir la próxima puesta a punto con Microsoft o liberar licencias para aplicaciones adicionales. line-of-business La consolidación y el dimensionamiento adecuado de las cargas de trabajo de SQL Server suelen ser las áreas en las que se obtienen las mayores ganancias financieras.

AWS puede ayudarlo a clasificar los sistemas en grupos de modernización. Algunos sistemas son antiguos y no son viables desde el punto de vista financiero, mientras que otros pueden modernizarse para convertirlos en contenedores o aplicaciones sin servidor, donde se obtienen los ahorros más significativos. La conversación con su AWS equipo pasa de temas generalizados sobre lo que permite la nube a debates más específicos sobre cómo y por qué se deben modernizar cargas de trabajo específicas. AWS también le ayuda a explorar posibles oportunidades de innovación.

Pasos a seguir a continuación

Si está empezando su proceso de optimización de costes para las cargas de trabajo de Microsoft que se ejecutan en entornos locales o en entornos locales AWS, póngase en contacto con su equipo de AWS cuentas y solicite una AWS OLA. AWS los miembros del equipo pueden responder a tus preguntas y ayudarte a decidir si un AWS OLA es, en última instancia, la mejor opción para ti y tu organización. Como alternativa, puede solicitar un AWS OLA en línea.

Recursos adicionales

- AWS Evaluación de la optimización y la concesión de licencias (AWS documentación)
- AWS re:Invent 2022: cómo ahorrar costes y optimizar las cargas de trabajo de Microsoft en AWS (ENT205) () YouTube

Windows en Amazon EC2

Amazon Elastic Compute Cloud (Amazon EC2) es una plataforma de computación en la nube altamente flexible y escalable que es ideal para ejecutar sus cargas de trabajo de Windows. Puede usar Amazon EC2 para implementar, administrar y escalar sus cargas de trabajo de Windows Server en la infraestructura segura, confiable, de alta disponibilidad y adaptable del Nube de AWS. Tenga en cuenta las siguientes ventajas clave de ejecutar cargas de trabajo de Windows en Amazon EC2:

- Escalabilidad: Amazon EC2 le permite escalar fácilmente sus cargas de trabajo de Windows para adaptarlas a los requisitos cambiantes. Puede crear nuevas EC2 instancias rápidamente para hacer frente al aumento de la demanda y cancelar las instancias con la misma facilidad cuando ya no las necesite. Solo paga por los recursos que realmente utiliza.
- Flexibilidad: Windows en Amazon EC2 admite una amplia gama de tipos de instancias que están diseñadas para satisfacer diversos requisitos de carga de trabajo, desde instancias de uso general hasta instancias optimizadas para memoria o procesamiento. Esta flexibilidad garantiza que pueda elegir el mejor tipo de instancia para sus aplicaciones específicas basadas en Windows, lo que maximiza el rendimiento y minimiza los costes.
- Seguridad: AWS proporciona varios niveles de seguridad para sus cargas de trabajo de Windows, incluidos los firewalls de red, el cifrado de datos y los controles de acceso seguros. Esto significa que puede confiar en que sus aplicaciones y datos están protegidos y, al mismo tiempo, tener el control total de sus ajustes y configuraciones de seguridad.
- Rentabilidad: el modelo de pay-as-you-go precios le permite pagar solo por los recursos que utilice, lo que elimina la necesidad de realizar inversiones iniciales en hardware y software. Este modelo también le permite optimizar sus costos, reducir los gastos de capital y aumentar la eficiencia operativa. Es un modelo de precios ideal para empresas de todos los tamaños.

En esta sección de la guía se tratan los siguientes temas:

- Automatice los horarios de parada e inicio
- Cargas de trabajo de Windows del tamaño correcto
- Seleccione el tipo de instancia adecuado para las cargas de trabajo de Windows
- Traiga licencias para cargas de trabajo de Windows y SQL Server
- Optimice el gasto para Windows en Amazon EC2
- Supervise los costes mediante herramientas AWS

Automatice los horarios de parada e inicio

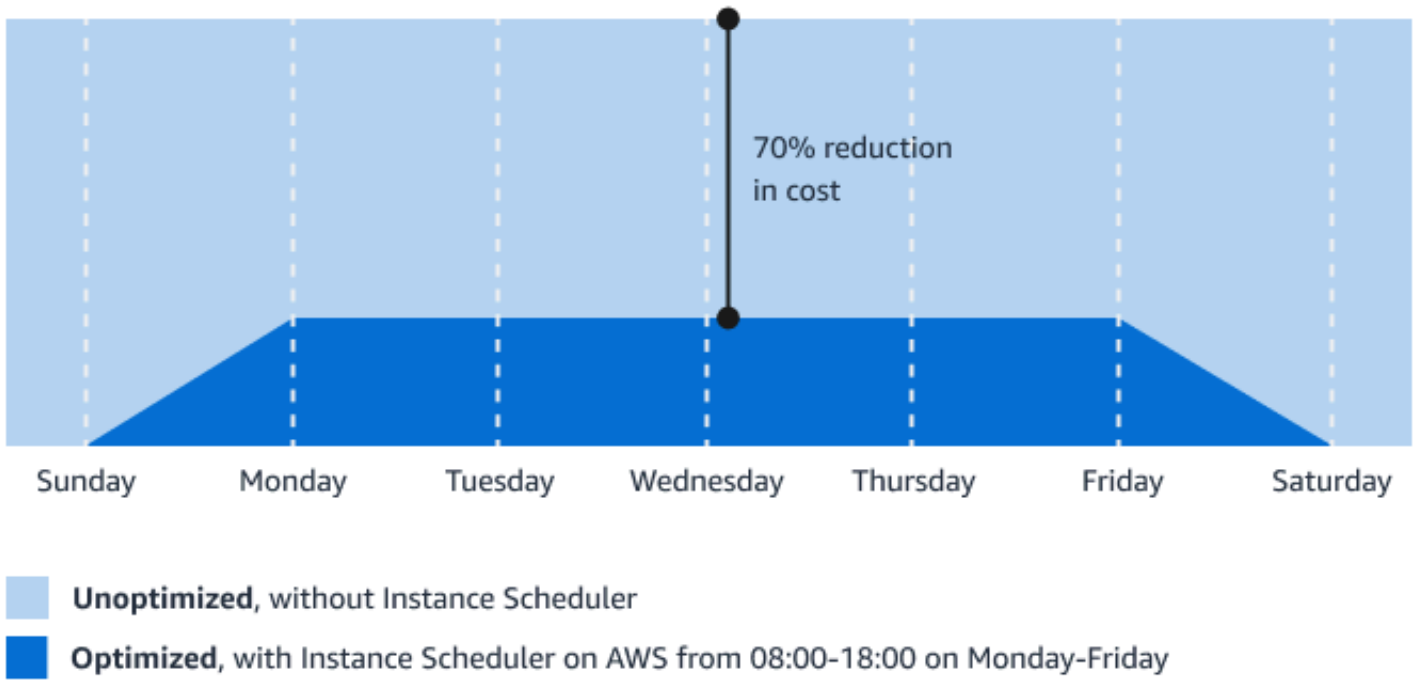
Descripción general

El programador de instancias activado AWS puede ayudarlo a reducir los costos operativos al automatizar el inicio y la detención de las instancias de Amazon EC2 y Amazon Relational Database Service (Amazon RDS). Si deja todas sus instancias ejecutándose a pleno rendimiento de forma continua, podría terminar pagando por los recursos que no se utilizan. El programador de instancias activado AWS le permite desactivar las instancias en momentos en que no se necesitan, como fuera del horario laboral, los fines de semana u otros períodos en los que el uso es bajo. Con el tiempo, esto puede suponer un importante ahorro de costes.

El programador de instancias AWS también ofrece la programación de instancias entre cuentas, el etiquetado automatizado y la posibilidad de configurar programas o períodos mediante una interfaz de línea de comandos o el período de mantenimiento. AWS Systems Manager Estas funciones pueden ayudarle a gestionar sus instancias de forma más eficaz y precisa, a realizar un seguimiento y a asignar los costes entre distintos proyectos o equipos.

Casos prácticos

Considere el ejemplo de una empresa que utiliza Instance Scheduler AWS en un entorno de producción para detener automáticamente las instancias fuera del horario laboral todos los días. Si esta empresa deja todas sus instancias funcionando a pleno rendimiento, puede ahorrar hasta un 70 por ciento en los costos de las instancias que solo son necesarias durante el horario laboral habitual. En el siguiente gráfico, se muestra cómo se reduce la utilización semanal de 168 a 50 horas.

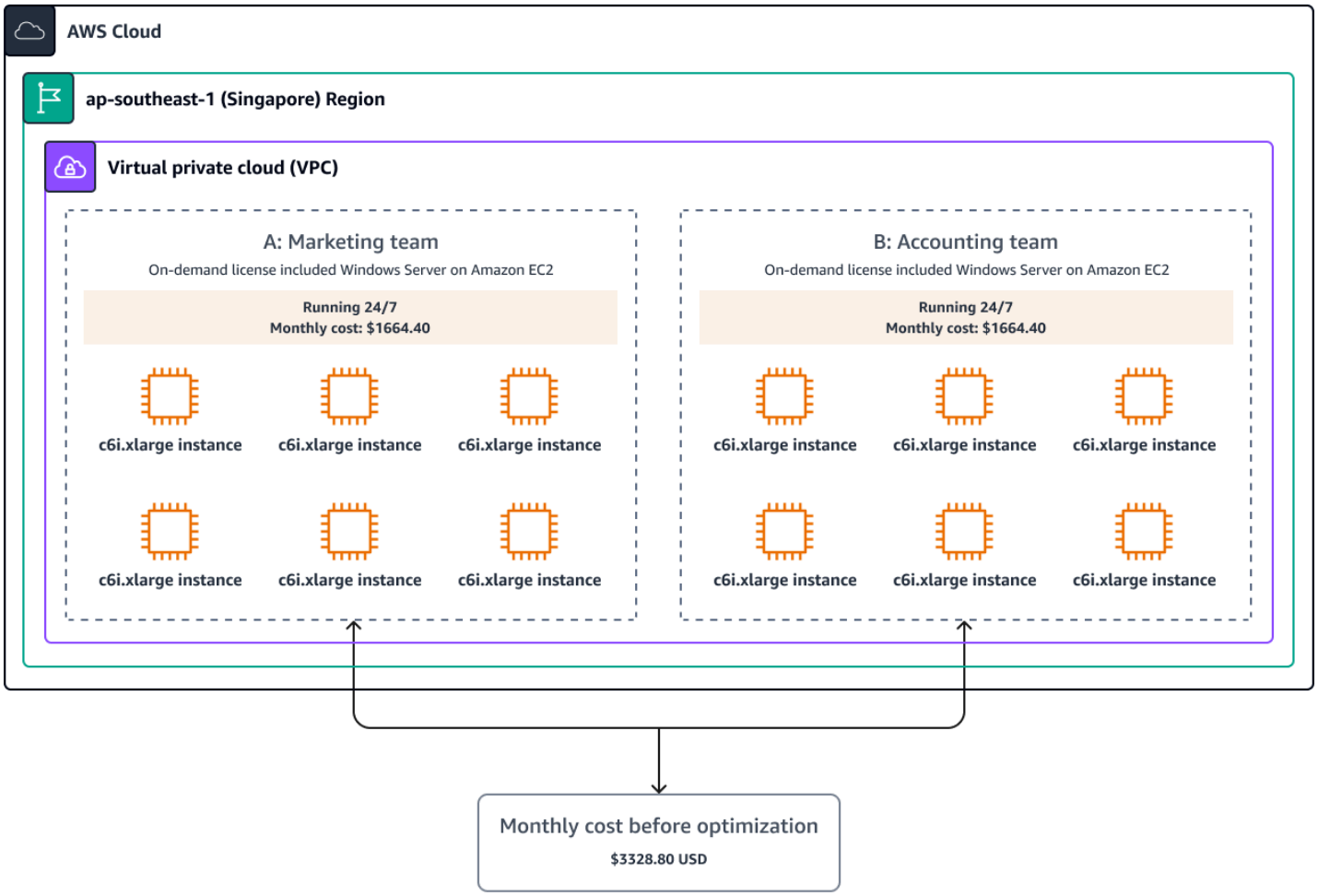


Considere otro ejemplo. La empresa eléctrica Jamaica Public Service Company Limited (JPS) migró su base de datos a Amazon RDS. Ahora, JPS usa Amazon EC2 para alojar servicios de API y ejecutar otras aplicaciones. Para JPS, Instance Scheduler on AWS se convirtió en la herramienta clave para gestionar entornos no productivos. JPS utilizó el programador de instancias AWS para reducir los costos de desarrollo y administrar las EC2 instancias en función de las necesidades del equipo y los horarios de trabajo. Esto ayudó a JPS a reducir los costos en un 40 por ciento. Para obtener más información, consulte el AWS caso práctico El servicio público de Jamaica migra eficientemente a la nube y reduce los costos en un 40% con AWS Instance Scheduler.

Escenario de optimización de costos

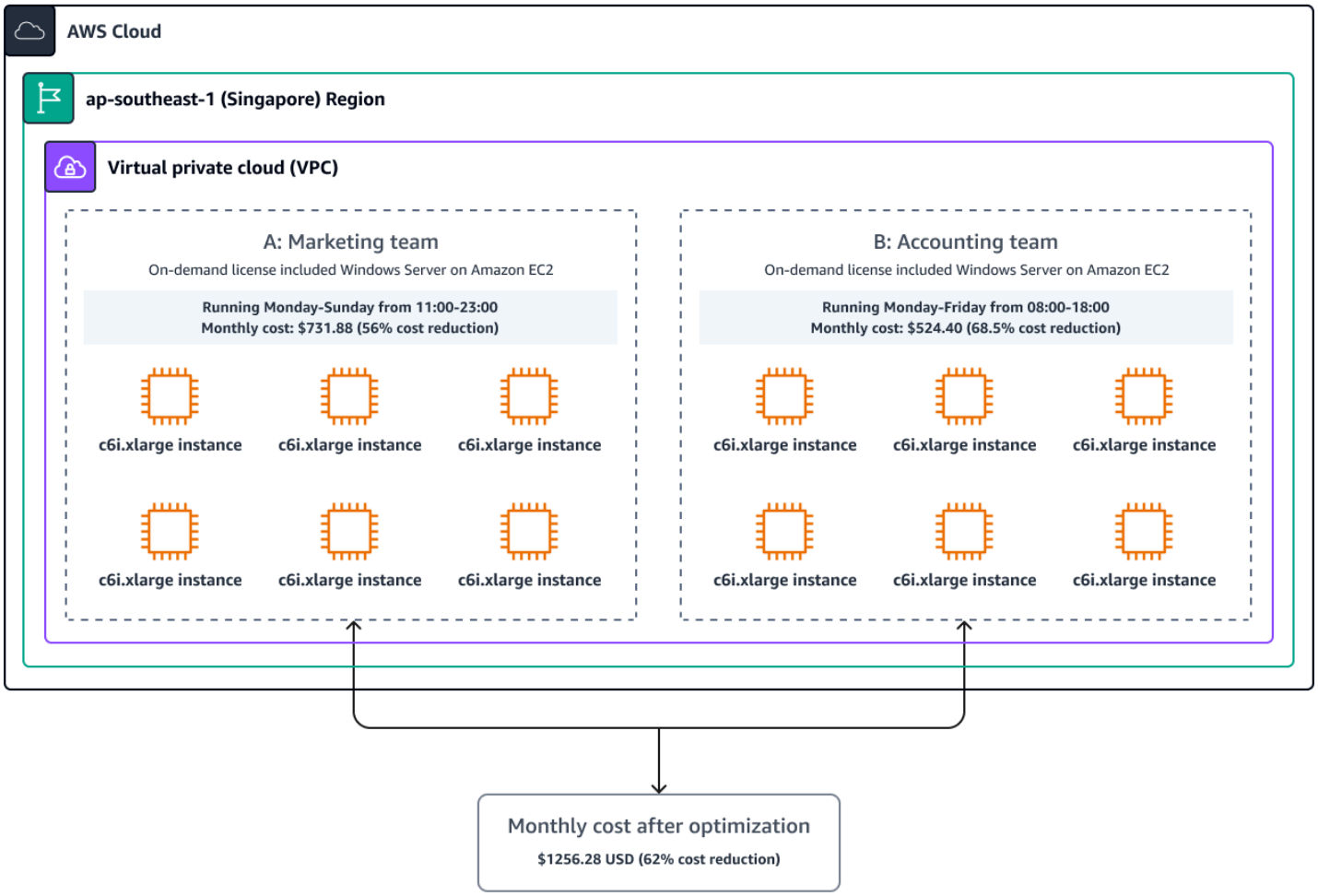
El siguiente escenario de ejemplo ayuda a ilustrar las ventajas económicas de utilizar Instance Scheduler en AWS. En este escenario, una importante empresa minorista de Singapur despliega dos entornos Windows en Amazon EC2. El equipo de marketing utiliza el primer entorno, conocido como carga de trabajo A, para analizar en tiempo real las transacciones en las tiendas mientras las tiendas están abiertas. El segundo entorno, conocido como carga de trabajo B, está reservado para el equipo de contabilidad, que solo trabaja durante el horario laboral habitual. El cronograma operativo actual de ambos entornos (24 horas al día, 7 días a la semana) no es ideal dados los patrones de uso actuales y requiere una optimización para reducir los costos operativos de la empresa.

El siguiente diagrama muestra el costo mensual antes de la optimización.



Por ejemplo, hay 31 días en el mes de marzo, de los cuales 23 son días laborables. Si el equipo de marketing usa el programador de instancias AWS activado y opera sus instancias solo cuando es necesario (es decir, 321 horas al mes en lugar de 730 horas al mes), podría ahorrarse 932,52$ al mes. Esto equivale a una reducción del 56 por ciento en los costos operativos. El equipo de contabilidad también puede disfrutar de importantes ventajas, ya que el tiempo de uso de las instancias se reduce de 730 horas al mes a 230 horas. Esto se traduce en una reducción de 1140$, o un 68,5 por ciento. La empresa podría ahorrar un total combinado de 2.072,52 dólares al mes (lo que equivale a una reducción del 62 por ciento), o 24.870,24 dólares al año.

El siguiente diagrama muestra el coste mensual tras la optimización.



> **Note**
>
> El precio de este ejemplo se determinó utilizando el Calculadora de precios de AWSde marzo de 2023.

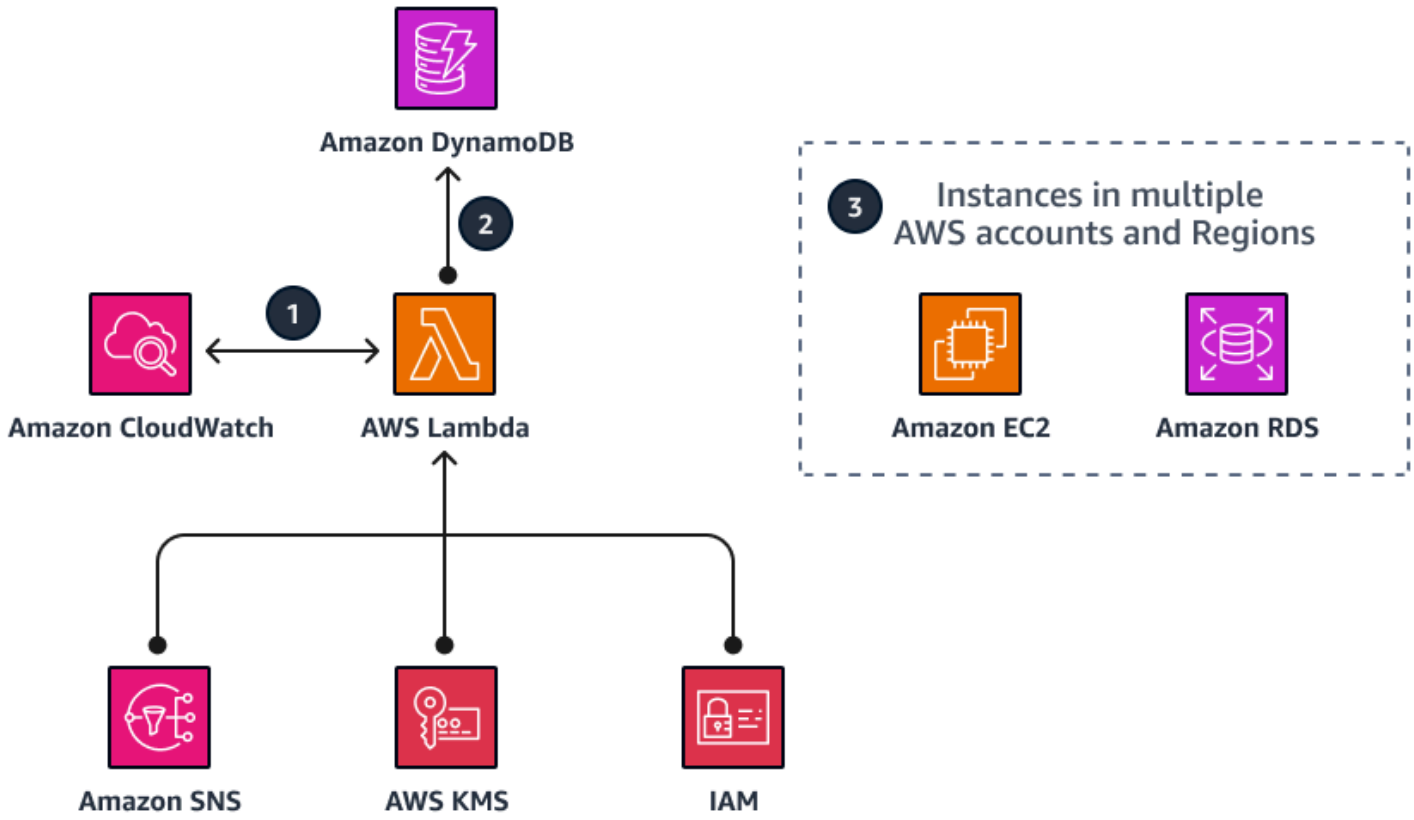
Recomendaciones de optimización de costos

En esta sección se explica cómo implementar y configurar el programador de instancias en AWS función del escenario de ejemplo descrito en la sección anterior sobre el escenario de optimización de costos. Le recomendamos que siga los siguientes pasos para optimizar sus costes mediante el uso del programador de instancias en: AWS

1. Lance la pila Instance Scheduler
2. Configure los períodos

3. Configure los horarios
4. Etiqueta de instancias .

El siguiente diagrama de arquitectura muestra lo que crea en la Nube de AWS pila Instance Scheduler.



El diagrama muestra los siguientes pasos del flujo de trabajo:

1. Una AWS CloudFormation plantilla configura un CloudWatch evento de Amazon en un intervalo que tú definas. Este evento invoca una AWS Lambda función. Durante la configuración, se definen las cuentas Regiones de AWS y. También debe definir una etiqueta personalizada que Instance Scheduler AWS utiliza para asociar los cronogramas con las instancias de Amazon, EC2 las instancias de Amazon RDS y los clústeres aplicables.
2. Los valores de configuración de la programación se almacenan en Amazon DynamoDB y la función Lambda los recupera cada vez que se ejecuta. A continuación, puede aplicar la etiqueta personalizada a las instancias correspondientes.
3. Durante la configuración inicial del programador de instancias, debe definir una clave de etiqueta para identificar las instancias de Amazon EC2 y Amazon RDS aplicables. Al crear un programa,

el nombre que especifique se utiliza como valor de etiqueta que identifica el programa que desea aplicar al recurso etiquetado.

Lanza la pila del programador de instancias

En esta sección, se muestra cómo lanzar la CloudFormation pila del programador de instancias en. AWS

> **Note**
>
> Eres responsable del coste de lo que Servicios de AWS se utilice mientras se ejecuta Instance Scheduler. AWS A partir de enero de 2023, el costo de ejecutar esta solución con la configuración predeterminada en la `us-east-1` región es de aproximadamente 9,90 USD al mes en el caso de los cargos de Lambda, o menos si tiene un crédito de uso mensual de la capa gratuita de Lambda. Para obtener más información, consulte la sección sobre costos de la guía de AWS implementación del programador de instancias en la biblioteca de soluciones. AWS

Para lanzar la pila de programadores de instancias, sigue estos pasos.

1. Inicia sesión AWS Management Consoley selecciona Launch solution (plantilla descargable) para lanzar la `instance-scheduler-on-aws.template` CloudFormation plantilla.

   > **Note**
   >
   > También puede descargar la plantilla para usarla como punto de partida para su propia implementación.

2. La plantilla se lanza en la región Este de EE. UU. (Norte de Virginia) de forma predeterminada. Para lanzar Instance Scheduler en una región diferente, utilice el selector de regiones de la barra de navegación de la consola.

   > **Note**
   >
   > En este ejemplo se utiliza la región Asia Pacífico (Singapur).

3. En la página Crear pila, en la sección Requisitos previos: Preparar plantilla, compruebe que esté seleccionada la opción La plantilla está lista. En la sección Fuente de la plantilla, compruebe que esté seleccionada la opción URL de Amazon S3.
4. Compruebe que la URL de la plantilla correcta esté en el cuadro de texto URL de Amazon S3 y, a continuación, seleccione Siguiente.
5. En la página Especificar los detalles de la pila, especifique un nombre para la pila. Para obtener información sobre las limitaciones de nombres de caracteres, consulte los límites de IAM y STS en la AWS Identity and Access Management documentación (IAM). El nombre de la pila para el ejemplo de esta guía se llama. `MyInstanceScheduler`

   > **Note**
   >
   > El nombre de la pila no puede contener más de 28 caracteres.

6. En Parámetros, revise los parámetros de la plantilla y modifíquelos según sea necesario.
7. Elija Next (Siguiente). En la página Configurar opciones de pila, elija Siguiente.
8. En la página de revisión, revise y confirme la configuración. Seleccione la casilla para confirmar que la plantilla creará recursos de IAM.
9. Elija Crear para implementar la pila.

Configure los períodos

Tras implementar la CloudFormation plantilla, la solución crea una tabla de DynamoDB que contiene ejemplos de reglas y programas de períodos que puede utilizar como referencia para crear sus propias reglas y programas de períodos personalizados. Para ver un ejemplo de configuración de períodos, consulte los ejemplos de cronogramas en el programador de instancias en la documentación. AWS

Para completar el paso de este escenario, debe generar períodos que correspondan a cada carga de trabajo y que satisfagan sus necesidades específicas. Por ejemplo:

```
Period 1 (Workload A):
    Name: retail-hours
    Days: Monday to Sunday
    Hours: 1100 - 2300
Period 2 (Workload B):
    Name: office-hours
    Days: Monday to Friday
```

```
    Hours: 0800 - 1800
```

Para configurar los períodos, complete los siguientes pasos:

1. Inicie sesión en la consola de DynamoDB y asegúrese de estar en la misma región en la que lanzó la plantilla para CloudFormation el programador de instancias. AWS
2. En el panel de navegación, elija Tablas y, a continuación, seleccione la tabla denominada. ConfigTable
3. Elija Explorar los elementos de la tabla.
4. Para crear un período para el horario de oficina, seleccione un período para el elemento de horario de oficina.
5. En la página Editar artículo, cambie el valor de la hora de inicio a 0800 y la hora de finalización a 1800. Deje el valor predeterminado para los días de la semana.

> **Note**
>
> Los valores de hora de inicio y finalización determinan cuándo se deben iniciar y detener las instancias, mientras que el valor de los días de la semana determina a qué días de la semana se aplica esta programación (de lunes a viernes, en este ejemplo).

6. Elija Guardar cambios.
7. Para duplicar el período de horario de oficina y usarlo para crear un nuevo período para el horario comercial, seleccione el período para el elemento de horario de oficina. A continuación, en el menú Acciones, selecciona Duplicar elemento.
8. Modifique los atributos para que se ajusten a sus necesidades. Los siguientes atributos se utilizan para cumplir los requisitos del escenario de ejemplo:

```
type: period
name: retail-hours
begintime: 11:00
description: Retail hours
endtime: 23:00
weekdays: mon-sun
```

9. Seleccione Crear elemento.
10. En ConfigTableDynamoDB, identifique los dos períodos que acaba de crear que aparecen en las listas de elementos.

Configure los horarios

En el contexto de Instance Scheduler on AWS, los cronogramas se refieren a la aplicación de uno o más períodos y a la zona horaria correspondiente. Luego, estos cronogramas se asignan a sus instancias como etiquetas. En esta sección, se muestra cómo crear dos programas (que se muestran a continuación) para adaptarse a los diferentes patrones de tiempo de las dos cargas de trabajo de ejemplo y, a continuación, asociar los programas a los períodos que creó en la sección anterior.

```
Schedule 1:
    Name: singapore-office-hours
    Period: office-hours
    Timezone: Asia/Singapore
Schedule 2:
    Name: singapore-retail-hours
    Period: retail-hours
    Timezone: Asia/Singapore
```

Para crear y configurar los horarios, complete los siguientes pasos:

1. Inicie sesión en la consola de DynamoDB y asegúrese de estar en la misma región en la que lanzó la plantilla para CloudFormation el programador de instancias. AWS
2. En el panel de navegación, elija Tablas y, a continuación, seleccione la tabla denominada. ConfigTable
3. Elija Explorar los elementos de la tabla.
4. Para duplicar el horario de oficina del Reino Unido y usarlo para crear un nuevo horario para su horario de oficina (horario de oficina de Singapur, por ejemplo), seleccione horario para uk-office-hours el elemento. A continuación, en el menú Acciones, selecciona Duplicar elemento.
5. Modifique los atributos para que se ajusten a sus necesidades. Los siguientes atributos se utilizan para cumplir los requisitos del escenario de ejemplo:

   ```
   type: schedule
   name: singapore-office-hours
   description: Office hours in Singapore
   periods: office-hours
   timezone: Asia/Singapore
   ```

6. Seleccione Crear elemento.
7. Repita los pasos 4 a 6 para crear un horario para el horario comercial de Singapur con los siguientes valores de atributo:

```
type: schedule
name: singapore-retail-hours
description: Retail hours in Singapore
periods: retail-hours
timezone: Asia/Singapore
```

8. En ConfigTableDynamoDB, identifique los dos programas y los dos períodos que ha creado.

Etiqueta de instancias .

Después de establecer las programaciones, debe usar etiquetas para asignar las programaciones a las instancias específicas que desee usar. Puedes usar el editor de etiquetas AWS Resource Groupsque contiene para generar y asignar etiquetas a tus EC2 instancias de Amazon.

1. Inicia sesión en AWS Management Consoley asegúrate de estar en la misma región en la que lanzaste la CloudFormation plantilla anteriormente.
2. Abra la consola Resource Groups. En el panel de navegación, expanda Etiquetado y, a continuación, elija Editor de etiquetas.
3. En la sección Buscar recursos para etiquetar, en el caso de las regiones, selecciona tus regiones. Para los tipos de recursos, elija Amazon EC2 o Amazon RDS. Este escenario se centra en las EC2 instancias de Amazon en la carga de trabajo A. El equipo de marketing utiliza la carga de trabajo A en la región de Singapur. Los recursos para esta carga de trabajo ya están etiquetados con una clave de departamento y un valor de marketing. Puede usar esta etiqueta para buscar las instancias.
4. Seleccione Buscar recursos.
5. Seleccione las instancias que desee incluir en la programación de la lista de resultados de la búsqueda y, a continuación, elija Administrar las etiquetas de los recursos seleccionados.
6. En la sección Editar etiquetas de todos los recursos seleccionados, elija Añadir etiqueta para añadir las etiquetas de programación del programador de instancias a sus EC2 instancias. Puede usar las claves y los valores de las etiquetas que coincidan con la tabla de planificación (creada anteriormente en DynamoDB).
7. En Tag key, añada Schedule. En Valor de etiqueta, introduzca singapore-retail-hours.
8. Seleccione Revisar y aplicar cambios en la etiqueta.
9. Para aplicar la etiqueta a todas las EC2 instancias que ha seleccionado, elija Aplicar los cambios a todas las seleccionadas.

10. Repita los pasos 3 a 9 para cualquier programa adicional que desee aplicar.

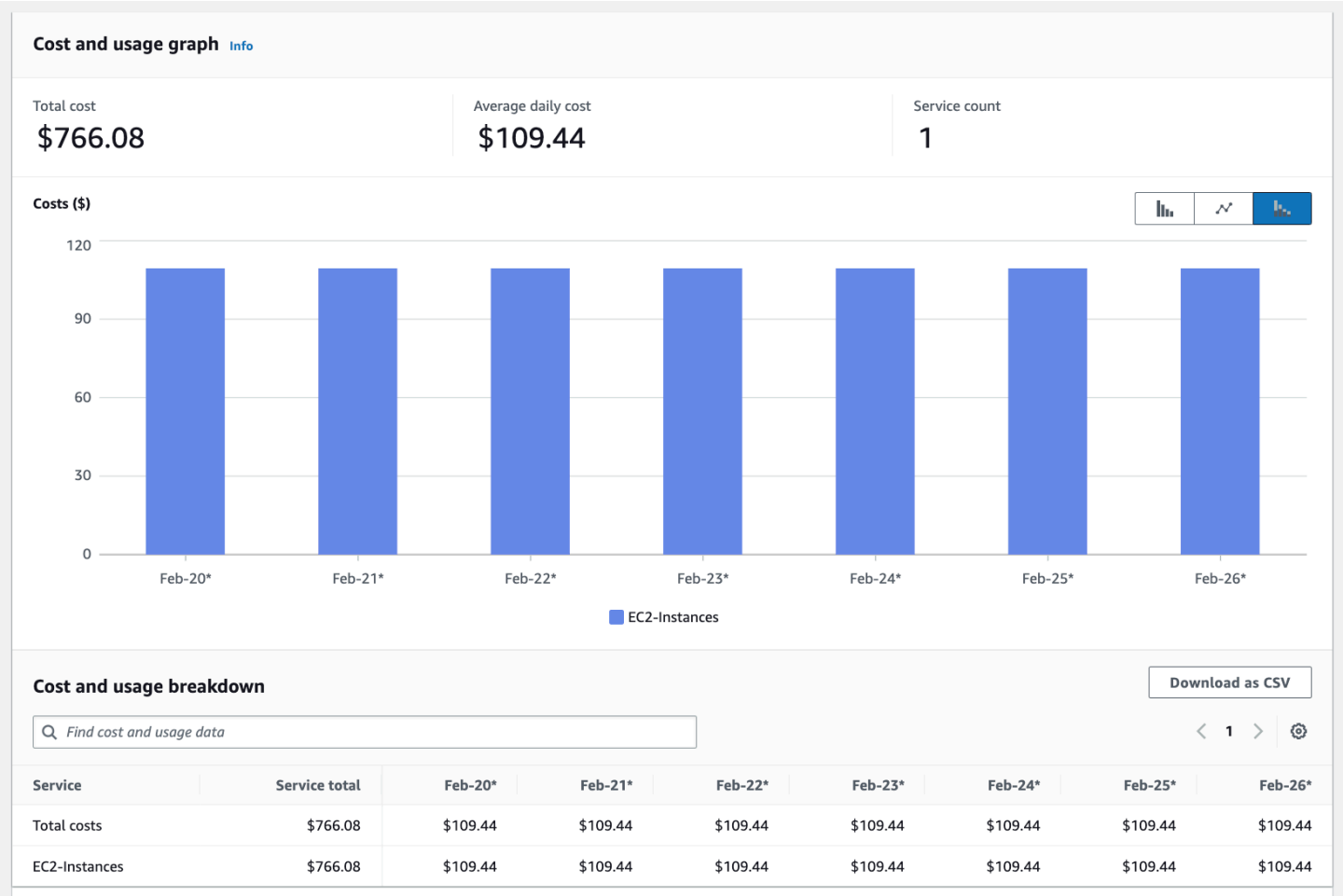
Validar los resultados

Le recomendamos que utilice Instance Scheduler AWS Cost Explorer para medir los costos y beneficios de usar Instance Scheduler en Internet. AWS Puede usar Cost Explorer para hacer lo siguiente:

- Vea y analice los costos asociados a sus EC2 instancias, incluidas las instancias administradas por Instance Scheduler.
- Filtre la vista de Cost Explorer por etiquetas para poder centrarse en cargas de trabajo específicas y obtener una visión pormenorizada de los ahorros de costes que se obtienen al utilizar Instance Scheduler.
- Obtenga información sobre el impacto financiero del uso de Instance Scheduler.
- Identifique oportunidades para una mayor optimización de los costos y tome decisiones basadas en datos para optimizar sus AWS gastos.

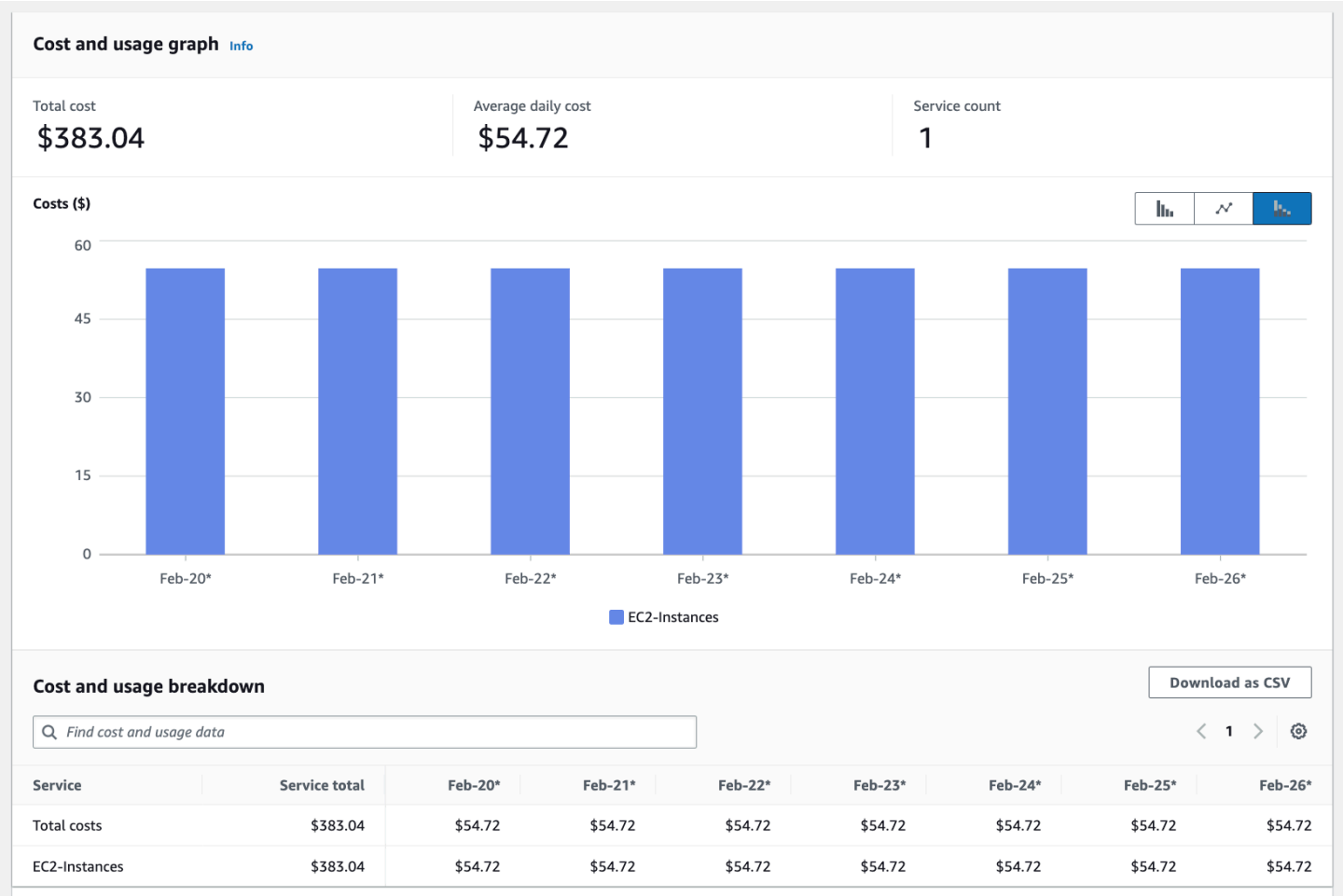
Los siguientes gráficos ilustran el costo de operar las cargas de trabajo A y B durante un período de siete días (de lunes a domingo) antes de la optimización mediante Instance Scheduler.

Gastos totales combinados de las cargas de trabajo A y B



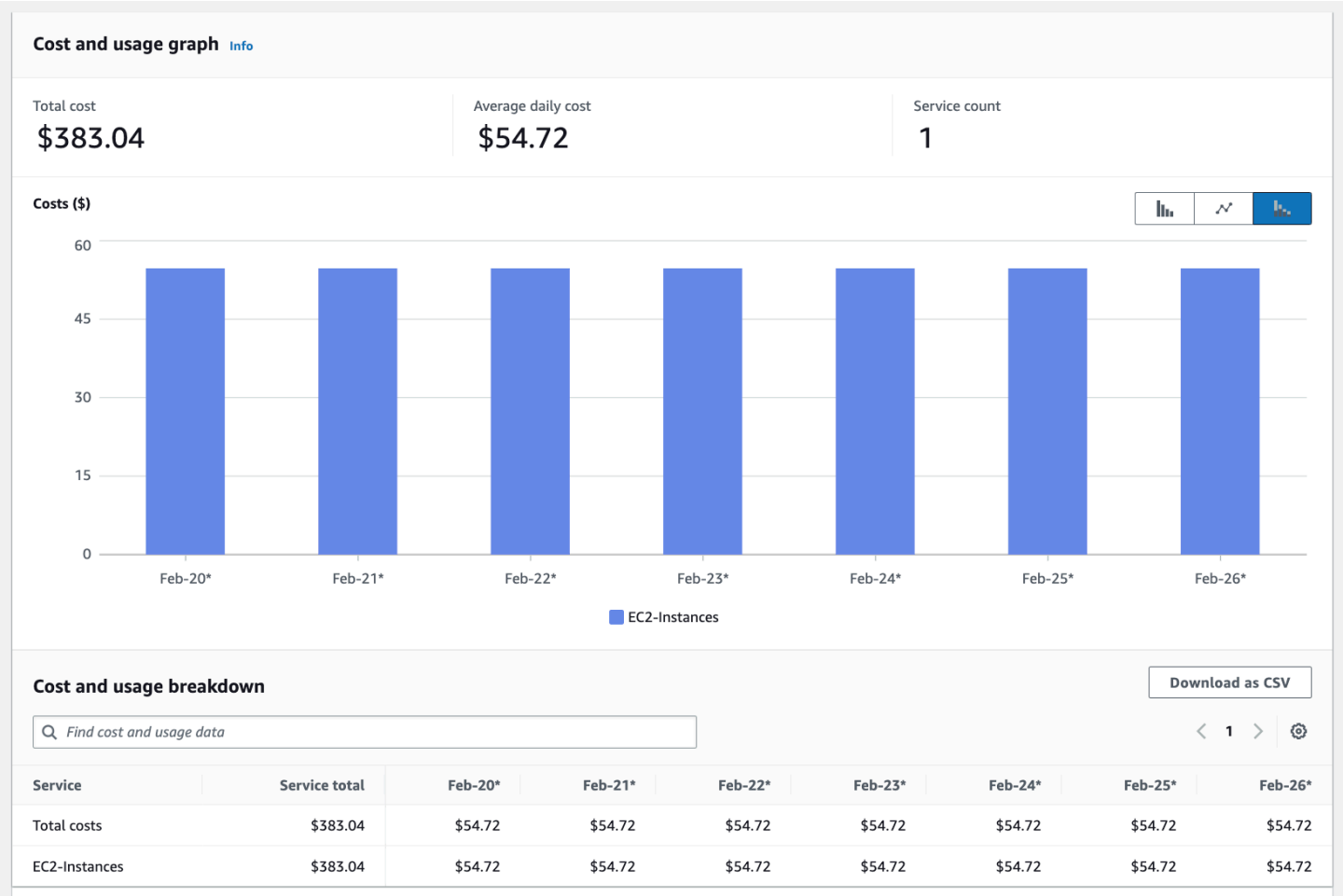
Service	Service total	Feb-20*	Feb-21*	Feb-22*	Feb-23*	Feb-24*	Feb-25*	Feb-26*
Total costs	$766.08	$109.44	$109.44	$109.44	$109.44	$109.44	$109.44	$109.44
EC2-Instances	$766.08	$109.44	$109.44	$109.44	$109.44	$109.44	$109.44	$109.44

Gastos de carga de trabajo A



Service	Service total	Feb-20*	Feb-21*	Feb-22*	Feb-23*	Feb-24*	Feb-25*	Feb-26*
Total costs	$383.04	$54.72	$54.72	$54.72	$54.72	$54.72	$54.72	$54.72
EC2-Instances	$383.04	$54.72	$54.72	$54.72	$54.72	$54.72	$54.72	$54.72

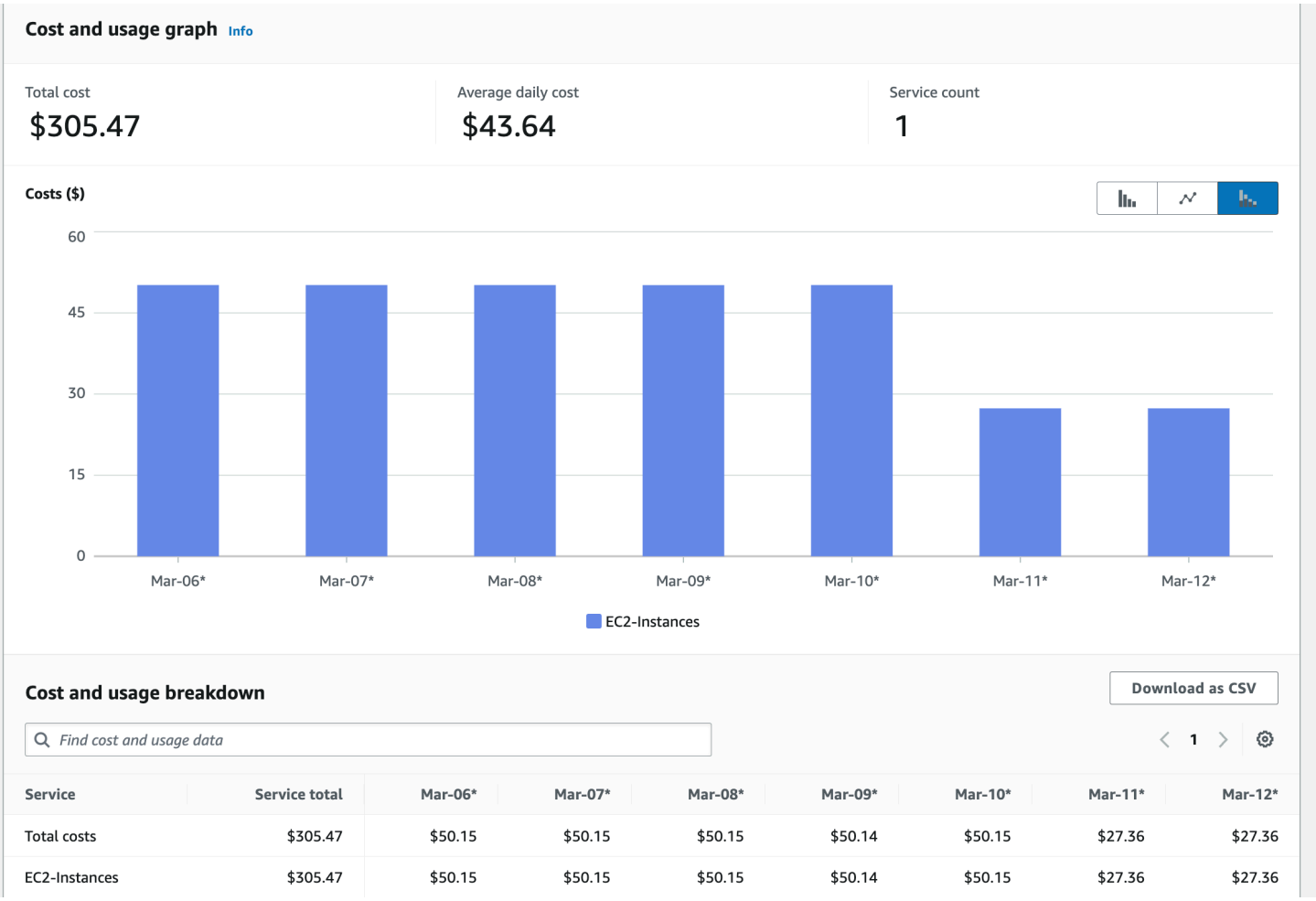
Gastos de carga de trabajo B



Service	Service total	Feb-20*	Feb-21*	Feb-22*	Feb-23*	Feb-24*	Feb-25*	Feb-26*
Total costs	$383.04	$54.72	$54.72	$54.72	$54.72	$54.72	$54.72	$54.72
EC2-Instances	$383.04	$54.72	$54.72	$54.72	$54.72	$54.72	$54.72	$54.72

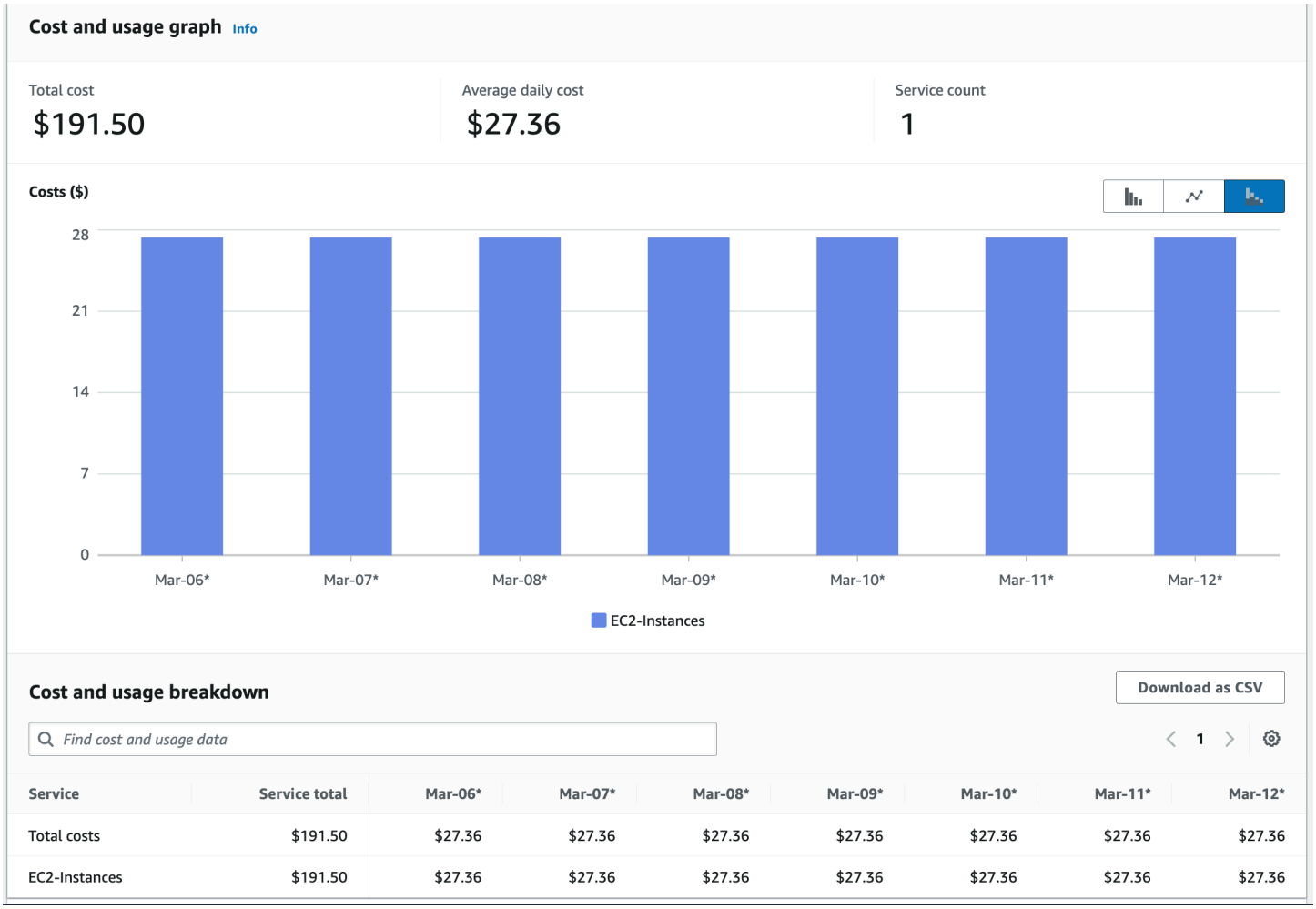
En este escenario, Cost Explorer muestra las reducciones de costos que se obtienen al implementar Instance Scheduler en AWS. Los siguientes gráficos muestran los costos operativos de las cargas de trabajo A y B durante un período de siete días (de lunes a domingo) después de la optimización.

Gastos totales combinados de las cargas de trabajo A y B



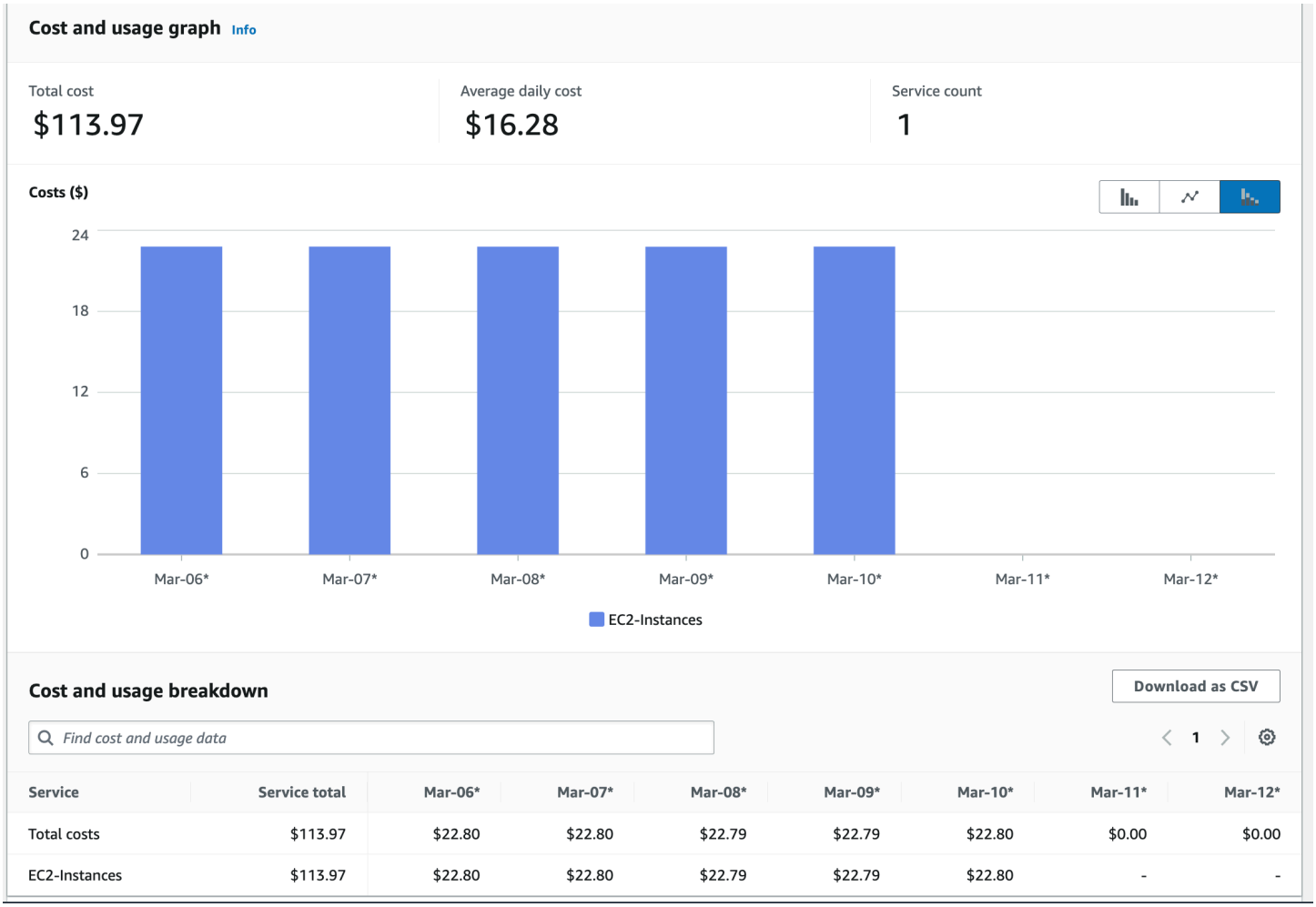
Service	Service total	Mar-06*	Mar-07*	Mar-08*	Mar-09*	Mar-10*	Mar-11*	Mar-12*
Total costs	$305.47	$50.15	$50.15	$50.15	$50.14	$50.15	$27.36	$27.36
EC2-Instances	$305.47	$50.15	$50.15	$50.15	$50.14	$50.15	$27.36	$27.36

Gastos de carga de trabajo A



Service	Service total	Mar-06*	Mar-07*	Mar-08*	Mar-09*	Mar-10*	Mar-11*	Mar-12*
Total costs	$191.50	$27.36	$27.36	$27.36	$27.36	$27.36	$27.36	$27.36
EC2-Instances	$191.50	$27.36	$27.36	$27.36	$27.36	$27.36	$27.36	$27.36

Gastos de carga de trabajo B



Service	Service total	Mar-06*	Mar-07*	Mar-08*	Mar-09*	Mar-10*	Mar-11*	Mar-12*
Total costs	$113.97	$22.80	$22.80	$22.79	$22.79	$22.80	$0.00	$0.00
EC2-Instances	$113.97	$22.80	$22.80	$22.79	$22.79	$22.80	-	-

Recursos adicionales

- Automatice el inicio y la detención de AWS instancias (véase el programador de instancias en la AWS documentación)
- Volver a lo básico: uso de un programador de instancias para controlar los costos de recursos de Amazon EC2 y Amazon RDS () YouTube
- Etiquetar sus AWS recursos (Guía del usuario sobre el etiquetado AWS de recursos)
- Analizar sus costos con AWS Cost Explorer(documentación)Administración de facturación y costos de AWS

Cargas de trabajo de Windows del tamaño correcto

Descripción general

El dimensionamiento correcto es una de las herramientas de ahorro de costes más potentes. AWS ofrece varios métodos para recopilar información sobre el tamaño correcto, desde revisar las posibles cargas de trabajo mediante una evaluación de AWS optimización y licencias (AWS OLA) hasta revisar las cargas de trabajo existentes mediante el uso de. AWS Cost Explorer

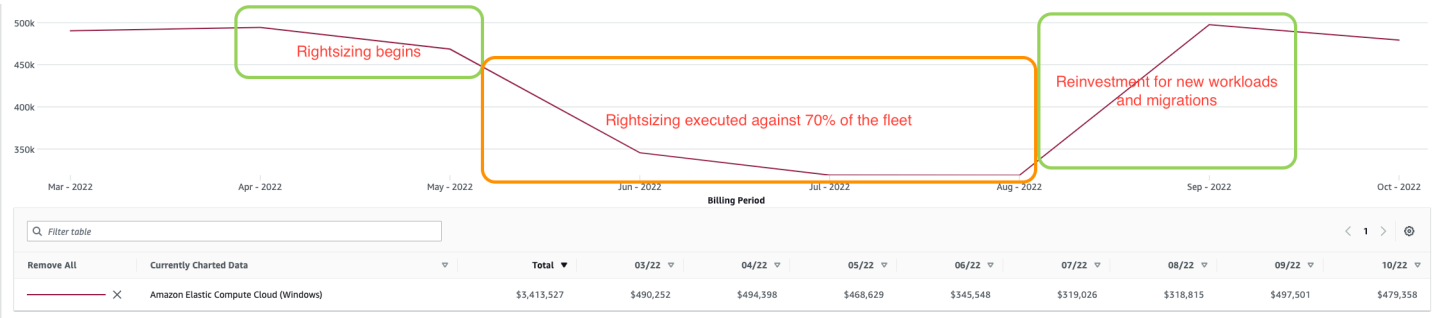
En esta sección se muestra cómo utilizar Amazon AWS Compute Optimizerpara identificar las oportunidades de talla EC2 adecuada. Compute Optimizer ayuda a evitar el sobreaprovisionamiento y el subaprovisionamiento de los siguientes tipos de recursos: AWS

- Tipos de instancias de Amazon Elastic Compute Cloud (Amazon EC2)
- Volúmenes de Amazon Elastic Block Store (Amazon EBS)
- Servicios de Amazon Elastic Container Service (Amazon ECS) en AWS Fargate
- AWS Lambdafunciones basadas en los datos de uso proporcionados por Amazon CloudWatch

Escenario de optimización de costes

Medir la eficacia del tamaño correcto puede ser un desafío, ya que los esfuerzos de dimensionamiento correcto pueden dirigirse a una aplicación específica, a un equipo o a toda la organización. Por ejemplo, pensemos en una organización que migra varios miles de instancias a cargas de trabajo de Windows AWS, donde el 90 por ciento de su flota está compuesta por cargas de trabajo de Windows. La organización puede emplear Compute Optimizer para analizar su flota y descubrir un exceso significativo de aprovisionamiento en sus cuentas y. Regiones de AWS Luego, pueden usar la AWS Systems Manager automatización para ajustar el tamaño de su flota a través de varios períodos de mantenimiento. Como resultado, la organización logra ajustar el tipo de instancia del tamaño correcto para el 70 por ciento de su flota y logra un ahorro de costos del 35 por ciento.

El siguiente panel ilustra los ahorros logrados durante varios meses, ya que esta organización de ejemplo implementó estratégicamente las recomendaciones de tamaño correctas de Compute Optimizer. Su objetivo era operar sus cargas de trabajo existentes de la manera más eficiente posible para reanudar una migración estancada desde un centro de datos de ubicación cercana a la finalización de su contrato.



Recomendaciones de optimización de costos

Te recomendamos que sigas los siguientes pasos para optimizar tus costos mediante Compute Optimizer:

- Activar Compute Optimizer
- Habilite la recopilación de métricas de memoria para los nodos de Windows
- Recomendaciones de Consume Compute Optimizer
- Etiquete las instancias para obtener el tamaño correcto
- Habilite la etiqueta de asignación de costos para que funcione con las herramientas AWS de facturación
- Implemente las recomendaciones de tamaño adecuadas con Automation AWS Systems Manager
- Considere métodos alternativos de redimensionamiento
- Revise los costes anteriores y posteriores en Cost Explorer

Activar Compute Optimizer

Puedes habilitar Compute Optimizer a nivel de organización o de cuenta única en. AWS Organizations La configuración de toda la organización proporciona informes continuos sobre las instancias nuevas y existentes en toda la flota y para todas las cuentas de los miembros. Esto permite que el dimensionamiento correcto sea una actividad recurrente en lugar de una actividad. point-in-time

Nivel de organización

Para la mayoría de las organizaciones, la forma más eficiente de usar Compute Optimizer es a nivel de la organización. Esto proporciona una visibilidad de la organización desde múltiples cuentas y regiones y centraliza los datos en una sola fuente para su revisión. Para habilitar esto a nivel de la organización, haga lo siguiente:

1. Inicie sesión en su cuenta de administración de Organizations con un rol que tenga los permisos necesarios y elija habilitar todas las cuentas de esta organización. Su organización debe tener habilitadas todas las características.
2. Una vez habilitada la cuenta de administración, puede iniciar sesión en ella, ver las demás cuentas de los miembros y consultar sus recomendaciones.

> **Note**
>
> Se recomienda configurar una cuenta de administrador delegado para Compute Optimizer. Esto te permite aplicar el principio del privilegio mínimo. De esta forma, puede minimizar el acceso a la cuenta de administración de la organización y, al mismo tiempo, seguir proporcionando acceso al servicio de toda la organización.

Nivel de cuenta único

Si te diriges a una cuenta con costes elevados pero a la que no tienes acceso AWS Organizations, puedes habilitar Compute Optimizer para esa cuenta y región. Para obtener más información sobre el proceso de suscripción, consulta Cómo empezar con AWS Compute Optimizer en la documentación de Compute Optimizer.

Habilita la recopilación de métricas de memoria para los nodos de Windows

Las métricas de memoria proporcionan a Compute Optimizer las métricas esenciales necesarias para hacer recomendaciones bien informadas sobre el tamaño correcto en tu organización. Esto se debe al análisis de la CPU, la memoria, la red y el almacenamiento que se lleva a cabo antes de ofrecer una recomendación.

Para pasar las métricas de memoria de las EC2 instancias de Windows a Compute Optimizer, debes habilitar el CloudWatch agente y configurar las métricas de memoria para que se recopilen cada 60 segundos. El uso de métricas de memoria con CloudWatch.

Habilite el CloudWatch agente y configure las métricas de memoria

Descargue el ComputeOptimizearchivo.yml. Puedes usar este archivo para habilitar la recopilación de memoria en todas las instancias de tu cuenta. El archivo de plantilla genera los siguientes componentes:

- AWS Systems Manager Almacén de parámetros: almacena la configuración del CloudWatch agente necesaria para recopilar las métricas de memoria.
- AWS Identity and Access Management Función (IAM) con políticas AWS administradas AWS Systems Manager adjuntas: esto es para el documento de automatización de Systems Manager.
- AWS Systems Manager documentos: esto instala y configura el CloudWatch agente (sustituyendo cualquier configuración existente CloudWatch).
- AWS Systems Manager Asociación de administradores de estados: permite que los documentos de Systems Manager se ejecuten en todas las instancias de su cuenta.

> **Important**
>
> La ejecución de esta plantilla sobrescribe cualquier CloudWatch configuración existente en las instancias.

Después, haga lo siguiente:

1. Inicie sesión en la CloudFormation consola AWS Management Console y ábrala.
2. En el panel de navegación, seleccione Stacks (Pilas).
3. Elija Create stack (Crear pila) y, a continuación, seleccione With existing resources (import resources) (Con recursos existentes (importar recursos)).
4. Elija Next (Siguiente).
5. En el recurso de plantilla, selecciona Cargar un archivo de plantilla.
6. Elige un archivo y, a continuación, sube el `ComputeOptimize.yml` archivo.
7. Elija Next (Siguiente).
8. En la página Especificar los detalles de la pila, en Nombre de la pila, introduce un nombre para la pila y, a continuación, selecciona Siguiente.
9. En la página Identificar recursos, introduce los valores de los identificadores de los recursos que vas a importar.
10. Selecciona Importar recursos.
11. Una vez desplegada la pila, selecciona la pestaña Salidas para buscar la clave, el valor y la descripción de la asociación.

Supervise el progreso de la asociación

1. Una vez completada la implementación de la CloudFormation pila, abra la consola de Systems Manager.
2. En el panel de navegación, en la sección Administración de nodos, elija State Manager.
3. En la página de asociaciones, elija el ID de asociación de su asociación.
4. Elija la pestaña Execution history (Historial de ejecución).
5. En la columna ID de ejecución, elige el ID de ejecución de tu asociación. El estado debe ser Éxito.

Vea las métricas en CloudWatch

Te recomendamos que esperes al menos cinco minutos para que se completen las métricas. CloudWatch

1. Abra la consola de CloudWatch.
2. En el panel de navegación, expande la sección Métricas y, a continuación, selecciona Todas las métricas.
3. Confirma que las métricas aparecen en el espacio de CWAgentnombres.

> **Note**
>
> Para aplicar la configuración a cualquier instancia nueva, vuelve a ejecutar la asociación.

Recomendaciones de Consume Compute Optimizer

Considere un ejemplo que se centre en realizar los cambios de tamaño correctos en una sola cuenta y una sola región. En este ejemplo, Compute Optimizer está habilitado a nivel de organización en todas las cuentas. Ten en cuenta que el dimensionamiento correcto es un proceso disruptivo que, en la mayoría de los casos, los propietarios de las aplicaciones llevan a cabo con precisión durante un período de mantenimiento programado que dura varias semanas.

Si accedes a Compute Optimizer desde la cuenta de administración de una organización (como se muestra en los pasos siguientes), puedes elegir la cuenta que deseas investigar. En este ejemplo, hay seis instancias que se ejecutan en una sola cuenta en la `us-east-1` región. Las seis instancias están sobreaprovisionadas. El objetivo es cambiar el tamaño de las instancias según las recomendaciones de Compute Optimizer.

Identifica las instancias sobreaprovisionadas y exporta los detalles de las recomendaciones

1. Inicia sesión en la consola de Compute Optimizer AWS Management Console y ábrela.
2. En el panel de navegación, elija Panel.
3. En el cuadro de búsqueda de la página del panel de control, escribe Region=US East (Virginia del Norte). Luego, ingresa Findings=Over-provisioned. Estos filtros le permiten ver todas las instancias sobreaprovisionadas en la región. `us-east-1`
4. Para consultar las recomendaciones detalladas sobre las EC2 instancias sobreaprovisionadas, desplázate hacia abajo hasta la tarjeta de EC2instancias y, a continuación, selecciona Ver recomendaciones.
5. Seleccione Exportar y guarde el archivo para usarlo en el futuro.
6. Para el bucket de S3, introduzca el nombre del bucket de Amazon S3 que desea que sea el destino del archivo de exportación.

> **Note**
>
> Para guardar las recomendaciones para revisarlas en el futuro, debes tener un bucket de S3 disponible para que Compute Optimizer escriba en cada región. Para obtener más información, consulte la política de bucket de Amazon S3 AWS Compute Optimizer en la documentación de Compute Optimizer.

7. En la sección Exportar filtros, selecciona la casilla Incluir recomendaciones para todas las cuentas de los miembros de la organización.
8. En Tipo de recurso, elija EC2 instancias.
9. En la sección Columnas que se van a incluir, active la casilla de verificación Seleccionar todo.
10. Seleccione Exportar.

Elija las instancias en función de las recomendaciones

Las recomendaciones de instancias se basan en las métricas de rendimiento recopiladas y analizadas por Compute Optimizer. Es esencial conocer las cargas de trabajo que se ejecutan en la instancia para asegurarte de elegir la mejor instancia. En este ejemplo se supone que puede elegir entre la última generación de instancias EC2 R6i, R5 y T3 de Amazon. Las instancias T3 son estables en ráfagas y tienen capacidades de ancho de banda de red más bajas. Las instancias R5 y R6 tienen el mismo costo por hora y son prácticamente idénticas. Sin embargo, la instancia R6 tiene

una mayor capacidad de ancho de banda de red, cuenta con la última generación de procesadores Intel y ofrece el mismo espacio informático que la R5. En este ejemplo, R6 es la mejor opción para cambiar el tamaño.

1. En la consola de Compute Optimizer, selecciona Recomendaciones para EC2 instancias en la barra de navegación. En esta página, se muestra una comparación del tipo de instancia actual con las opciones recomendadas para reemplazarla.
2. Para obtener el ID de la instancia cuyo tamaño desea ajustar, abra la consola Amazon S3 desde la cuenta de administración en AWS Organizations.
3. En el panel de navegación, selecciona Buckets y, a continuación, elige el bucket que vas a usar para almacenar los resultados exportados.
4. En la pestaña Objetos, selecciona el archivo de exportación de la lista de objetos y, a continuación, selecciona Descargar.
5. Para extraer la información de la instancia del archivo, puede utilizar el botón Texto a columnas de la pestaña Datos de Microsoft Excel.

> **Note**
>
> IDs Las instancias se representan como Amazon Resource Names (ARNs). Asegúrese de establecer el delimitador en «/» y extraer el ID de la instancia. Como alternativa, puede escribir un script o usar un entorno de desarrollo integrado (IDE) para recortar el ARN.

6. En Excel, filtre la columna de búsqueda para mostrar solo las instancias OVER_PROVISIONED. Estas son las instancias a las que te diriges para que tengan el tamaño correcto.
7. Guarda la instancia IDs en un editor de texto para poder acceder a ella fácilmente más adelante.

Etiquete las instancias para obtener el tamaño correcto

Etiquetar tus cargas de trabajo es una herramienta poderosa para organizar tus recursos. AWS Las etiquetas le permiten obtener una visibilidad detallada de los costes y facilitar la devolución de los gastos. Para obtener más información sobre las estrategias y los métodos para añadir etiquetas a AWS los recursos, consulte el AWS documento técnico Mejores prácticas para etiquetar los recursos. AWS Para este ejemplo, puede usar el editor de AWS etiquetas para realizar ajustes de etiquetado en las instancias sobreaprovisionadas a las que desee redimensionar durante un período de mantenimiento. También puedes usar esta etiqueta para ver los costes antes y después del cambio.

1. Inicia sesión en la AWS Resource Groups consola de la AWS Management Console cuenta que contiene las instancias a las que se va a cambiar el tamaño y ábrela.
2. En la barra de navegación, en la sección Etiquetado, selecciona Editor de etiquetas.
3. En el caso de las regiones, selecciona la región de destino.
4. Para los tipos de recursos, elija AWS::EC2::Instance.
5. Seleccione Buscar recursos.
6. En la página de resultados de la búsqueda de recursos, selecciona todas las instancias a las que quieras ajustar el tamaño y, a continuación, selecciona Administrar las etiquetas de los recursos seleccionados.
7. Seleccione Agregar etiqueta.
8. En Tag key, ingresa Rightsizing. En Valor de etiqueta, escriba enabled. A continuación, selecciona Revisar y aplicar los cambios en la etiqueta.

> **Note**
>
> Puede incluir metadatos adicionales, como el equipo o la unidad de negocio, para facilitar el filtrado más adelante en Cost Explorer.

Tras crear y aplicar las etiquetas definidas por el usuario a sus recursos, las etiquetas pueden tardar hasta 24 horas en aparecer en la página de etiquetas de asignación de costes para su activación. Una vez que haya seleccionado las etiquetas para activarlas, las etiquetas podrían tardar otras 24 horas en activarse.

Para los usuarios avanzados, puedes usar AWS CloudShellla cuenta y la región de destino para etiquetar varias instancias. Por ejemplo:

```
bash
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="type-m5"
# Get a list of instance IDs
INSTANCE_IDS=$(aws ec2 describe-instances –query
 "Reservations[].Instances[].InstanceId" –output text)
# Loop through each instance ID and add the tag
for INSTANCE_ID in $INSTANCE_IDS; do
 aws ec2 create-tags –resources $INSTANCE_ID –tags Key=$TAG_KEY,Value=$TAG_VALUE
```

```
done
```

Habilita la etiqueta de asignación de costes para que funcione con las herramientas AWS de facturación

Recomendamos activar la etiqueta de asignación de costes definida por el usuario. Esto permite reconocer y filtrar la etiqueta Rightsizing en las herramientas de AWS facturación (por ejemplo, Cost Explorer y). AWS Cost and Usage Report Si no lo habilitas, la opción de filtrado de etiquetas y los datos no estarán disponibles. Para obtener información sobre el uso de etiquetas de asignación de costes, consulte Activar etiquetas de asignación de costes definidas por el usuario en la Administración de facturación y costos de AWS documentación.

1. Inicie sesión en la AWS Billing consola AWS Management Console y ábrala.
2. En el panel de navegación, en la sección Facturación, selecciona las etiquetas de asignación de costes.
3. En la pestaña Etiquetas de asignación de costes definidas por el usuario, introduzca Rightsizing.
4. Seleccione la tecla de etiqueta RightSizing y, a continuación, elija Activar.

Transcurridas 24 horas, la etiqueta debería aparecer en Cost Explorer.

Implemente las recomendaciones de tamaño adecuadas con Systems Manager Automation

El cambio de tamaño es un escenario que requiere detener e iniciar una instancia. En este escenario, es posible que tengas que gestionar esta interrupción en un período de mantenimiento y que necesites que diferentes equipos se encarguen de su propio redimensionamiento. Antes de cambiar un tipo de instancia, consulta las Consideraciones sobre los tipos de instancias compatibles en la EC2 documentación de Amazon.

Los pasos de ejemplo de esta sección implementan las recomendaciones de tamaño correcto por cuenta y región mediante un documento de automatización de Systems Manager denominado AWS-ResizeInstance. Este enfoque es típico de la mayoría de las organizaciones, ya que la mayoría de las organizaciones requieren diferentes tipos de instancias para diferentes propósitos. También puedes usar el mismo documento de `AWS-ResizeInstance` automatización para segmentar las implementaciones de una o varias cuentas.

1. Inicie sesión en la consola de Systems Manager AWS Management Console y ábrala.

2. En el panel de navegación, en la sección Recursos compartidos, elija Documentos.
3. En la barra de búsqueda, escribe AWS- yResizeInstance, a continuación, selecciona AWS- en los resultados ResizeInstance de la búsqueda.
4. Elija Ejecutar automatización.
5. En la página Ejecutar el manual de automatización, elija Ejecución simple.
6. En la sección Parámetros de entrada, introduzca InstanceIdy InstanceType. Conserve el resto de los valores predeterminados.
7. Selecciona Ejecutar y, a continuación, espera a que la automatización complete los pasos necesarios para cambiar el tipo de instancia.

Considera métodos de redimensionamiento alternativos

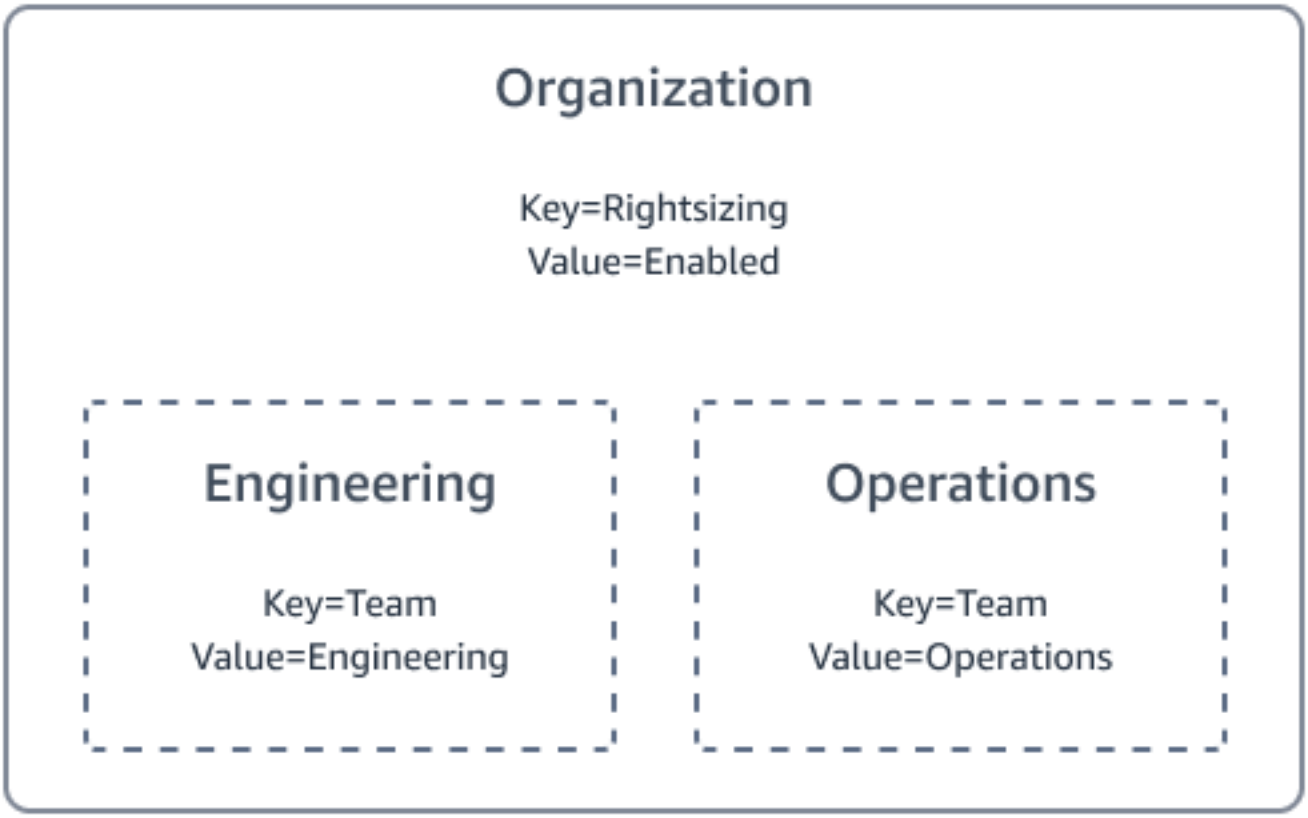
Si utilizas una plantilla de lanzamiento para implementar tus instancias, puedes actualizar la plantilla de lanzamiento con el tipo de instancia del tamaño correcto y, a continuación, realizar una actualización de la instancia para reemplazar las instancias por la versión del tamaño correcto.

Si planea usar el proceso de dimensionamiento correcto en varias cuentas y regiones, debe crear un documento personalizado de Systems Manager Automation. Este documento le permite introducir varias instancias como parámetro y trasladar las instancias de destino al mismo tipo de instancia de destino (por ejemplo, todas las instancias en transición a t3a.medium, independientemente del tipo de instancia de origen).

Revise los costes anteriores y posteriores en Cost Explorer

Una vez que haya dimensionado correctamente sus recursos, puede usar el Explorador de costos para mostrar los costos anteriores y posteriores mediante la etiqueta Rightsizing. Recuerde que puede usar etiquetas de recursos para realizar un seguimiento de los costos. Al utilizar varias capas de etiquetas, puede lograr una visibilidad pormenorizada de sus costes. En el ejemplo que se describe en esta guía, la etiqueta Rightsizing se utiliza para aplicar una etiqueta genérica a todas las instancias objetivo. A continuación, se utiliza una etiqueta de equipo para organizar aún más los recursos. El siguiente paso consiste en introducir etiquetas de aplicación para mostrar mejor el impacto en los costes de funcionamiento de una aplicación específica.

El siguiente diagrama muestra la estructura de etiquetas de una organización.



Considere el ejemplo de una empresa que dimensiona correctamente los servidores web de producción propiedad del equipo de operaciones. En Cost Explorer, la etiqueta Rightsizing está configurada como habilitada y la etiqueta Team está configurada como operaciones. En este ejemplo, el dimensionamiento correcto reduce los costos operativos de 0,89 centavos a 0,28 centavos por hora. Suponiendo que haya 744 horas al mes, el coste anual antes del dimensionamiento correcto es de 7.945,92$. Con el tamaño correcto, el costo anual se reduce a 2.499,84 dólares. Esto se traduce en una disminución del 68,5 por ciento en los costes anuales de la carga de trabajo. Imagine el impacto de esto en una organización grande. Tenga en cuenta que esto se hace en un entorno de muestra y las instancias están en su mayoría inactivas. En un entorno de producción, puede obtener ahorros de entre un 10 y un 35 por ciento.

Ahora, considere el impacto que tendría el dimensionamiento correcto del bastión de producción que es propiedad del equipo de ingeniería. En Cost Explorer, la etiqueta Rightsizing está configurada como habilitada y la etiqueta Team está configurada como ingeniería. En este ejemplo, el dimensionamiento correcto reduce los costos de 0,75 centavos a 0,44 centavos por hora. Suponiendo que haya 744 horas al mes, el coste anual antes del dimensionamiento correcto es de 6.696,00$. Con el tamaño correcto, el costo anual se reduce a 3.928,32 dólares.

Si utilizas varias etiquetas, puedes filtrar los datos para obtener detalles detallados de los costes. En este ejemplo, la etiqueta Equipo reduce el ruido para que puedas ver el impacto a nivel de equipo. Como la etiqueta Rightsizing está habilitada, también puedes filtrar por cualquier instancia que tenga esa etiqueta con el valor activado o sin ningún valor. Esto puede proporcionar una visión global de sus esfuerzos de dimensionamiento correctos, especialmente si se ven en la cuenta de administración (pagadora) en el nivel de Cost Explorer. Esta vista le permite ver todas las cuentas e instancias.

Considere un ejemplo a nivel de cuenta única en el que la etiqueta Rightsizing esté configurada como habilitada. Los costos operativos bajan de 1,64 dólares por hora a 0,72 centavos por hora. Suponiendo que haya 744 horas al mes, el coste anual, sin tener en cuenta el tamaño correcto, es de 14.641,92 dólares. Con el tamaño correcto, el costo anual se reduce a 6.428,16 dólares. Esto se traduce en una disminución del 56 por ciento en los costos de cómputo de esta cuenta.

Antes de embarcarse en la elección del tamaño correcto, tenga en cuenta lo siguiente:

- AWS ofrece muchas opciones para reducir los costos. Esto incluye AWS OLA, en el que se AWS revisan las instancias locales antes de mudarlas a AWS ellas. El AWS OLA también le proporciona recomendaciones sobre el tamaño correcto y orientación sobre licencias.
- Complete todas las tallas correctas antes de comprar Savings Plans. Esto puede ayudarlo a evitar compras excesivas en su compromiso con Savings Plans.

Recomendaciones

Recomendamos los siguientes pasos:

1. Revise su panorama actual y considere la posibilidad de convertir los volúmenes gp2 de Amazon EBS en volúmenes gp3.
2. Revise Savings Plans.

Recursos adicionales

- AWS Compute Optimizer(AWS documentación)
- Mejores prácticas para etiquetar AWS recursos (documentos técnicos AWS)
- ¿Cómo recopilar datos de AWS Compute Optimizer y AWS Trusted Advisor entre sus AWS Organizations () YouTube

- Optimización del rendimiento y reducción de los costes de licencias: aprovechamiento de AWS Compute Optimizer las instancias de Amazon EC2 SQL Server (Microsoft Workloads en el AWS blog)

Seleccione el tipo de instancia adecuado para las cargas de trabajo de Windows

Descripción general

Una diferencia importante entre las cargas de trabajo que funcionan en la nube y las que funcionan en entornos locales es la práctica del sobreaprovisionamiento. Al comprar hardware físico para uso local, se realiza un gasto de capital que se prevé que dure un período predeterminado, normalmente de 3 a 5 años. Para adaptarse al crecimiento previsto durante la vida útil del hardware, éste se adquiere con más recursos de los que su carga de trabajo requiere actualmente. En consecuencia, el hardware físico suele estar sobreaprovisionado, lo que supera con creces las necesidades de la carga de trabajo real.

La tecnología de máquinas virtuales (VM) surgió como un medio eficaz de utilizar los recursos de hardware sobrantes. Los administradores tenían un aprovisionamiento excesivo VMs de v CPUs y RAM, lo que permitía al hipervisor gestionar el uso de los recursos físicos entre los servidores ocupados e inactivos mediante la asignación de los recursos no utilizados a cada máquina virtual. Al administrar VMs, los recursos de vCPU y RAM asignados a cada máquina virtual funcionaban más como reguladores de recursos que como indicadores del uso real. La sobreasignación de recursos de las máquinas virtuales podría superar fácilmente el triple de los recursos informáticos disponibles.

Amazon Elastic Compute Cloud (Amazon EC2) se abstiene de aprovisionar VMs en exceso el hardware subyacente, ya que no es necesario. La computación en la nube es un gasto operativo, no un gasto de capital, y usted solo paga por lo que utiliza. Si su carga de trabajo requiere más recursos en el futuro, aprovisiónelos cuando realmente los necesite, en lugar de hacerlo de forma preventiva.

Existen cientos de opciones para elegir los tipos de EC2 instancias de Amazon correctos. Si tienes pensado migrar una carga de trabajo de Windows a la nube, AWS ofrece un AWS OLA para ayudarte a entender mejor tu carga de trabajo actual y proporcionarte un ejemplo de su rendimiento en ella AWS. El análisis de AWS OLA tiene como objetivo hacer coincidir el tipo y el tamaño de EC2 instancia adecuados con el uso real en las instalaciones.

Si ya tiene cargas de trabajo ejecutándose en Amazon EC2 y busca estrategias de optimización de costes, esta sección de la guía le ayudará a identificar las diferencias entre las EC2 instancias de Amazon y su aplicabilidad a las cargas de trabajo típicas de Windows.

Recomendaciones de optimización de costos

Para optimizar los costes de sus tipos de EC2 instancias, le recomendamos que haga lo siguiente:

- Elija la familia de instancias adecuada para su carga de trabajo
- Comprenda las variaciones de precios entre las arquitecturas de procesadores
- Comprenda las diferencias entre precio y rendimiento entre generaciones EC2
- Migre a instancias más nuevas
- Utilice instancias con capacidad de ráfaga

Elija la familia de instancias adecuada para su carga de trabajo

Es importante elegir la familia de instancias adecuada para su carga de trabajo.

Las EC2 instancias de Amazon se dividen en los siguientes grupos:

- Fin general
- Optimizada para computación
- Optimizada para memoria
- Computación acelerada
- Optimizada para almacenamiento
- Optimizado para HPC

La mayoría de las cargas de trabajo de Windows se clasifican en las siguientes categorías:

- Fin general
- Optimizada para computación
- Optimizada para memoria

Para simplificarlo aún más, considere una EC2 instancia de referencia en cada categoría:

- Optimizado para cómputo: C6i

- Uso general: M6i
- Memoria optimizada: R6i

La generación anterior de EC2 instancias presentaba pequeñas diferencias en los tipos de procesadores. Por ejemplo, las instancias optimizadas para cómputo C5 tienen procesadores más rápidos que las instancias M5 de uso general o las instancias R5 optimizadas para memoria. Todas las EC2 instancias de última generación (C6i, M6i, R6i, C6a, M6a y R6a) utilizan el mismo procesador en todas las familias de instancias. Como el procesador es uniforme en la última generación de instancias, la diferencia de precio entre las familias de instancias ahora depende más de la cantidad de RAM. Cuanta más RAM tenga una instancia, más cara será.

En el siguiente ejemplo, se muestra el precio por hora de una instancia de 4 vCPU basada en Intel que se ejecuta en `us-east-1` la región.

instancia	v CPUs	RAM	Precio por hora
c6i.xlarge	4	8	0,17$
m6i.xlarge	4	16	0,19$
r6i.xlarge	4	32	0,25 DÓLARES

> **Note**
>
> Los precios se basan en los precios por hora a pedido en la `us-east-1` región.

Instancias explotables

Si bien es una práctica recomendada en la computación en nube desactivar los recursos de cómputo no utilizados para evitar cargos, no todas las cargas de trabajo se pueden apagar y encender cada vez que se necesitan. Algunas cargas de trabajo permanecen inactivas durante períodos prolongados, pero deben estar accesibles las 24 horas del día.

Las instancias de fragmentación (T3) ofrecen una forma de mantener en línea las cargas de trabajo con picos de actividad o de bajo consumo durante todo el día y, al mismo tiempo, mantener bajos los costos de procesamiento. EC2 Las instancias Burstable tienen una cantidad máxima de recursos

de vCPU que la instancia puede usar durante períodos breves. Estas instancias utilizan un sistema basado en créditos de CPU estables. Estos créditos se acumulan durante los períodos de inactividad a lo largo del día. Las instancias Burstable ofrecen vCPU-to-RAM proporciones variables, lo que las convierte en alternativas a las instancias optimizadas para la computación en algunos casos y a otras instancias de uso general en otros.

En el siguiente ejemplo, se muestra el precio por hora de una instancia T3 (es decir, una instancia de fragmentación) que se ejecuta en la región. `us-east-1`

instancia	v CPUs	RAM (GB)	Precio por hora
t3.nano	2	0,5	0,0052$
t3.micro	2.	1	0,0104$
t3.small	2	2.	0,0208$
t3.medium	2	4	0,0416$
t3.large	2	8	0,0832$
t3.xlarge	4	16	0,1664$
t3.2xlarge	8	32	0,3328$

> **Note**
>
> Los precios se basan en los precios por hora bajo demanda en la `us-east-1` región.

Comprenda las variaciones de precios entre las arquitecturas de procesadores

Los procesadores Intel han sido el estándar en EC2 algunos casos desde su creación. Las EC2 instancias de generaciones anteriores, como la C5, la M5 y la R5, no indicaban que la arquitectura de procesador fuera Intel (ya que era la predeterminada). Las generaciones más recientes de EC2 instancias, como la C6i, la M6i y la R6i, incluyen una «i» para indicar el uso de un procesador Intel.

El cambio en la anotación sobre la arquitectura del procesador se debe a la introducción de opciones de procesador adicionales. El procesador más comparable a Intel es AMD (indicado con una «a»).

Los procesadores AMD EPYC utilizan la misma arquitectura x86 y ofrecen un rendimiento similar al de los procesadores Intel, pero a un precio inferior. Como se demuestra en los siguientes ejemplos de precios, las EC2 instancias AMD ofrecen un descuento de aproximadamente un 10 por ciento en los costos de cómputo en comparación con sus homólogas de Intel.

Instancia de Intel	Precio por hora	Instancia AMD	Precio	% de diferencia
c6i.xlarge	0,17$	c6a.xlarge	0,153$	10%
m6i.xlarge	0,192$	m6a.xlarge	0,1728$	10%
r6i.xlarge	0,252$	r6a.xlarge	0,2268$	10%

> **Note**
>
> Los precios se basan en los precios por hora bajo demanda en la `us-east-1` región.

La tercera opción principal de arquitectura de procesador son los procesadores AWS Graviton (marcados con una «g») en las EC2 instancias. Diseñados por AWS, los procesadores Graviton ofrecen la mejor relación precio-rendimiento de Amazon EC2. Los procesadores Graviton actuales no solo son un 20 por ciento más baratos que sus homólogos de Intel, sino que también ofrecen un aumento del rendimiento del 20 por ciento o más. Se espera que la próxima generación de procesadores Graviton amplíe aún más esta diferencia de rendimiento, y las pruebas muestran un aumento adicional del rendimiento del 25 por ciento.

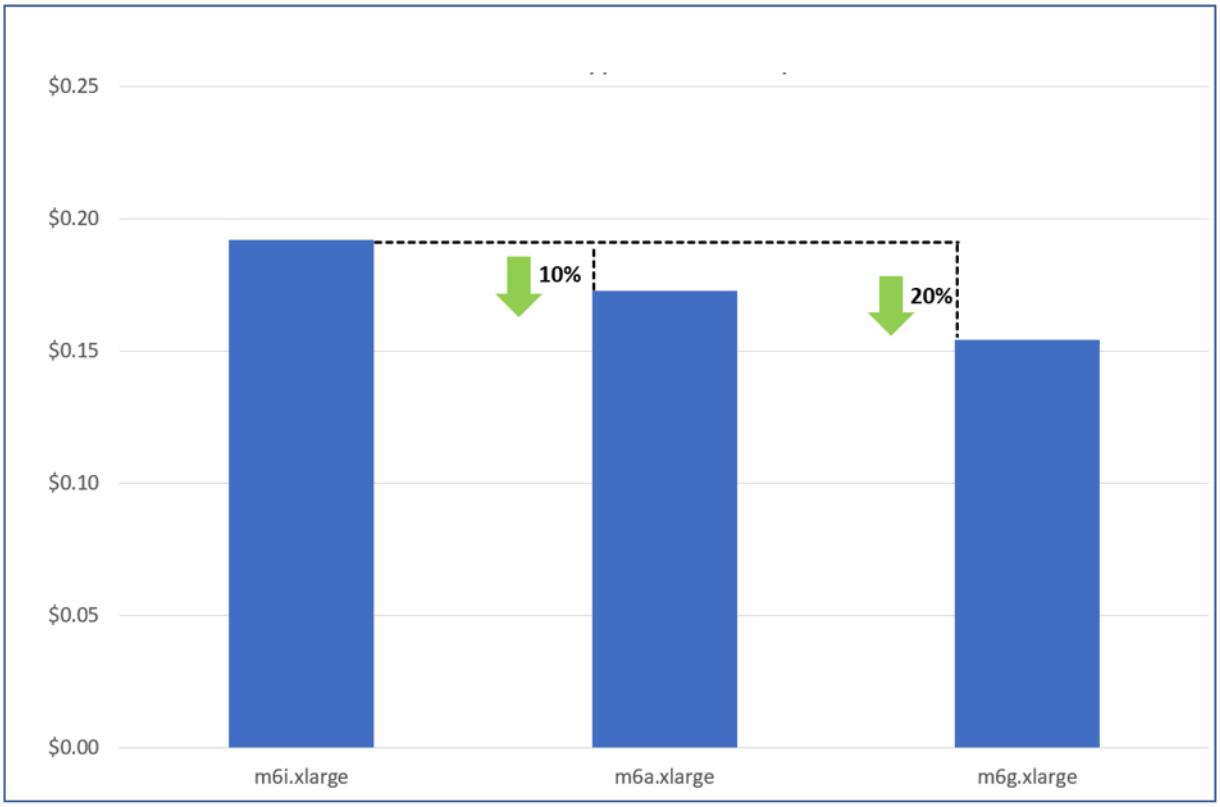
Windows Server no se puede ejecutar en los procesadores Graviton, que se basan en la arquitectura ARM. De hecho, Windows Server solo funciona con procesadores x86. Si bien no puede lograr un aumento del precio y el rendimiento del 40 por ciento con las instancias basadas en Graviton para Windows Server, sí puede usar los procesadores Graviton con cargas de trabajo específicas de Microsoft. Por ejemplo, las versiones más recientes de .NET pueden ejecutarse en Linux. Esto significa que estas cargas de trabajo pueden usar procesadores ARM y beneficiarse de instancias Graviton EC2 más rápidas y asequibles.

En el siguiente ejemplo, se muestra el precio por hora de una instancia de Graviton que se ejecuta en la región. `us-east-1`

Instancia de Intel	Precio por hora	Instancia de Graviton	Precio por hora	% de diferencia
c6i.xlarge	0,17$	c6g.xlarge	0,136$	20%
m6i.xlarge	0,192$	m6g.xlarge	0,154$	20%
r6i.xlarge	0,252$	r6g.xlarge	0,2016$	20%

> **Note**
>
> Los precios se basan en los precios por hora bajo demanda en la `us-east-1` región.

En el siguiente gráfico se comparan los precios de las instancias de la serie M.



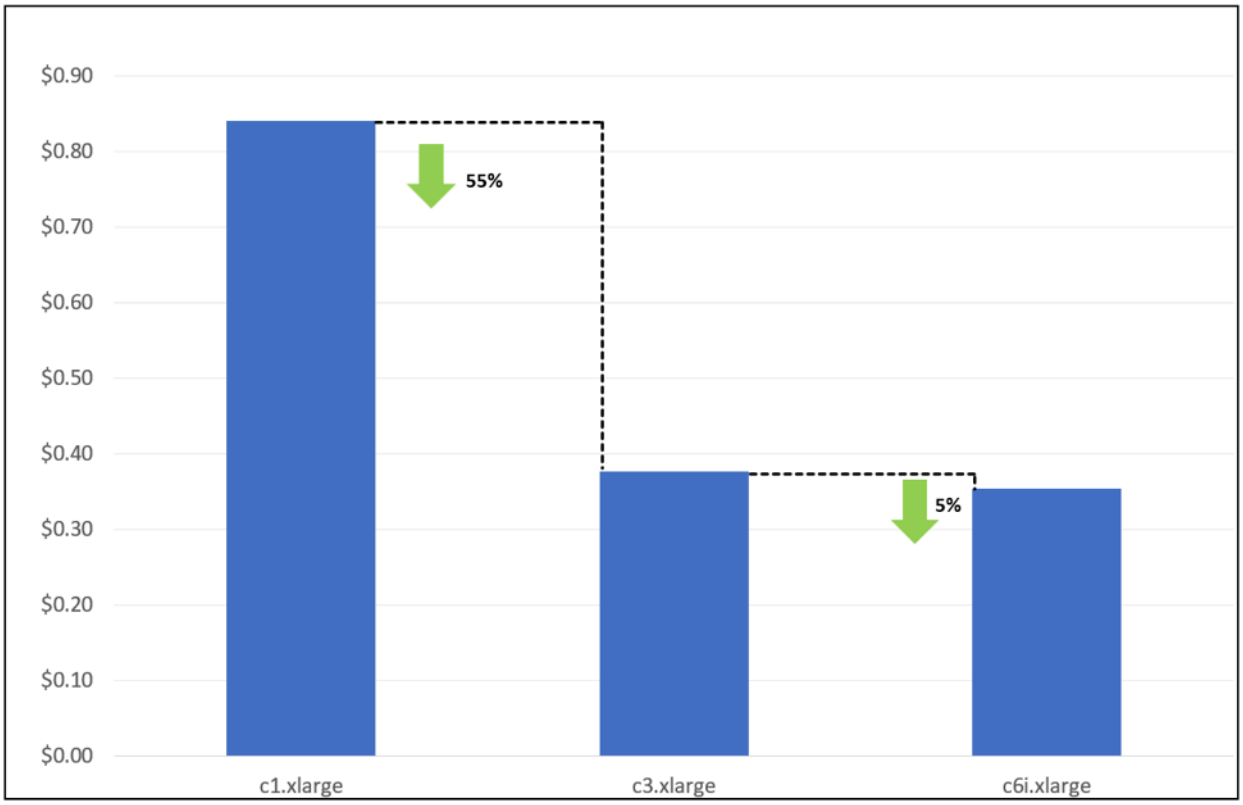
Comprenda las diferencias de precio y rendimiento entre EC2 generaciones

Una de las características más consistentes de Amazon EC2 es que cada nueva generación ofrece una mejor relación precio-rendimiento que su predecesora. Como se muestra en la siguiente tabla, el precio de las EC2 instancias de nueva generación disminuye con cada versión posterior.

Instancia optimizada para computación	Precio por hora	Instancia de uso general	Precio por hora	Instancia optimizada para memoria	Precio por hora
C1.xlarge	0,52$	M1.xLarge	0,35$	r1.xlarge	n/a
C3.xLarge	0,21$	M3.x Large	0,266$	r3.xlarge	0,333$
C5.xLarge	0,17$	M5.xLarge	0,192$	r5.xlarge	0,252$

> **Note**
>
> Los precios se basan en los precios por hora bajo demanda en la `us-east-1` región.

En el siguiente cuadro se comparan los costos de las distintas generaciones de instancias de la serie C.



Sin embargo, la sexta generación de instancias tiene el mismo precio que la quinta generación, como se muestra en la siguiente tabla.

Instancia optimizada para computación	Precio por hora	Instancia de uso general	Precio por hora	Instancia optimizada para memoria	Precio por hora
C5.xlarge	0,17$	M5.xLarge	0,192$	r5.xlarge	0,252$
C6i.XLarge	0,17$	M6. X grande	0,192$	r6i.xlarge	0,252$

> **Note**
>
> Los precios se basan en los precios por hora bajo demanda en la `us-east-1` región.

A pesar de tener el mismo coste, la nueva generación ofrece una relación precio-rendimiento superior gracias a los procesadores más rápidos, la mejora del rendimiento de la red y el aumento del rendimiento y las IOPS de Amazon Elastic Block Store (Amazon EBS).

Una de las mejoras de precio-rendimiento más importantes es la mejora de la instancia X2i. Esta generación de instancias ofrece una relación precio-rendimiento hasta un 55 por ciento superior a la de la generación anterior. Como se muestra en la siguiente tabla, la x2iedn demuestra una mejora en todos los aspectos del rendimiento (todo al mismo precio que la generación anterior).

instancia	Precio por hora	v CPUs	RAM	Velocidad del procesador	Almacenamiento de la instancia	Red	Rendimiento de Amazon EBS	IOPS DE EBS
x1e.2xlarge	1,66$	8	244	2.3 GHz	SSD DE 237 GB	10 Gbps	125 MB/s	7400
x1 iedn.2 x grande	1,66$	8	256	3.5 GHz	SSD de 240 GB NVMe	25 Gbps	2500 MB/s	65000

> **Note**
>
> Los precios se basan en los precios por hora bajo demanda en la `us-east-1` región.

Ejemplos de escenarios de

Pensemos en el ejemplo de una empresa de análisis que hace un seguimiento de los vehículos de entrega y desea mejorar el rendimiento de SQL Server. Una vez que una pyme de MACO analiza los problemas de rendimiento de esta empresa, la empresa pasa de las instancias x1e.2xlarge a las instancias x2iedn.xlarge. El tamaño de la nueva instancia es más pequeño, pero las mejoras introducidas en las instancias x2 permiten aumentar el rendimiento y la optimización de SQL Server mediante el uso de Buffer Pool Extensions. Esto permite a la empresa pasar de la edición Enterprise de SQL Server a la edición Standard de SQL Server. También permite a la empresa reducir sus licencias de SQL Server de 8 CPUs a 4 versiones. CPUs

Antes de la optimización:

Server	EC2 instancia	Edición de SQL Server	Costo mensual
Prod DB1	x1e.2xlarge	Enterprise	3.918,64 DÓLARES
Prod DB2	x1e.2xlarge	Enterprise	3.918,64 DÓLARES
Total			7.837,28 DÓLARES

Tras la optimización:

Server	EC2 instancia	Edición de SQL Server	Costo mensual
Prod DB1	x2iedn.xlarge	Estándar	1.215,00$
Orgulloso DB2	x2iedn.xlarge	Estándar	1.215,00$
Total			2.430,00 DÓLARES

En conjunto, el cambio de las instancias x1e.2xlarge a las instancias x2iedn.xlarge permite a la empresa, en el escenario del ejemplo, ahorrar 5.407$ al mes en sus servidores de bases de datos de producción. Esto reduce el costo total de la carga de trabajo en un 69 por ciento.

> **Note**
>
> Los precios se basan en los precios por hora a pedido en la `us-east-1` región.

Migre a instancias más nuevas

Las generaciones anteriores de Amazon EC2 utilizan el hipervisor Xen, mientras que las generaciones más recientes utilizan el sistema AWS Nitro. El sistema Nitro entrega casi todos los recursos de cómputo y memoria del hardware anfitrión a sus instancias. Esto se traduce en un rendimiento general mejorado. Al migrar de instancias de Xen a instancias basadas en Nitro, se

deben tener en cuenta aspectos especiales. Por ejemplo, AWS Windows AMIs está configurado con la configuración y las personalizaciones predeterminadas que utilizan los medios de instalación de Microsoft. Las personalizaciones incluyen controladores y configuraciones que admiten los tipos de instancias de última generación (instancias basadas en el sistema Nitro).

Si va a lanzar instancias desde Windows personalizado AMIs o desde Windows AMIs proporcionado por Amazon que se crearon antes de agosto de 2018, le recomendamos que complete los pasos de Migración a tipos de instancias de última generación que se indican en la EC2 documentación de Amazon.

Usa instancias con capacidad de fragmentación

Si bien las instancias con capacidad de fragmentación son una buena forma de ahorrar costes de procesamiento, le recomendamos que las evite en los siguientes escenarios:

- Las especificaciones mínimas de Windows Server with the Desktop Experience requieren 2 GB de RAM. Evite usar instancias t3.micro o t3.nano con Windows Server porque carecen de la cantidad mínima de RAM.
- Si tu carga de trabajo es intensa pero no permanece inactiva el tiempo suficiente para acumular créditos de ráfaga, usar EC2 instancias normales es más eficiente que usar instancias con ráfagas. Te recomendamos que monitorices tus créditos de CPU para verificarlo.
- En la mayoría de los casos, se recomienda evitar el uso de instancias fragmentables con SQL Server. La licencia de SQL Server se basa en el número de v CPUs asignado a una instancia. Si SQL Server está inactivo la mayor parte del día, tendrá que pagar por licencias de SQL que no utilice en su totalidad. En estos escenarios, le recomendamos que consolide varias instancias de SQL Server en un servidor más grande.

Pasos a seguir a continuación

Le recomendamos que siga los siguientes pasos para optimizar los costes de las instancias de Amazon EC2 Windows:

- Utilice la EC2 instancia de última generación para obtener la mejor relación precio-rendimiento.
- Use EC2 instancias con procesadores AMD para reducir los costos de cómputo en un diez por ciento.
- Maximice la utilización de los recursos al elegir un tipo de EC2 instancia que se adapte a su carga de trabajo.

En la siguiente tabla, se muestran ejemplos de puntos de partida típicos para las cargas de trabajo de Windows. Hay opciones adicionales disponibles, como volúmenes de almacenamiento de instancias para mejorar las cargas de trabajo de SQL Server o EC2 instancias con proporciones mucho mayores vCPU-to-RAM. Te recomendamos que pruebes tus cargas de trabajo exhaustivamente y que utilices herramientas de supervisión AWS Compute Optimizer para ayudarte a realizar los ajustes necesarios.

Carga de trabajo	Típico	Opcional
Active Directory	T3, M6i	R6i
Servidores de archivos	T3, M6i	C6i
Servidores web	T3, C6i	M6i, R6i
SQL Server	R6i	x2iedn, x2iezn

Si debe cambiar el tipo de EC2 instancia, el proceso normalmente implica un simple reinicio del servidor. Para obtener más información, consulta Cambiar el tipo de instancia en la EC2 documentación de Amazon.

Antes de cambiar el tipo de instancia, le recomendamos que tenga en cuenta lo siguiente:

- Debe detener las instancias respaldadas por Amazon EBS antes de poder cambiar su tipo de instancia. Asegúrese de planificar el tiempo de inactividad mientras la instancia esté detenida. El detenimiento y el cambio de tipo de instancia puede tardar unos minutos y el tiempo que se tarda en reiniciar la instancia es variable, en función de los scripts de inicio de la aplicación. Para obtener más información, consulta Detener e iniciar tu instancia en la EC2 documentación de Amazon.
- Cuando detiene e inicia una instancia, AWS mueve la instancia a un nuevo hardware. Si la instancia tiene una IPv4 dirección pública, AWS libera la dirección y proporciona a la instancia una nueva IPv4 dirección pública. Si necesitas una IPv4 dirección pública que no cambie, usa una dirección IP elástica.
- No puedes cambiar el tipo de instancia si la hibernación está habilitada en la instancia.
- No puede cambiar el tipo de instancia de una instancia de spot.
- Si la instancia está en un grupo de Auto Scaling, Amazon EC2 Auto Scaling marca la instancia detenida como en mal estado y puede cerrarla y lanzar una instancia de reemplazo. Para evitar esto, puede suspender los procesos de escalado del grupo mientras cambia el tipo de instancia.

Para obtener más información, consulte Suspender y reanudar un proceso para un grupo de Auto Scaling en la documentación de Amazon EC2 Auto Scaling.

- Cuando cambias el tipo de instancia de una NVMe instancia con volúmenes de almacenes de instancias, la instancia actualizada puede tener volúmenes de almacenes de instancias adicionales, ya que todos los volúmenes de almacenes de NVMe instancias están disponibles incluso si no están especificados en Amazon Machine Image (AMI) o en la asignación de dispositivos de bloques de instancias. De lo contrario, la instancia actualizada tiene la misma cantidad de volúmenes de almacén de instancias que especificó cuando lanzó la instancia original.

Recursos adicionales

- Tipos de EC2 instancias de Amazon (AWS documentación)
- AWS Evaluación de optimización y licencias (AWS documentación)

Traiga licencias para cargas de trabajo de Windows y SQL Server

Descripción general

Si tiene inversiones importantes en cargas de trabajo de Microsoft y acuerdos de licencia empresarial existentes, puede elegir entre varias AWS opciones para respaldar estas cargas de trabajo, incluidas las opciones de licencia incluida (proporcionada por AWS) y Bring Your Own License (BYOL). Puede utilizar Amazon EC2 Dedicated Hosts para aprovechar al máximo los acuerdos de licencia de Microsoft existentes y llevar Windows Server a AWS. Esto puede ahorrarle hasta un 50 por ciento en los costes de las EC2 instancias de Amazon. Dado que las licencias de Windows representan aproximadamente la mitad de los costes de las instancias, instalar Windows Server AWS en hosts dedicados puede suponer un importante ahorro de costes. Dado que Windows Server no puede tener un arrendamiento predeterminado (compartido), los hosts dedicados son la opción ideal si desea utilizar sus licencias actuales de Windows Server en AWS ellas.

Los hosts dedicados no son solo para las instancias BYOL de Windows Server. También le ofrecen la flexibilidad necesaria para adaptar sus licencias locales a las cargas de trabajo de SQL Server existentes. Los hosts dedicados exponen los núcleos físicos del servidor subyacente y le permiten licenciar SQL Server a nivel de núcleo físico. Esto no es posible en el arrendamiento predeterminado (compartido), donde las licencias de SQL Server se basan en la cantidad de unidades virtuales CPUs asignadas a la instancia. Esta función le permite licenciar cargas de trabajo de SQL Server AWS de forma coherente con su estrategia de licencias locales. Por lo tanto, puede ahorrar hasta

un 50 por ciento en los costos de licencias de SQL Server en comparación con el arrendamiento predeterminado (compartido), además de ahorrar en los costos de las instancias, si utiliza licencias de Windows aptas. Para obtener más información sobre este escenario, consulte la sección Descripción de las licencias de SQL Server de esta guía.

Hosts EC2 dedicados de Amazon

Un host EC2 dedicado de Amazon es básicamente el mismo EC2 host que AWS utiliza para ejecutar sus ofertas EC2 informáticas. La diferencia es que estos hosts están totalmente dedicados a un solo cliente y proporcionan acceso exclusivo a la infraestructura física subyacente. Puede usar hosts dedicados para ejecutar sus instancias en un hardware totalmente dedicado a su uso, en lugar de compartir recursos con otros AWS clientes. Esto le da un mayor control sobre los recursos de la nube y le permite reducir los costos al incorporar sus propias licencias de software, como Windows Server y SQL Server AWS.

Tenga en cuenta lo siguiente:

- Un host dedicado es un servidor físico totalmente dedicado a un solo cliente. Obtiene visibilidad de los sockets y los núcleos físicos del host dedicado para poder cumplir con los requisitos de conformidad con las licencias, como los acuerdos de licencia de software por socket, por núcleo o por máquina virtual.
- Los hosts dedicados que admiten varios tamaños de instancias de la misma familia de instancias se denominan hosts dedicados heterogéneos. Estas familias de instancias incluyen T3, A1, C5, M5, R5, C5n, R5n y M5n. Por el contrario, otras familias de instancias admiten solo un tamaño de instancia en el mismo host dedicado. Se denominan hosts dedicados homogéneos.
- Los hosts dedicados se facturan por host. Esto significa que se le cobrará por host dedicado, independientemente del número de instancias que se ejecuten en él. El precio del host dedicado varía en función de la familia de instancias, la región y la opción de pago seleccionada. Puede elegir la configuración óptima para su carga de trabajo a fin de lograr los resultados de rendimiento y costos deseados.

Este diagrama ilustra las diferencias entre las instancias de arrendamiento compartido y los hosts dedicados.

Hosts dedicados homogéneos

Considere un escenario en el que se utilice un host dedicado M6i. Los hosts dedicados M6i y R6i tienen dos sockets, 64 núcleos físicos y admiten tipos de instancias del mismo tamaño. Se denominan hosts dedicados homogéneos. Esto significa que la cantidad de instancias que puede lanzar en un único host dedicado M6i depende del tamaño de la instancia.

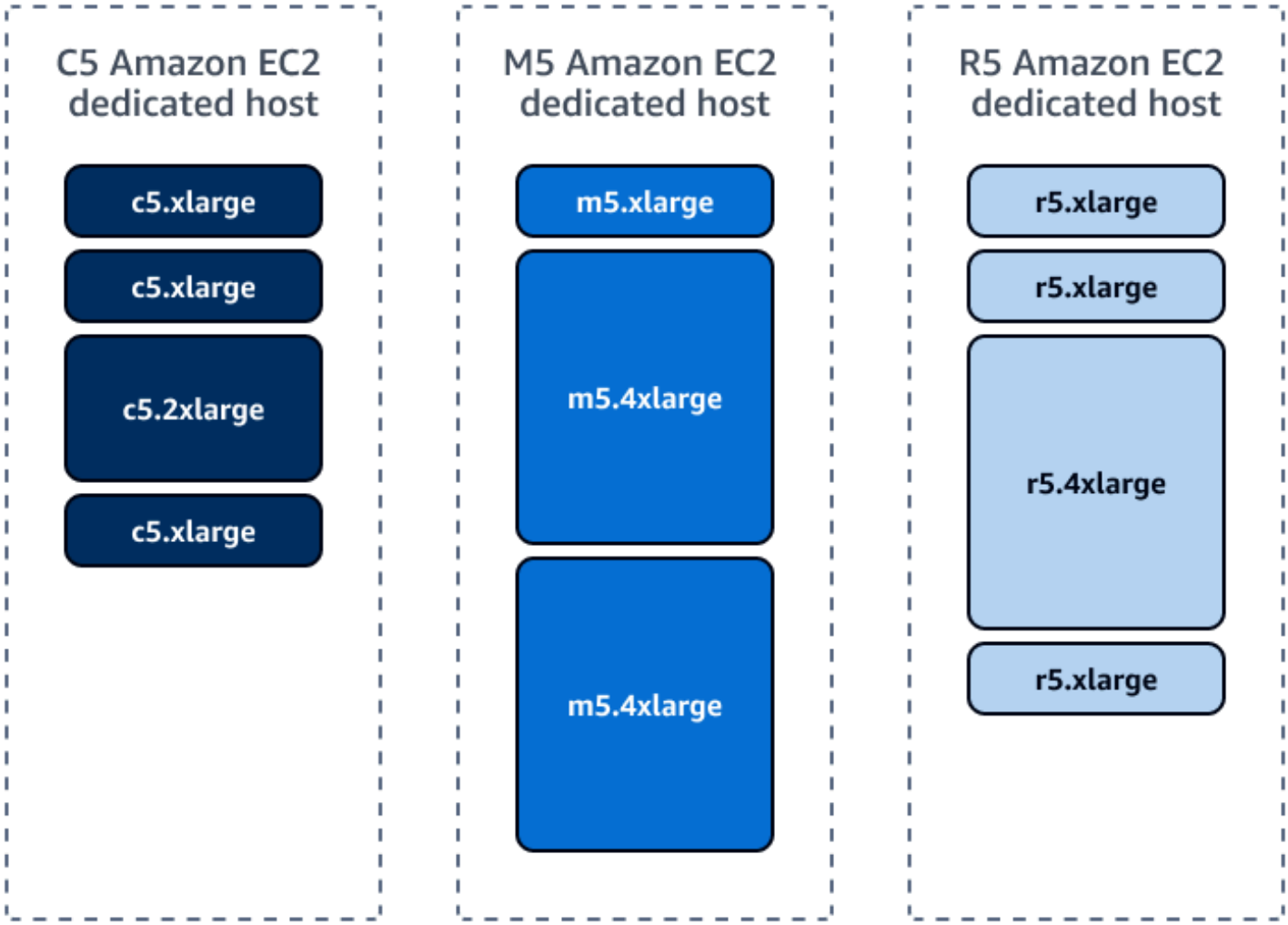
Por ejemplo:

- En el caso de xlarge (4 vCPUs), puede lanzar un máximo de 32 instancias m6i.xlarge en este host dedicado.
- En el caso de 8xlarge (32 vCPUs), puede lanzar un máximo de 4 instancias m6i.8xlarge en este host dedicado.
- En el caso de Metal (128 vCPUs), puede lanzar un máximo de 1 instancia m6i.metal en este host dedicado.

En el siguiente diagrama, se muestran las opciones de host dedicado para las instancias M6.

Hosts dedicados heterogéneos

Los hosts dedicados que admiten varios tamaños de instancia en el mismo host se denominan hosts EC2 dedicados de Amazon heterogéneos. En el siguiente diagrama, se muestra un ejemplo de hosts dedicados C5, M5 y R5 con varios tamaños de instancia, como 2xlarge, xlarge y 4xlarge.



Administración de hosts dedicados

Le recomendamos que tenga en cuenta lo siguiente en relación con la gestión de Amazon EC2 Dedicated Hosts:

- Para aprovechar al máximo los hosts dedicados, puede compartir un único host entre varias cuentas de su organización. El uso compartido de hosts permite la optimización de los recursos y puede suponer un ahorro de costes al utilizar todos los espacios disponibles en el host. Al compartir un host dedicado entre las unidades de negocio, puede centralizar su infraestructura de TI y mejorar la utilización de los recursos, sin dejar de mantener la separación entre las cargas de trabajo. Si forma parte de una organización AWS Organizations y el uso compartido está habilitado dentro de su organización, los consumidores de su organización tendrán acceso automático al host dedicado compartido. De lo contrario, los consumidores reciben una invitación para unirse al

recurso compartido y se les concede acceso al host dedicado compartido después de aceptar la invitación.

- Puede ejecutar Windows Server 2022 en hosts dedicados con el modelo de licencia incluida, ya que Windows Server 2019 es la última versión en la que se puede utilizar el BYOL. Si quiere usar Windows Server 2022 en hosts dedicados, debe usar instancias de Windows Server 2022 con licencia incluida.
- AWS License Manageres una solución integral para administrar licencias de software de varios proveedores en entornos locales AWS y locales. Si usa License Manager, puede obtener una mayor visibilidad y control sobre cómo se usan las licencias de software, lo que se traduce en ahorros de costos y mejora el cumplimiento. Puede usar License Manager para establecer reglas que emulen sus condiciones de licencia exclusivas. Esto le permite hacer cumplir esas reglas y evitar el uso indebido de las licencias. Esto puede reducir el riesgo de incumplimiento y mejorar los procesos de administración de licencias.
- Puede usar License Manager para automatizar la ubicación, la publicación y la recuperación de los hosts mediante grupos de recursos de hosts. Esto puede aumentar la productividad y reducir la sobrecarga de administración. License Manager también proporciona una vista centralizada del uso de las licencias en los entornos locales AWS y en función de las normas de licencia, lo que facilita la gestión de las compras incrementales de licencias, el cumplimiento y las auditorías de los proveedores en toda la organización. Además, License Manager se integra con AWS Organizations y AWS Resource Access Manager (AWS RAM) para compartir las configuraciones de licencia entre cuentas y regiones. Esto le permite crear informes para todo su entorno en función de un cronograma y gestionar las normas de licencia de forma centralizada y unificada Cuenta de AWS. En última instancia, esto puede mejorar la gobernanza y reducir la complejidad.
- Al diseñar una alta disponibilidad para hosts dedicados en una sola región, asegúrese de haber asignado un mínimo de dos hosts dedicados en un mínimo de dos zonas de disponibilidad para las cargas de trabajo críticas para la producción. Para obtener más información, consulte Amazon EC2 Dedicated Hosts para Microsoft Windows sobre la implementación de AWS referencia.
- Para cada familia de instancias de host dedicado, existe un límite en la cantidad de instancias que puede ejecutar para cada tamaño de instancia. Para obtener más información, consulte la tabla de configuración de hosts dedicados en la EC2 documentación de Amazon.

AWS opciones de licencia

Las licencias se clasifican en las siguientes categorías principales:

- Licencia incluida: esta opción de licencia le permite comprar y utilizar licencias a pedido, pagando únicamente por lo que utilice. Es ideal para los casos de uso en los que busca flexibilidad en el uso de las licencias y desea evitar costes iniciales. Puede elegir entre una variedad de productos de Windows Server, SQL Server y otros productos de Microsoft.
- Productos BYOL con movilidad de licencias: si ya tiene licencias existentes y desea usarlas en la nube, esta opción de licencia le permite llevar sus propias licencias a la nube a través del programa Microsoft License Mobility. Los productos con movilidad de licencias, como SQL Server with Software Assurance (SA), se pueden alquilar de forma compartida o exclusiva. Esto reduce los costos de las AWS instancias.
- Productos BYOL sin movilidad de licencias: para los productos de Microsoft, como Windows Server, que carecen de movilidad de licencias, AWS ofrece opciones específicas para usar estos productos en la nube. Además, los hosts dedicados permiten conceder licencias a nivel del núcleo físico, lo que permite ahorrar un 50 por ciento o más en las licencias necesarias para ejecutar sus cargas de trabajo. Los hosts dedicados son una opción excelente para cargas de trabajo estables y predecibles que se ejecutan la mayor parte del tiempo.

Incorpora licencias de Windows Server

Incorporar sus propias licencias de Windows es una de las estrategias más eficaces para la optimización de las licencias, ya que le permite aprovechar las inversiones existentes y reducir sus AWS gastos. Los escenarios BYOL específicos no requieren beneficios de SA o movilidad de licencias, pero siempre es necesaria una infraestructura EC2 dedicada de Amazon. Para cumplir los requisitos, debe haber adquirido las licencias perpetuas antes del 1 de octubre de 2019 o haberlas añadido a modo de prueba en el marco de una inscripción empresarial activa en vigor antes del 1 de octubre de 2019. En estos escenarios específicos de BYOL, solo puede actualizar las licencias a versiones disponibles antes del 1 de octubre de 2019. Por ejemplo, si dejó de usar SA en 2017, solo tiene los derechos para implementarlas hasta Windows Server 2016, no hasta 2019. Sin embargo, 2019 es la última versión apta para BYOL. AWS Para obtener más información, consulte Licencias: Windows Server en la AWS documentación.

La incorporación de licencias puede afectar significativamente al costo de ejecutar las cargas de trabajo de Microsoft. AWS Si utilizas tus propias licencias, no estás obligado a pagar costes de licencia adicionales por las instancias que se ejecutan en la nube, lo que puede suponer un ahorro de costes considerable.

En la siguiente tabla se muestra el coste mensual bajo demanda que supone ejecutar una única instancia de c5.xlarge las 24 horas del día, los 7 días de la semana, en varias configuraciones.

Configuración	Coste mensual (USD)
Edición Windows Server + SQL Server Enterprise	1.353,00$ (LI)
Edición estándar de Windows Server y SQL Server	609,00$ (LI)
Solo Windows Server	259,00$ (LI)
Solo computación (Linux)	127,00$

Puede usar sus licencias actuales para reducir los costos de licencia y ahorrar dinero en su factura total AWS .

Para poder optar a BYOL en Amazon EC2 Dedicated Hosts, debe traer sus propias licencias de software, como las de Windows Server y SQL Server. BYOL le permite utilizar sus licencias existentes AWS y, por lo tanto, puede ahorrarle costes. Para disponer de sus propias licencias, debe contar con los derechos de licencia del proveedor del software y, además, debe proporcionar el soporte de instalación o la imagen del software. El medio de instalación o la imagen se pueden usar para lanzar instancias en hosts dedicados. Para obtener más información sobre la creación de una AMI BYOL, consulte Cómo crear Windows Server Bring-Your-Own-License AMIs desde un entorno local con VM Import/Export en el blog Microsoft Workloads on. AWS

> **Note**
>
> Un tipo de licencia establecido en Automático equivale a una opción con licencia incluida.AWS Esta opción puede provocar gastos a pedido no deseados. Tendrá que cambiar los tipos de licencia.

Escenarios de optimización de costes

El tamaño correcto y la optimización de las licencias son un componente clave de la optimización de costes en AWS. Si implementa las estrategias correctas, puede reducir los costos de licencias, mantener el cumplimiento y obtener el mejor valor posible de su inversión en licencias mediante Amazon EC2 Dedicated Hosts y la opción BYOL.

En esta sección se describen los siguientes escenarios de ejemplo:

- Ahorro de costes con los hosts dedicados T3
- Comparación de la tenencia compartida con los hosts dedicados con SQL Server BYOL
- Implementaciones de SQL Server de alta disponibilidad

Ahorro de costes con los hosts dedicados T3

Los hosts dedicados T3 se diferencian de otros hosts EC2 dedicados de Amazon que tradicionalmente proporcionan recursos de CPU fijos. Los hosts dedicados T3, por el contrario, admiten instancias en ráfagas que son capaces de compartir los recursos de la CPU, proporcionar un rendimiento básico de la CPU y reproducirse en ráfagas cuando es necesario. Compartir los recursos de CPU, también conocido como sobresuscripción, es lo que permite que un único host dedicado T3 admita hasta cuatro veces más instancias que los hosts dedicados de uso general comparables.

Los hosts dedicados T3 generan un menor TCO al ofrecer una mayor densidad de instancias que cualquier otro host EC2 dedicado de Amazon. Las instancias T3 Burstable le permiten consolidar un mayor número de instancias con un uso low-to-moderate medio de la CPU en un menor número de hosts que nunca. Los hosts dedicados T3 también ofrecen tamaños de instancia más pequeños en un mayor número de combinaciones de vCPU y memoria que otros hosts dedicados de EC2 Amazon. Los tamaños de instancia más pequeños pueden contribuir a reducir el TCO y ayudar a ofrecer índices de consolidación equivalentes o superiores a los de los hosts locales.

Los hosts dedicados T3 son los más adecuados para ejecutar software BYOL con uso de low-to-moderate CPU y licencias de software elegibles por socket, por núcleo o por máquina virtual, incluidas las bases de datos Microsoft Windows Desktop, Windows Server, SQL Server y Oracle.

Utilice los hosts dedicados T3 para reducir las licencias de los centros de datos de Windows Server (por núcleo)

En los entornos locales, está aprovechando el hecho de que puede sobresuscribir fácilmente sus VMware hosts físicos CPUs en exceso y lograr altos niveles de consolidación.

Considere el siguiente ejemplo. Actualmente utiliza VMware hosts de 10 x 36 núcleos y 384 GB de RAM en un entorno local. Además, cada host ejecuta máquinas virtuales Windows Server de 96x2 vCPU y 4 GB de RAM con un uso medio de CPU bajo.

Ahora puede lograr niveles de consolidación mucho más altos si mueve sus máquinas virtuales a hosts dedicados T3, que tienen el doble de RAM en comparación con los hosts locales actuales.

VMware Puede ejecutar la misma cantidad de servidores en hosts dedicados T3 con un 50 por ciento menos de costo de alojamiento. Esto puede ayudarle a reducir los costes de licencias de Windows Server en un 33 por ciento. En la siguiente tabla se destacan los ahorros derivados del uso de hosts dedicados T3.

	Hosts locales VMware	Hosts dedicados T3	Ahorro
Servidores físicos	10	5	
Núcleos físicos por host	36	48	
RAM por host (GB)	384	768	
2 vCPU, 4 GB de RAM por host VMs	96	192	
Número total de VMs	960	960	
Número total de licencias de centros de datos de Windows Server (por núcleo) = (número de servidores * recuento de núcleos físicos)	10 * 36 = 360	5 * 48 = 240	33%

Comparación entre la tenencia compartida y los hosts dedicados con SQL Server BYOL

Considere un ejemplo práctico para demostrar el valor de Amazon EC2 Dedicated Hosts. En este escenario, una organización ejecuta una carga de trabajo de SQL Server en un entorno local con 240 núcleos y desea implementar la misma carga de trabajo de forma rentable. AWS Si esta organización ofrece sus propias licencias (BYOL), seguirá pagando por el SA y reducir el número de núcleos repercute directamente en sus costes.

En el siguiente diagrama se comparan los AWS ahorros entre los derechos de Microsoft y los de SQL Server.

Microsoft entitlements (Enterprise Agreements)	
	Number of cores
SQL Server Enterprise edition	208
SQL Server Standard edition	32
Total SA cost	**$341,000**

SQL Server savings with AWS	
AWS shared vCPUs	AWS BYOL/Dedicated Hosts cores
120	96
20	-
$197,418	**$151,355**

Al dimensionar correctamente las instancias en régimen de arrendamiento AWS compartido, puede reducir las licencias de SQL Server a 140 núcleos. Esto se traduce en un coste de SA de 197 000$.

Los hosts EC2 dedicados de Amazon le permiten licenciar SQL Server a nivel de núcleo físico. Esto no es posible en la tenencia compartida, donde las licencias de SQL Server se basan en la cantidad de v CPUs asignada a la instancia. Por lo tanto, si utiliza dos hosts dedicados R5 con 48 núcleos cada uno, solo necesitará cubrir 96 núcleos en lugar de los 140 v que CPUs se requieren en el arrendamiento compartido. Al implementar hosts dedicados R5 y licenciar las cargas de trabajo a nivel físico, puede reducir la cantidad requerida de licencias de la edición SQL Server Enterprise a 96 núcleos. Esto significa que puede implementar hasta 192 núcleos (teniendo en cuenta los subprocesos múltiples) de cargas de trabajo de SQL Server y, al mismo tiempo, cumplir con los requisitos de licencia y lograr importantes ahorros de costos.

En este caso, la organización paga aproximadamente 341 000$ al año en concepto de costes de SA. Tras dimensionar correctamente el arrendamiento compartido, reducen los costos a 197 000 dólares con 140 v. CPUs Amazon EC2 Dedicated Hosts reduce aún más los costes hasta alcanzar los 151 000$ (una disminución de aproximadamente un 56 por ciento).

Implementaciones de SQL Server de alta disponibilidad

En este ejemplo, se analiza cómo el costo puede influir en una implementación de SQL Server AWS teniendo en cuenta diversas consideraciones de licencia. Supongamos que una organización necesita implementar seis servidores de SQL Server Enterprise AWS para admitir tres aplicaciones. Estos servidores requieren una alta disponibilidad y tienen 16 V CPUs y 256 GB de RAM cada uno. Consulte los siguientes detalles del escenario:

- Servidor: SQL Server
- Edición de sistema operativo: Windows Server Datacenter 2019
- Edición SQL Server: SQL Server Enterprise 2019
- vCPU: 16
- Memoria (GB): 256

- Cantidad — 6

Para optimizar los costes AWS sin sacrificar el rendimiento, le recomendamos que asigne el tamaño adecuado a las instancias en función del uso de la CPU, la memoria, la red y el disco (IOPS/BW). Tras dimensionar correctamente las cargas de trabajo, colóquelas en el tipo de instancia x2iedn.4xlarge, que ofrece 16 v. CPUs Sin embargo, este tipo de instancia también incluye el doble de la memoria necesaria para las cargas de trabajo. Aún es posible realizar una mayor optimización.

Escenario 1

Una organización implementa seis servidores SQL Server Enterprise en régimen de arrendamiento AWS compartido mediante la opción de licencia incluida para Windows y SQL Server. Con esta opción, el costo de las licencias de Windows y SQL Server se incorpora al precio de la instancia. Consulte los siguientes detalles del escenario:

- Arrendamiento compartido (instancia): x2iedn.4xlarge
- Coste por hora (USD): 10.0705$
- Coste mensual por unidad (USD): 7.351,47$
- Número de servidores: 6
- CPU: 16
- Memoria: 512
- Coste mensual de 6 servidores: 44.108 dólares

Escenario 2

Una organización tiene SA y BYOL para SQL Server en régimen de arrendamiento compartido. Esto significa que la organización usa la opción con licencia incluida para Windows, pero proporciona sus propias licencias de SQL Server en función del número de v CPUs asignado a la instancia. Como la organización tiene seis servidores SQL Server Enterprise con 16 v CPUs cada uno, CPUs se requiere un total de 96 v. Consulte los siguientes detalles del escenario:

- Arrendamiento compartido (instancia): x2iedn.4xlarge
- Coste por hora (USD): 4.0705$
- Coste mensual por unidad (USD): 2971,47$
- Número de servidores: 6
- CPU: 16

- Memoria: 512
- Núcleos BYOL: 96
- Coste mensual de 6 servidores: 17.828 dólares

Al incorporar sus propias licencias de SQL Server con SA, la organización en este escenario puede ahorrar costos en comparación con el uso de la opción de SQL Server con licencia incluida. El ahorro de costes preciso depende del precio y las condiciones del acuerdo de licencia específico. En este escenario, AWS los costes se reducen en 26 280$ al mes si se cambian las licencias de SQL Server Enterprise a. AWS

Escenario 3

Una organización tiene BYOL tanto para Windows como para SQL Server en Amazon EC2 Dedicated Hosts. Esto significa que la organización asignará las licencias a nivel de núcleo físico, lo que le permitirá licenciar únicamente los núcleos físicos del anfitrión. La concesión de licencias a nivel de núcleo físico le permite implementar el número máximo de instancias sin que ello afecte a las licencias requeridas. Este modelo de licencia se suele utilizar con las ediciones Windows Server Datacenter y SQL Server Enterprise.

En este escenario se utilizan dos hosts dedicados de Amazon EC2 X2iEZN. Cada host tiene 24 núcleos físicos y 48 V. CPUs Esto proporciona una capacidad adecuada para los seis servidores SQL Server Enterprise con 16 v CPUs y 256 GB de RAM cada uno. Consulte los siguientes detalles del escenario:

- Número de hosts dedicados: 2
- Familia de instancias: x2iezn
- Coste por hora (USD): 11,009 USD
- Coste mensual por unidad (USD): 8.036 dólares
- Núcleo físico: 48
- vCPU disponible: 96
- Se requieren 24 licencias principales de Windows Server
- Se requieren licencias para los núcleos de SQL Server Enterprise: 24
- Coste mensual: 16.073

El costo total de dos hosts dedicados de EC2 Amazon de la familia X2iEZN es de 16.073 USD al mes. Para obtener más información sobre los precios, consulte la estimación para este escenario.

Calculadora de precios de AWS En este escenario, la organización puede ahorrar 1.755,65$ al mes si adquiere sus licencias de Windows. Si utilizan Amazon EC2 Dedicated Hosts, también pueden reducir el número de licencias de SQL Server necesarias. En régimen de arrendamiento compartido, necesitarían 96 licencias de SQL Server Enterprise para cubrir los seis servidores de SQL Server Enterprise de 16 V CPUs cada uno. Sin embargo, al utilizar Amazon EC2 Dedicated Hosts y licenciar a nivel de núcleo físico, pueden reducir el número de licencias necesarias a 48 núcleos.

En los siguientes detalles se comparan los costes del ejemplo 3 y se muestra cuánto puede ahorrar al implementar cargas de trabajo en Amazon EC2 Dedicated Hosts con la opción BYOL en comparación con otros escenarios.

- Servidor local: SQL Server
- vCPU: 16
- Memoria: 256
- Número de servidores: 6
- Coste mensual para el escenario 1: Windows (LI) + SQL Server Enterprise (LI): 44.108 dólares
- Coste mensual para el escenario 2: Windows (LI) + SQL Server Enterprise (BYOL): 17.828$
- Coste mensual para el escenario 3: Windows (LI) + SQL Server Enterprise (BYOL) en Amazon EC2 Dedicated Host: 16.073 $

> **Note**
>
> El costo se basa en los precios bajo demanda. Puede reducir aún más los costos mediante el uso de Savings Plans o Dedicated Reserved Instances. Estas opciones ofrecen un modelo de precios flexible con importantes ahorros de costos en comparación con los precios bajo demanda. Con estos planes, puedes comprometerte a un plazo de uno o tres años. Para obtener más información, consulta la EC2 sección Optimizar el gasto para Windows en Amazon de esta guía.

Considera las siguientes opciones de pago para Amazon EC2 Dedicated Hosts:

- Hosts dedicados (EC2 documentación de Amazon)
- Reservas para anfitriones dedicados (EC2documentación de Amazon)
- Savings Plans (EC2 documentación de Amazon)

Calculadora de precios de AWSAhora es compatible con los precios de alojamiento dedicado. Esto puede ayudarle a elegir el host dedicado subyacente adecuado.

Recomendaciones de optimización de costos

Le recomendamos que siga los siguientes pasos para optimizar sus costes mediante AWS Cost Explorer:

1. Habilitar Cost Explorer.
2. Use Cost Explorer para ver y analizar los costos y el uso de sus implementaciones de Amazon EC2 Dedicated Host.
3. Valide que está ejecutando BYOL. Puedes mostrar los siguientes detalles de la plataforma y los valores de las operaciones de uso en las instancias o páginas de AMI de la EC2 consola de Amazon, o en la respuesta que devuelve el `describe-instances` comando `describe-images` o.
   - Detalles de la plataforma: Windows, operación de uso ::0002 RunInstances (licencia incluida)
   - Detalles de la plataforma: Windows BYOL, Operación de uso: :0800 RunInstances

Recursos adicionales

- Tipos de licencia aptos para la conversión de tipos de licencia (documentación)AWS License Manager
- AWS License Manager y taller exclusivo para anfitriones (AWS License Manager taller)
- Amazon EC2 Dedicated Hosts FAQs (documentación de AWS)
- Cómo crear Windows Server Bring-Your-Own-License AMIs de forma local con VM Import/Export (Microsoft Workloads en el blog) AWS
- Importación/exportación de máquinas virtuales (documentación)AWS
- Amazon Web Services y Microsoft: Preguntas frecuentes (AWS documentación)
- Conversiones de tipos de licencia en License Manager (AWS License Manager documentación)
- Implementación de SQL Server de alta disponibilidad en hosts EC2 dedicados de Amazon (blog sobre operaciones y migraciones AWS en la nube)

Optimice el gasto para Windows en Amazon EC2

Descripción general

Una de las principales preocupaciones a la hora de migrar servidores AWS son los costes de infraestructura. Es cierto que uno de los beneficios de la nube es pagar por los recursos a pedido, pero hay cargas de trabajo de producción que deben estar disponibles las 24 horas del día, los 7 días de la semana, los 365 días del año. Los Savings Plans están diseñados para ahorrar dinero en su AWS uso estable en todos los EC2 casos AWS Lambda, y. AWS Fargate

Los Savings Plans ofrecen un modelo de precios flexible y pueden ayudarle a reducir los precios del uso de Amazon EC2, Fargate, Lambda y Amazon SageMaker AI a cambio de un compromiso con una cantidad constante de uso (por ejemplo, 10 USD por hora). Te comprometes a una cantidad constante de gastos informáticos por hora durante uno o tres años y, a cambio, recibes un descuento por ese uso.

Puede elegir entre tres opciones de pago diferentes con Savings Plans:

- La opción sin pago por adelantado no requiere ningún pago por adelantado y su compromiso se cobra únicamente de forma mensual.
- La opción Partial Upfront ofrece precios más bajos en los Savings Plans. Se le cobrará al menos la mitad de su compromiso por adelantado y el resto se cobrará mensualmente.
- La opción All Upfront ofrece los precios más bajos y todo el compromiso se cobra en un solo pago.

Puede hacer un seguimiento de los vencimientos de sus Savings Plans y de los próximos Savings Plans en cola en. AWS Cost Explorer Puede usar las alertas de Savings Plans para recibir alertas anticipadas por correo electrónico 1, 7, 30 o 60 días antes de la fecha de vencimiento de su plan, o cuando haya un compromiso pendiente de compra. Estas notificaciones también le avisan sobre la fecha de vencimiento. Puedes enviar notificaciones a un máximo de 10 destinatarios de correo electrónico.

Descripción de Savings Plans

Cada tipo de uso de cómputo tiene una tarifa bajo demanda y una tarifa de Savings Plans. Si te comprometes a usar el cómputo de 10$ por hora, obtendrás precios de Savings Plans para todo uso de hasta 10$ a la tarifa de Savings Plans. Cualquier uso que supere el compromiso de gasto informático se cobrará según las tarifas habituales bajo demanda. Puede empezar a utilizar Savings Plans utilizando Cost Explorer en el AWS Management Console.

Puede comprometerse fácilmente con Savings Plans utilizando las recomendaciones que se proporcionan en Cost Explorer para obtener los mayores ahorros. El compromiso por hora recomendado se basa en su historial de uso bajo demanda y en el tipo de plan, duración y opción de pago que haya elegido. Savings Plans se aplica primero a la cuenta que compró el plan y, después, se comparte con otras cuentas de la familia de facturación unificada.

> **Note**
>
> La opción de compartir en Savings Plans AWS Organizations está habilitada de forma predeterminada. Puedes rechazar esta opción en la AWS Billing consola de la cuenta del pagador. Puedes visitar la página de recomendaciones para ver los Savings Plans que AWS recomiendan ayudarte a ahorrar en el uso que cumpla con los requisitos. Estas recomendaciones se pueden actualizar en cualquier momento para facilitarle la compra de los Savings Plans óptimos.

Savings Plans para computación

Compute Savings Plans ofrece la mayor flexibilidad y ayuda a reducir los costos. Estos planes se aplican automáticamente al uso de las EC2 instancias, independientemente de la familia de instancias, el tamaño, la zona de disponibilidad, la región, el sistema operativo o el arrendamiento. También se aplican al uso de Fargate o Lambda. Por ejemplo, con Compute Savings Plans, puedes cambiar de instancias C4 a M5, transferir una carga de trabajo de la UE (Irlanda) a la UE (Londres) o mover una carga de trabajo de EC2 Fargate o Lambda en cualquier momento. Seguirá pagando automáticamente el precio de Savings Plans.

EC2 Instance Savings Plans

EC2 Los Instance Savings Plans ofrecen los mayores descuentos a cambio de un compromiso de uso de familias de instancias individuales en una región (por ejemplo, comprometiéndose a un nivel constante de uso de M5 en el norte de Virginia). Esto le proporciona automáticamente descuentos en el precio bajo demanda de la familia de instancias seleccionada en esa región, independientemente de la zona de disponibilidad, el tamaño, el sistema operativo o el arrendamiento. EC2 Los Instance Savings Plans le permiten cambiar el uso entre las instancias de una familia de esa región. Por ejemplo, puede pasar de c5.xlarge con Windows a c5.2xlarge con Linux y beneficiarse automáticamente de los precios de Savings Plans.

Tanto los planes Compute como los EC2 Instance Savings se aplican a las EC2 instancias que forman parte de los clústeres de Amazon EMR, Amazon Elastic Kubernetes Service (Amazon EKS)

y Amazon Elastic Container Service (Amazon ECS). Savings Plans no cubre los cargos de Amazon EMR, Amazon EKS y Amazon ECS, pero sí las EC2 instancias subyacentes. EC2 Los Instance Savings Plans se aplican antes que Compute Savings Plans porque los Compute Savings Plans tienen una aplicabilidad más amplia.

> **Note**
>
> No puede cambiar un Savings Plan fácilmente después de haber asumido un compromiso. Le recomendamos que planifique detenidamente antes de comprometerse con cualquiera de las opciones de Savings Plans. Los Savings Plans ofrecen precios más bajos en comparación con los precios bajo demanda a cambio de un compromiso y no se pueden cancelar durante el plazo.

Ejemplo de compromiso por hora

Si compra un Savings Plan, asume un compromiso monetario por hora durante el plazo del plan. Si te comprometes a utilizar 10$ la hora de cómputo, el precio del Savings Plan se aplicará automáticamente a todo el uso, hasta un máximo de 10$ por hora. Cualquier uso que supere el compromiso se cobrará según las tarifas habituales bajo demanda. Puede usar la herramienta de recomendaciones de compra de Savings Plans en Cost Explorer para obtener compromisos recomendados que pueden maximizar sus ahorros. El compromiso financiero por hora de un plan específico no se puede modificar durante la vigencia del plan. Si desea aumentar su compromiso después de analizar el uso, puede adquirir un Savings Plan adicional para cubrir el exceso de uso.

Beneficios de los Savings Plans

En comparación con las instancias reservadas, los Savings Plans ofrecen un modelo de precios más flexible que le permite ahorrar dinero y, al mismo tiempo, aprovechar la amplia selección de opciones informáticas que ofrecen los Savings Plans. Los Savings Plans ofrecen descuentos, incluso a medida que cambien sus necesidades informáticas. Esto puede ayudarlo a mantenerse al día con su entorno dinámico en constante cambio sin incurrir en gastos de administración adicionales. Estos son algunos otros beneficios de usar Savings Plans:

- Fácil de usar: reciba descuentos automáticos a cambio de un compromiso monetario.
- Flexibilidad: un compromiso único que se aplica a varios tipos de uso.
- Ahorros potenciales: hay varias formas de ahorrar. Considere los siguientes ejemplos:

- Ahorro del 60 por ciento en cargas de trabajo de Windows Server con Compute Savings Plans (d2.8xlarge, 3 años, todo por adelantado, Windows, arrendamiento compartido, us-east-2)
- Ahorro del 73 por ciento en cargas de trabajo de Windows Server con EC2 Instance Savings Plans (d2.8xlarge, 3 años, todo por adelantado, Windows, arrendamiento compartido, us-east-2)
- Entre un 28 y un 41 por ciento de ahorro en tipos de instancias no exóticas (familia t3, 3 años, todo por adelantado, ventanas, arrendamiento compartido, us-east-2)
- Entre un 25 y un 40 por ciento de ahorro promedio en servidores Windows

> **Note**
>
> EC2 Los Instance Savings Plans ofrecen un descuento mayor que los Compute Savings Plans debido a su flexibilidad reducida. Te comprometes a usarlo a un precio reducido.

Cada tipo de uso de cómputo tiene una tarifa del Plan de Ahorros y una tarifa bajo demanda. La siguiente tabla muestra los Savings Plans y las tarifas bajo demanda para cada tipo de sistema operativo. Se le cobrarán las tarifas de Savings Plans por el uso comprometido y cualquier uso que supere el compromiso se cobrará según las tarifas normales bajo demanda.

Nombre de instancia	Tarifa de Savings Plans	Ahorros a pedido	Tarifa bajo demanda	Sistema operativo	Región	Opción de pago	Duración del plazo
x2iedn.xlarge	0,32$	61%	0,83$	Linux	Este de EE. UU. (Norte de Virginia)	Sin pago inicial	3
x2iedn.xlarge	2,01 DÓLARES	50%	1,02 DÓLARES	Windows	Este de EE. UU. (Norte de Virginia)	Sin pago inicial	3
x2iedn.xlarge	1,02$	20%	2,52 DÓLARES	Incluye licencia de	Este de EE. UU.	Sin pago inicial	3

Nombre de instancia	Tarifa de Savings Plans	Ahorros a pedido	Tarifa bajo demanda	Sistema operativo	Región	Opción de pago	Duración del plazo
				Windows y edición SQL Server Enterprise	(Norte de Virginia)		
x2iedn.xlarge	0,32$	61%	0,83$	BYOL	Este de EE. UU. (Norte de Virginia)	Sin pago inicial	3

Los Savings Plans incluyen el sistema operativo y tienen un descuento aparte para BYOL. Todos están desglosados en la calculadora Compute Savings Plans.

Modelo de precios de instancias reservadas

AWS tiene otro modelo de precios basado en el compromiso, conocido como instancias reservadas. Este modelo puede resultar problemático si su proceso cambia después de haber asumido un compromiso, lo que provoca que las instancias reservadas no se utilicen. Los Savings Plans están diseñados para ofrecer reducciones de costos similares a las de las instancias reservadas estándar y convertibles, pero con una flexibilidad mucho mayor. Compute Savings Plans ofrece precios más bajos en el EC2 uso de las instancias, independientemente de la familia, el tamaño, el sistema operativo, el arrendamiento o la región de la instancia. También permiten la máxima flexibilidad.

La siguiente tabla puede ayudarle a elegir entre Savings Plans o Reserved Instances.

	Reserved Instance	EC2 Instance Savings Plans	Savings Plans para computación
Descuento medio de 1 año	Hasta un 38%	Hasta un 29%	Hasta un 29%

	Reserved Instance	EC2 Instance Savings Plans	Savings Plans para computación
Descuento medio de 3 años	Hasta un 58%	Hasta un 73%	Hasta un 60%
Familia de instancias	Fixed	Fixed	Flexible
Tamaño de instancia	Fijo (no Linux)	Flexible	Flexible
Geography	1 región	1 región	Flexible
Sistema operativo	Fixed	Flexible	Flexible
Servicio	Amazon EC2 o Amazon RDS	Amazon EC2	Amazon EC2, Fargate, Lambda
Opciones de pago	Todo, parcial, sin anticipar	Todo, parcial, no anticipado	Todo, parcial, no anticipado
Límites de instancias	20 por zona de disponibilidad	Sin límite	Sin límite

> **Note**
>
> Los Savings Plans funcionan ofreciéndote un discount basado en un compromiso monetario por hora. El compromiso financiero por hora no se puede cancelar ni cambiar durante la vigencia de su plan, pero puede comprar Savings Plans adicionales para cubrir el uso adicional. Esto le permite mantener un compromiso horario constante a medida que su flota crece.

Puede utilizar herramientas como AWS Cost Explorerlos paneles de Nube de AWS inteligencia para hacer un seguimiento de su compromiso. Cost Explorer proporciona una línea objetivo de cobertura que puede ayudar a su organización a planificar su estrategia de cobertura de Savings Plans. Si el 75 por ciento de su carga de trabajo es estable, entonces el 75 por ciento es un buen objetivo. Esto deja el 25 por ciento del gasto a demanda o variable en función de las cargas de trabajo dinámicas.

Si necesita aumentar esa cobertura al 85 por ciento, puede comprar otro compromiso de Savings Plans para aumentar el compromiso monetario por hora.

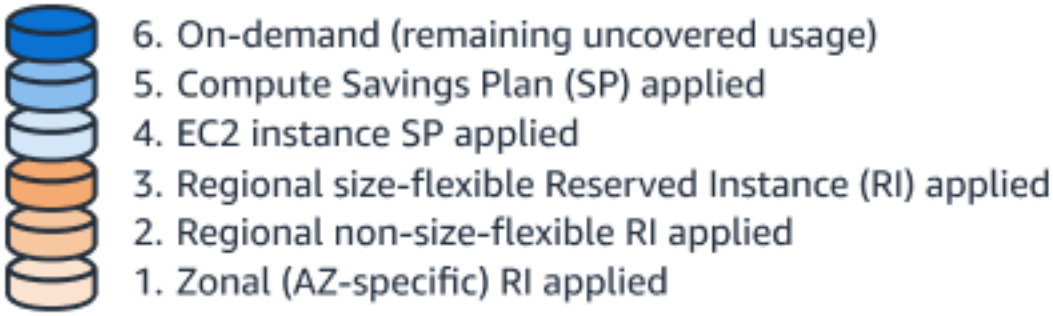
> **Note**
>
> Le recomendamos que compre Savings Plans en lugar de Reserved Instances, pero los dos modelos de compromiso pueden funcionar juntos si ya ha adquirido Reserved Instances.

Considere un ejemplo en el que compró una instancia reservada, pero quiere empezar a probar una opción de Savings Plans. Es lógico que esta combinación se aplique a su facturación final. Esta es una jerarquía que puedes aplicar a tu Cuentas de AWS:

1. La instancia reservada zonal se aplica a la cuenta que la posee. Si a una instancia reservada le quedan horas, se aplica al resto de la organización.
2. Las instancias reservadas regionales para Windows, que no son flexibles en cuanto al tamaño, se aplican al mismo uso en la cuenta propietaria. Todo lo que quede se transfiere al resto de la organización.
3. Las instancias reservadas regionales con flexibilidad de tamaño se aplican a la cuenta propietaria (primero la instancia más pequeña de la familia y luego a las instancias más grandes) y, después, al resto de la organización.
4. Las instancias reservadas regionales se aplican a cualquier reserva de capacidad bajo demanda no utilizada.
5. EC2 Los Instance Savings Plans se aplican a la cuenta que los compró.
6. Compute Savings Plans se aplica a la cuenta que lo compró.

> **Note**
>
> Los descuentos comienzan con el uso que se traduce en el descuento más alto y luego van bajando hasta el descuento más bajo. Tradicionalmente, las instancias de Windows tienen un potencial de descuento más bajo que las de Linux para los tipos de instancias más comunes (por ejemplo, T3, M6 y C5). Esto significa que las instancias de Linux se benefician más que las instancias de Windows en la mayoría de los casos.

El siguiente gráfico muestra el precio después de dividir Reserved Instances de Savings Plans. Tanto el Compute como el EC2 Instance Savings Plans se aplican primero a las instancias en ejecución y, después, a las reservas de capacidad bajo demanda no utilizadas.



Escenarios de optimización de costes

En esta sección se describen los escenarios de optimización de costes para los hosts EC2 dedicados de Amazon y EC2 las instancias de Amazon que utilizan un modelo de facturación con licencia incluida.

Hosts EC2 dedicados de Amazon

Considere un escenario en el que vaya a migrar sus cargas de trabajo de Windows locales a. AWS Su centro de datos tiene los siguientes servidores:

- Dos servidores con 16 vCPU y 128 GB de RAM
- Dos servidores con 32 vCPU y 164 GB de RAM
- Un servidor con 8 vCPU y 64 GB de RAM
- 16 servidores con vCPU y 32 GB de RAM

Además, suponga que puede traer su propia licencia AWS porque tiene suficientes licencias para traerla. En la siguiente tabla se muestran las instancias de servidor en las que puede utilizarlas AWS.

Tipo de instancia	CPU	RAM	Amount
r5.4xlarge	16	128	2
r5.8xlarge	32	256	2
r5.2xlarge	8	64	1

Tipo de instancia	CPU	RAM	Amount
r5.xlarge	4	32	16
			21

Un análisis muestra que estas 21 máquinas virtuales se pueden distribuir en dos hosts dedicados con un host de la familia de instancias R5. En la siguiente tabla se muestra el costo de estos dos hosts dedicados.

Escenario de host dedicado bajo demanda	Pago por adelantado	1 mes	1 año	3 años	Calculadora de precios de AWS
Bajo demanda	Ninguno	10.123 DÓLARES	121.475 DÓLARES	364.392 DÓLARES	Calculadora de precios de AWS estimación
Plan de Ahorros de 1 año	Ninguno	7.447$	89.362 DÓLARES	–	Calculadora de precios de AWS estimación
Plan de Ahorros de 3 años	Ninguno	5.476$	65.712 DÓLARES	197.128 DÓLARES	Calculadora de precios de AWS estimación
Savings Plan de 3 años con pago por adelantado	84.438$	2.755 DÓLARES	117.499 DÓLARES	183.618 DÓLARES	Calculadora de precios de AWS estimación

Si tiene los servidores a los que quiere migrar AWS, el precio final de un Plan de Ahorros de 1 año es de 89.362$, en lugar de 121.475$ del precio bajo demanda. Esto representa un descuento del 26,5 por ciento después de un año. Si está pensando en quedarse en casa AWS por un período más largo, puede elegir el Savings Plan de 3 años para ahorrar costos aún más. Al cabo de tres años, pagas 197.128 dólares en lugar de 364.392 dólares. Esto se traduce en un ahorro del 46 por ciento del importe total al cabo de tres años.

EC2 Instancias de Amazon con licencias incluidas

Considere un escenario en el que vaya a migrar una sola aplicación de tres niveles y desee utilizar las licencias que proporciona. AWS AWS Además, suponga que su aplicación funciona con los siguientes servidores:

- Dos servidores web con dos v CPUs y 4 GB de RAM
- Dos servidores de aplicaciones con ocho v CPUs y 16 GB de RAM
- Dos servidores de bases de datos con 16 v CPUs y 64 GB de RAM (con la edición SQL Server Standard)

En la siguiente tabla se muestran las instancias de servidor en las que puede utilizarlas AWS.

Tipo de instancia	CPU	RAM	Amount
c5.large	2	4	2
c5.2xlarge	8	16	2
r5.2xlarge	8	64	2
			6 servidores

La siguiente tabla muestra el costo de estos servidores en AWS.

Licencia incluida por AWS	Pago por adelantado	1 mes	1 año	3 años	Calculadora de precios de AWS
Bajo demanda	Ninguno	3.912 DÓLARES	46.950 DÓLARES	140.849 DÓLARES	Calculadora de precios de AWS estimación
Plan de Ahorros de 1 año	Ninguno	3.466$	41.952 DÓLARES		Calculadora de precios de AWS estimación
Savings Plan de 3 años sin pago por adelantado	Ninguno	3.189$	38.264 DÓLARES	114.804 DÓLARES	Calculadora de precios de AWS estimación
Savings Plan de 3 años con pago por adelantado	112.110$	Ninguno	Ninguna	Ninguno	Calculadora de precios de AWS estimación

Si desea utilizar estos servidores para entornos de producción (24 horas al día, 7 días a la semana) con precios bajo demanda, tendrá que pagar un coste mensual de 3.912$. El pago de este costo mensual equivale a 46.950 dólares al cabo de un año y a un total de 140.849 dólares al cabo de tres años.

Si eliges el Plan de Ahorros de 1 año sin pago por adelantado, el costo mensual se reduce a 3.466$. Al final del primer año, pagas 41.952 dólares. Se trata de un descuento total del 11 por ciento. Si eliges el Plan de Ahorros de 3 años sin pago por adelantado, el costo mensual se reduce a $3,189. Al cabo de tres años, usted paga 114.804 dólares. Eso le da un ahorro del 18,5 por ciento.

Recomendaciones de optimización de costos

Ambos escenarios le ayudan a ahorrar dinero a la hora de planificar y pronosticar sus cargas de trabajo. AWS Es importante reconocer que el descuento en el segundo escenario es menor en comparación con el primer escenario. En el segundo escenario, el precio de la licencia está incluido en el precio del servidor en la nube. AWS no ofrece descuentos en el precio de la licencia, pero siempre puede traer sus licencias (en situaciones específicas) y siempre AWS puede garantizar el mejor precio de cómputo/instancia.

Te recomendamos que hagas lo siguiente para controlar tus AWS gastos en recursos de procesamiento e instancias:

- Recomendaciones de acceso
- Personalice las recomendaciones según sus necesidades
- Revisa el compromiso por hora

Recomendaciones de acceso

Puedes usar la EC2 consola de Amazon para acceder a las recomendaciones de tu Savings Plan. Incluso puedes descargar tus recomendaciones para revisarlas más tarde en formato CSV. Para obtener más información, consulte Monitoring your Savings Plans en la documentación de Savings Plans.

Personalice las recomendaciones según sus necesidades

Abre la EC2 consola de Amazon, expande la sección Instances y, a continuación, selecciona Savings Plans. En esta página, se muestran los precios de las instancias y los cálculos antes y después de hacer una recomendación. También puedes ajustar los siguientes factores para tu recomendación:

- Plazo: por ejemplo, de 1 a 3 años
- Opción de pago: por ejemplo, pago por adelantado, pago parcial o sin pago inicial
- Historial: por ejemplo, los últimos 7, 30 o 60 días

Revisa el compromiso por hora

Con el mismo ejemplo, supongamos que tienes una instancia que funciona las 24 horas del día, los 7 días de la semana. La recomendación es utilizar un Savings Plan. Según el tamaño, tienes un precio

bajo demanda de 120 dólares/hora. Tienes la opción de comprometer 90$ por hora, pero esto puede variar en función de la región, el caso y la opción de compra. En este ejemplo, puedes ahorrar un 25 por ciento en comparación con el coste bajo demanda. También puede realizar un seguimiento de su utilización y cobertura, si están por debajo del umbral que ha definido, y configurar una alerta cuando se acabe el presupuesto.

Revise las recomendaciones

Le recomendamos que revise detenidamente las recomendaciones del Savings Plan. AWS no cambiará nada sin su permiso. Estas son solo recomendaciones y depende de usted aplicarlas o no.

Adquiera un plan

Abre la EC2 consola de Amazon, expande la sección Instances y, a continuación, selecciona Savings Plans. Luego, elija Purchase Savings Plans. Según sus requisitos, puede seleccionar las siguientes opciones: plazo, región, familia de instancias, compromiso por hora, opción de pago e incluso fecha de inicio. Puede elegir entre Compute Savings Plans, EC2 Instance Savings Plans y SageMaker AI Savings Plans. Para obtener más información, consulte Purchasing Savings Plans en la documentación de Savings Plans.

Obtenga un informe de utilización

Después de comprar un Savings Plan, puede obtener un informe de utilización. El informe le ayuda a comprobar su utilización, comprobar si el plan adquirido es suficiente para cubrir y maximizar el descuento, y cancelar o añadir nuevos descuentos. Este informe se puede exportar a otros formatos, como CSV. Para obtener más información, consulte Uso del informe de utilización en la documentación de Savings Plans.

Siga las prácticas recomendadas de compra

Le recomendamos que siga estas mejores prácticas antes de comprar Savings Plans:

- AWS Trusted AdvisorUtilícelo para eliminar EC2 los recursos inactivos.
- Realice la talla correcta antes de realizar las compras de Savings Plans.
- Establezca una tarifa por hora que mantenga de manera constante durante 30 a 60 días.
- Adquiera un compromiso para cubrir la mayor parte de la tarifa por hora constante que su organización pueda aceptar. Tenga en cuenta las fluctuaciones de la demanda o de la temporada.
- Elija un presupuesto de Savings Plans revisado trimestralmente para mantener una tasa constante (por ejemplo, un objetivo de cobertura del 70 por ciento para la cobertura de Savings Plans).

Si la tarifa cae por debajo de la cobertura deseada, compre un Savings Plan adicional como compensación para cumplir con su objetivo de cobertura.

Recursos adicionales

- Savings Plans for Amazon EC2 Reserved Instances (AWS documentos técnicos)
- Entender cómo se aplican los Savings Plans a su AWS consumo (documentación de Savings Plans)
- Anunciamos la facturación por segundo de las instancias de EC2 Windows Server y SQL Server (AWS documentación)
- AWS Serie sobre optimización de costes: vídeo sobre Savings Plans | Amazon Web Services () YouTube

Supervise los costes mediante herramientas AWS

Descripción general

La visibilidad de los costos es un factor clave para optimizar los costos en AWS. AWS tiene una serie de herramientas que puede utilizar para visualizar los costos y crear alertas en respuesta a esos costos. Estas incluyen herramientas que, por ejemplo AWS Budgets, te ayudan a llevar un registro de tus gastos y a reportarlos. En esta sección se describen formas específicas de supervisar los AWS gastos de Windows, de modo que pueda realizar un seguimiento de sus necesidades presupuestarias y reaccionar en consecuencia. Esto incluye añadir las etiquetas necesarias a EC2 los recursos de Windows. Estas etiquetas le permiten supervisar correctamente Windows EC2 y otros servicios de Microsoft mediante el uso de AWS Budgets.

Al monitorear los gastos y crear alertas con AWS herramientas, puede estar más informado sobre los gastos actuales, los gastos proyectados y las anomalías en los gastos. Si utiliza los Savings Plans para reducir el precio de las EC2 instancias por hora, le recomendamos que consulte la utilización y la cobertura generales del Savings Plan. Esto puede ayudarle a asegurarse de que está ahorrando continuamente. Puede usarlo AWS Cost Explorer para ver el inventario de Savings Plan y obtener recomendaciones de planes de ahorro adicionales en función del uso anterior. También puede realizar un seguimiento de gastos específicos mediante su uso AWS Budgetsy configuración AWS Cost Anomaly Detection.

Recomendaciones de optimización de costos

Le recomendamos que siga los siguientes pasos para optimizar sus costes mediante AWS Budgets el uso de Cost Explorer y la detección de anomalías:

- Etiquete los recursos de Windows EC2
- Configure alertas mediante AWS Budgets
- Habilite la detección de anomalías de costos
- Obtenga un análisis de gastos en tiempo real
- Vea los gastos de Windows con licencia incluida mediante Cost Explorer

Etiquete los recursos de Windows EC2

Para supervisar sus AWS gastos de forma eficaz, debe establecer una estrategia de etiquetado para las cargas de trabajo que desee supervisar. Esto es importante para poder agrupar los recursos de forma categórica y recibir notificaciones sobre gastos específicos, a diferencia de los gastos de uso general. Puede utilizar recursos de etiquetado que no solo ayudan a reducir los costes, sino que también se pueden utilizar para otros fines, como AWS Systems Manager la automatización. Además, le recomendamos que implemente alguna gestión para las etiquetas necesarias.

Para realizar un seguimiento de sus gastos en AWS Budgets Cost Explorer y Cost Anomaly Detection, debe asegurarse de que las etiquetas adecuadas estén colocadas. Puede usar las etiquetas para establecer un presupuesto específico para los artículos que coincidan con esas etiquetas, de modo que reciba alertas cuando aumenten los gastos.

Por ejemplo, puedes usar una etiqueta sencilla como Key=OS Value=Windows. De este modo, todas las instancias de Windows se agrupan en un grupo en el que puede realizar un seguimiento de los gastos. También puede usar etiquetas para otros elementos, como Systems Manager. Tras crear una etiqueta, debe activarla para el seguimiento de los costes. Considere la posibilidad de añadir una AWS Config regla que supervise las etiquetas adjuntas a determinados recursos. AWS Config puede avisarle si hay recursos en ejecución que no contengan las etiquetas adecuadas, lo que le proporcionará una representación precisa de sus EC2 gastos de Windows.

Una vez que haya colocado las etiquetas, podrá crear un presupuesto personalizado en AWS Billing. Esto proporciona visibilidad de sus EC2 gastos de Windows. Puede establecer un presupuesto diario o mensual.

Configure alertas mediante AWS Budgets

En este escenario de ejemplo, se crea un presupuesto diario para Windows EC2. Es un presupuesto recurrente que utiliza la opción de ajuste automático para realizar un seguimiento de sus gastos y ajustar el presupuesto en consecuencia. Si tienes un entorno estático, puedes usar un presupuesto fijo en su lugar. Asegúrese de elegir un intervalo de tiempo de referencia (por ejemplo, 30 días).

1. Inicie sesión en la AWS Cost Management consola AWS Management Console y ábrala.
2. En el panel de navegación, elija Budgets (Presupuestos).
3. En la parte superior de la página, elija Create budget.
4. En Configuración del presupuesto, seleccione Personalización (avanzada).
5. En Tipos de presupuesto, selecciona Presupuesto de costes. A continuación, elija Siguiente.
6. En Detalles, en Nombre del presupuesto, introduce el nombre del presupuesto. Por ejemplo, EC2gastos de Windows.
7. En Establecer importe presupuestario, en Período, selecciona Diariamente.
8. En el tipo de renovación del presupuesto, selecciona Presupuesto periódico para un presupuesto que se restablece después del período presupuestario.
9. En Fecha de inicio, selecciona la fecha o el período de inicio para empezar a hacer el seguimiento del importe presupuestado.
10. En Método de presupuestación, selecciona Ajuste automático (nuevo).
11. En Intervalo de tiempo previsto, selecciona Intervalo personalizado y, a continuación, introduce 30 días.
12. Elija Next (Siguiente).
13. En la sección Alcance del presupuesto, seleccione Filtrar dimensiones de AWS costo específicas. Aquí es donde se utilizan las etiquetas para crear las dimensiones adecuadas. AWS Budgets no admite el tipo de plataforma como opción en sus filtros. Por este motivo, debe aplicar las etiquetas del sistema operativo.
14. Elija Añadir filtro y, a continuación, seleccione la opción Etiqueta en Dimensiones.
15. Elija la etiqueta del sistema operativo y, a continuación, elija el valor de Windows para crear un presupuesto para la etiqueta.
16. Elija Next (Siguiente).
17. En la página Configurar alertas, seleccione Añadir un umbral de alerta. Aquí puede configurar dos alertas: una para un umbral del 50 por ciento y otra para un umbral del 100 por ciento. Si se supera el umbral del 50 por ciento antes de la mitad del mes, emitirá una advertencia. De esta

forma, podrás comprobar si tus gastos son superiores a los esperados y reaccionar antes de que acabe el mes.

18. En Umbral, introduzca 50 y seleccione el% del importe presupuestado.
19. En Trigger, elija Actual.
20. Para los destinatarios del correo electrónico, introduzca una dirección de correo electrónico. Agrega otra alerta para un umbral de 100.

> **Note**
>
> En este ejemplo, se utiliza una notificación por correo electrónico para la alerta, pero también se pueden utilizar otros enfoques, como Slack.

Habilite la detección de anomalías de costes

Puede usar sus etiquetas de costos para configurar alertas de gastos que sean una anomalía. Por ejemplo, puedes utilizarlas AWS Cost Anomaly Detectionpara crear monitores de tus gastos y recibir alertas cuando el sistema detecte gastos anormales en tu cuenta.

Para configurar un monitor y alertas para las etiquetas Key=OS y Value=Windows que creó anteriormente, haga lo siguiente:

1. Inicie sesión en la consola y ábrala AWS Management Console .AWS Cost Management
2. En el panel de navegación, elija Detección de anomalías en los costos.
3. Seleccione la pestaña Monitores de costes y, a continuación, seleccione Crear monitor.
4. En el paso 1, elija la etiqueta de asignación de costos como tipo de monitor.
5. En la clave de etiqueta de asignación de costos, elija Windows EC2 Spend.
6. Para el valor de la etiqueta de asignación de costos, elija Windows.
7. En Asigne un nombre a su monitor, escriba Windows EC2 spend.
8. Elija Next (Siguiente).
9. Para crear una suscripción para las alertas, selecciona Crear una nueva suscripción. Si tiene suscripciones existentes, seleccione Choose an existing subscription (Elegir una suscripción existente).
10. En el campo Nombre de la suscripción, introduce Windows EC2 Spend Anomaly.
11. En Frecuencia de alertas, selecciona Resúmenes diarios.

12. Para los destinatarios de las alertas, introduce tu dirección de correo electrónico.
13. Elija Agregar umbral. En Umbral, introduce 10 y, a continuación, selecciona el porcentaje por encima de la velocidad esperada.
14. Elija Crear monitor.

Obtenga una visión en tiempo real de los gastos

Una alerta es una herramienta útil para supervisar los EC2 gastos de Windows, pero debe utilizar Cost Explorer si quiere tener una visión en tiempo real de los gastos. Vea este vídeo para obtener información sobre cómo Cost Explorer le permite analizar y reducir sus EC2 costes. Para obtener más información, vea el vídeo AWS Supports You | Understanding and Reducing Your EC2 Costs publicado aquí YouTube.

Vea los gastos de Windows con licencia incluida

Puede ver los gastos de EC2 Windows en su cuenta mediante Cost Explorer. Para ver los gastos de Windows con licencia incluida, debe configurar los siguientes filtros correctos en Cost Explorer:

- En Plataforma, elija Windows (Amazon VPC). Para el funcionamiento de la API, elijaRunInstance: 0002. Este es el AWS Billing código para las instancias de Windows con licencia incluida. EC2
- Si quieres ver los gastos de tus instancias BYOL, cambia de: 0002 a: 0800. RunInstance RunInstance Este es el código de facturación BYOL de Windows. EC2

Con esta visibilidad en Cost Explorer, puede filtrar rápidamente sus costos para determinar exactamente lo que está gastando en Windows EC2. Si quiere profundizar aún más en sus AWS gastos, puede utilizar un filtro AWS Cost and Usage Report para reducir los gastos a nivel de instancia individual. También puedes generar informes que se pueden visualizar en Amazon QuickSight y crear paneles personalizados.

Para obtener más información, vea el vídeo Le AWS apoyamos: cómo visualizar sus informes de costos y uso en. YouTube

Recursos adicionales

- Cómo configurar las etiquetas necesarias con AWS Config (AWS Config documentación)
- AWS Budgets Tutorial: Configurar alertas para AWS Billing | Amazon Web Services (YouTube)
- AWS Cost and Usage Report Biblioteca de consultas (AWS Well-Architected Labs)

SQL Server

Los clientes han estado ejecutando cargas de trabajo de Microsoft AWS durante más de 15 años, más tiempo que cualquier otro proveedor de nube. Esto se debe en gran parte a que AWS tiene más experiencia con las aplicaciones de Microsoft en la nube y ofrece la mejor plataforma para Windows Server y Microsoft SQL Server en las siguientes áreas:

- Mayor rendimiento y fiabilidad
- Mayores servicios de seguridad e identidad
- Más apoyo a la migración
- Las capacidades más amplias y profundas
- Menor costo total de propiedad (TCO)
- Opciones de licencia flexibles

AWS admite todo lo necesario para crear y ejecutar aplicaciones de Windows que se basan en SQL Server, incluidas Active Directory, .NET, SQL Server, Windows Desktop como servicio y todas las versiones compatibles de Windows Server. Con una experiencia comprobada, AWS puede ayudarlo a levantar y cambiar, refactorizar o incluso modernizar fácilmente sus cargas de trabajo de Windows.

En esta sección de la guía se tratan los siguientes temas:

- Elija una solución de alta disponibilidad y recuperación ante desastres
- Comprenda las licencias de SQL Server
- Seleccione la EC2 instancia adecuada para las cargas de trabajo de SQL Server
- Consolide las instancias
- Compare las ediciones de SQL Server
- Evalúe la edición SQL Server Developer
- Evalúe SQL Server en Linux
- Optimice las estrategias de respaldo de SQL Server
- Modernice las bases de datos de SQL Server
- Optimice el almacenamiento para SQL Server
- Optimice las licencias de SQL Server mediante Compute Optimizer
- Optimice el tamaño de SQL Server mediante Compute Optimizer
- Revise Trusted Advisor las recomendaciones para las cargas de trabajo de SQL Server

Elija una solución de alta disponibilidad y recuperación ante desastres

Descripción general

Le recomendamos que diseñe una arquitectura para la implementación de SQL Server AWS que se adapte a las necesidades de su empresa y, al mismo tiempo, cumpla con sus objetivos de recuperación ante desastres (DR), incluidos su objetivo de tiempo de recuperación (RTO) y su objetivo de punto de recuperación (RPO). Las siguientes soluciones pueden ayudarlo a diseñar la arquitectura adecuada para SQL Server en Amazon Elastic Compute Cloud (Amazon EC2) y, al mismo tiempo, a optimizar los costos de sus cargas de trabajo de SQL Server.

- Grupos de disponibilidad de SQL Server Always On: los grupos de disponibilidad de SQL Server Always On proporcionan una alta disponibilidad y una solución de recuperación ante desastres (HA/DR) solutions for SQL Server databases. An availability group consists of a set of user databases that fail over together. Always On availability groups also provide redundancy at the database level, but don't require shared storage—each replica has its own local storage. You can deploy this feature as an HA/DRsolución). Para obtener más información, consulte ¿Qué es un grupo de disponibilidad de Always On? en la documentación de Microsoft.
- Instancias de clúster de conmutación por error (FCI) de SQL Server Always On: SQL Server Always On FCIs utiliza el clúster de conmutación por error de Windows Server (WSFC) para proporcionar alta disponibilidad a nivel de instancia de SQL Server. FCIs requieren almacenamiento compartido para alojar bases de datos. Puede utilizar el almacenamiento en bloques compartido o el almacenamiento de archivos compartido. Por ejemplo, puede utilizar Amazon FSx for Windows File Server o Amazon FSx for NetApp ONTAP como solución de almacenamiento compartido con varias zonas de disponibilidad. Para obtener más información, consulte Instancias de clúster de conmutación por error Always On (SQL Server) en la documentación de Microsoft.
- SIOS DataKeeper: SIOS DataKeeper puede ayudarlo a cumplir con los requisitos de alta disponibilidad y recuperación ante desastres al habilitar una FCI de SQL Server que abarque tanto las zonas de disponibilidad como las. Regiones de AWS SIOS DataKeeper crea una SAN virtual agrupada mediante volúmenes locales de Amazon Elastic Block Store (Amazon EBS) y utiliza la replicación sincrónica entre las zonas de disponibilidad para alta disponibilidad, a la vez que utiliza la replicación asíncrona entre regiones y para la recuperación ante desastres. Para obtener más información, consulte la sección Protección de alta disponibilidad para aplicaciones de Windows en la documentación de SIOS.

- Grupos de disponibilidad distribuida: los grupos de disponibilidad distribuidos son un tipo especial de grupo de disponibilidad que abarca dos grupos de disponibilidad Always On independientes. Un grupo de disponibilidad puede residir en dos regiones distintas (por ejemplo, `us-east-1` y `us-west-1`). Puede pensar en un grupo de disponibilidad distribuido como un grupo de disponibilidad de grupos de disponibilidad porque los grupos de disponibilidad Always On subyacentes están configurados en dos clústeres de WSFC diferentes. Se requiere la edición SQL Server Enterprise para implementar grupos de disponibilidad distribuidos. Para obtener más información, consulte Grupos de disponibilidad distribuida en la documentación de Microsoft.
- Transporte de registros: puede implementar el envío de registros para proteger sus bases de datos en varias regiones, en el raro caso de que una región se vea afectada y deje de estar disponible. En función de la frecuencia de envío de las transacciones y los registros, puede lograr el RPO y el RTO en cuestión de minutos. Para obtener más información, consulte Acerca del envío de registros (SQL Server) en la documentación de Microsoft.
- AWS Elastic Disaster Recovery— Elastic Disaster Recovery es una aplicación de software como servicio (SaaS) que administra la replicación de servidores desde cualquier infraestructura con AWS fines de recuperación ante desastres. También puede usar Elastic Disaster Recovery para replicar SQL Server en todas las regiones. Elastic Disaster Recovery es una solución basada en agentes que replica máquinas virtuales completas, incluido el sistema operativo, todas las aplicaciones instaladas y todas las bases de datos en un área de almacenamiento provisional. Para obtener más información, consulte ¿Qué es Elastic Disaster Recovery? en la documentación de Elastic Disaster Recovery.
- AWS Database Migration Service (AWS DMS): AWS DMS admite la migración de datos en tiempo real hacia y desde una región diferente AWS, incluida una región diferente. Puede usar esta función para configurar una instancia de SQL Server independiente en una región diferente para que sirva como base de datos de recuperación ante desastres. Para obtener más información, consulte ¿Qué es AWS Database Migration Service? en la AWS DMS documentación.

Grupos de disponibilidad de SQL Server Always On

Si utiliza la edición Enterprise de SQL Server solo para un grupo de disponibilidad Always On de alta disponibilidad, puede pasarse a la edición SQL Server Standard aprovechando los grupos de disponibilidad básicos. Puede reducir los costos entre un 65 y un 75 por ciento si utiliza grupos de disponibilidad básicos en lugar de grupos de disponibilidad de Always On.

> **Note**
>
> Para obtener información adicional sobre las diferencias de costo entre las distintas ediciones de SQL Server, consulte la sección Comparación de las ediciones de SQL Server de esta guía.

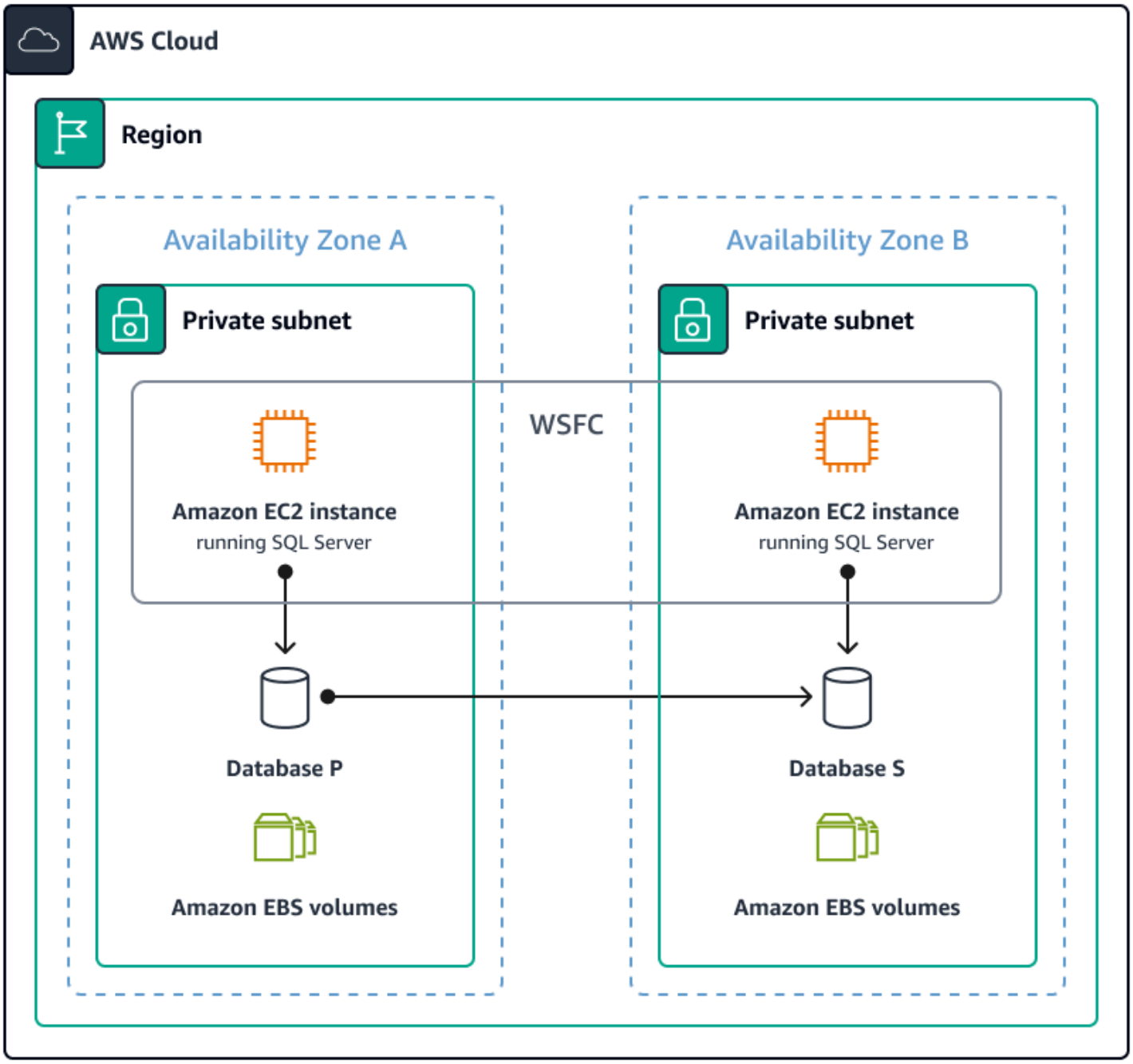
Características

- Disponible en la edición estándar de SQL Server
- Límite de dos réplicas (principal y secundaria)
- No hay acceso de lectura en la réplica secundaria
- No hay controles de integridad en las réplicas secundarias

Limitaciones

- Support solo para una base de datos de disponibilidad por grupo de disponibilidad
- Los grupos de disponibilidad básicos no pueden formar parte de un grupo de disponibilidad distribuido

El siguiente diagrama muestra un ejemplo de arquitectura para una solución de clúster de conmutación por error de Windows Server.



Instancias de clúster de conmutación por error de SQL Server Always On

Puede utilizar instancias de clúster de conmutación por error (FCIs) para garantizar la continuidad de las operaciones de la base de datos y, al mismo tiempo, minimizar el tiempo de inactividad y reducir el riesgo de pérdida de datos. FCIs ofrezca una solución fiable si busca una alta disponibilidad para su base de datos de SQL Server sin una configuración de réplica de lectura.

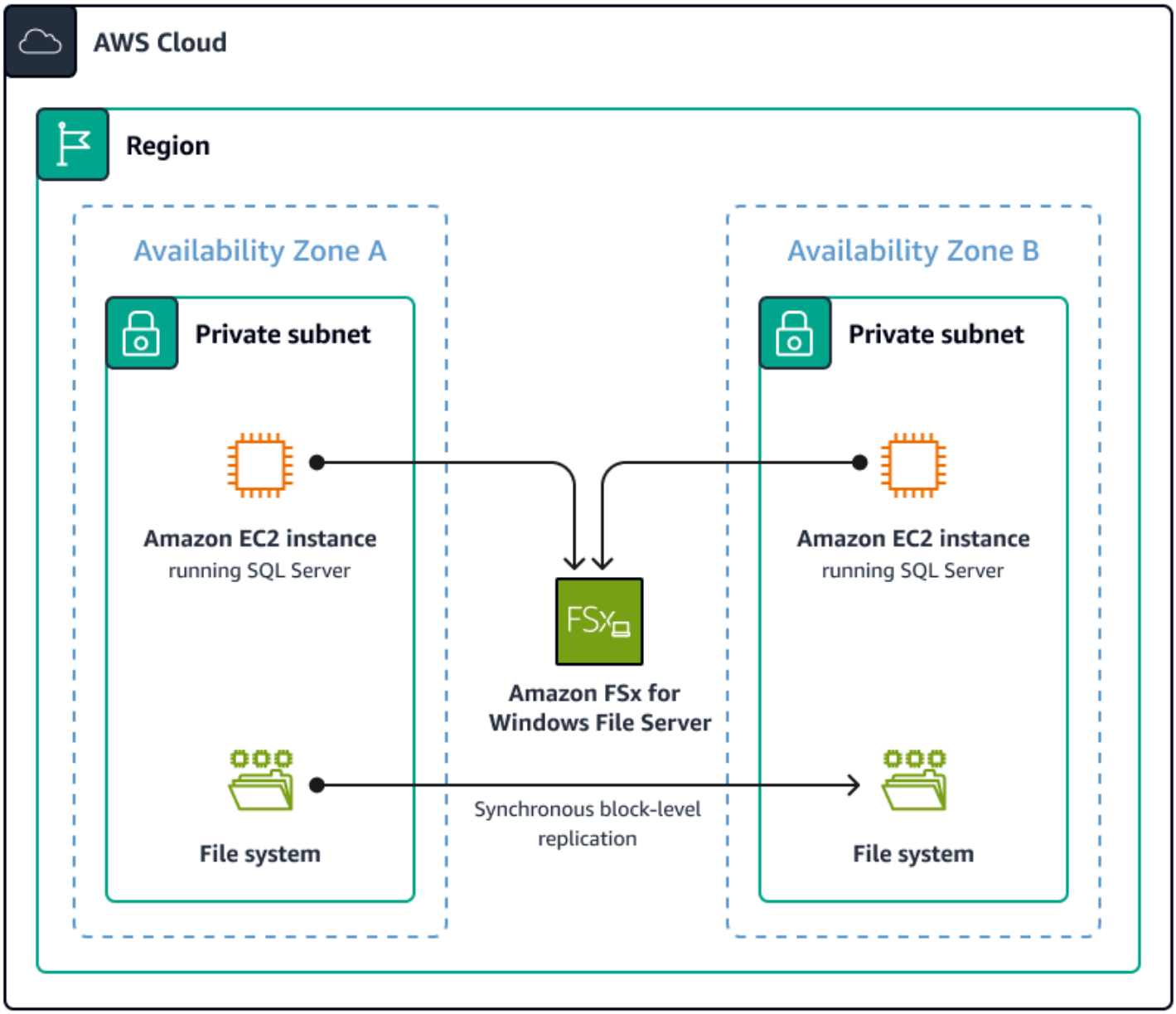
A diferencia de los grupos de disponibilidad, FCIs puede proporcionar una solución de conmutación por error fiable sin necesidad de la edición SQL Server Enterprise. En su lugar, solo FCIs requieren licencias de la edición estándar de SQL Server. Se puede utilizar FCIs para reducir los costes de licencias de SQL Server entre un 65 y un 75 por ciento.

> **Note**
>
> Para obtener información adicional sobre las diferencias de costo entre las ediciones de SQL Server, consulte la sección Comparación de las ediciones de SQL Server de esta guía.

Considere lo siguiente:

- Amazon FSx for Windows File Server ofrece una potente solución para cumplir con los requisitos de almacenamiento compartido FCI de SQL Server. Puede utilizar Windows File Server FSx para evitar la necesidad de adquirir una licencia para una solución de replicación del almacenamiento y gestionar el almacenamiento compartido por su cuenta. Esto puede resultar en un ahorro de costos significativo del 30 al 40 por ciento. Para obtener más información, consulte la publicación Simplifique las implementaciones de alta disponibilidad de Microsoft SQL Server con Amazon FSx for Windows File Server en el blog sobre AWS almacenamiento.
- Con el resumen de las ventajas de Software Assurance (PDF descargable) y el modelo Bring Your Own License (BYOL), puede aprovechar las ventajas de la conmutación por error pasiva, siempre y cuando el servidor secundario sea pasivo. Esto se traduce en un ahorro de costes en las licencias de SQL, ya que no es necesario proporcionar licencias al nodo pasivo del clúster.

El siguiente diagrama muestra un ejemplo de arquitectura para una FCI de SQL Server utilizada FSx para Windows File Server.



SIOS DataKeeper

Le recomendamos que tenga en cuenta los requisitos de almacenamiento compartido si planea implementar SQL Server FCIs en AWSél. Las instalaciones locales tradicionales suelen utilizar una red de área de almacenamiento (SAN) para cumplir con los requisitos de almacenamiento compartido, pero esta no es una opción viable. AWS Amazon FSx for Windows File Server es la solución de almacenamiento recomendada para SQL Server FCI on AWS, pero tiene limitaciones que impiden añadir servidores de clústeres en diferentes Regiones de AWS ubicaciones.

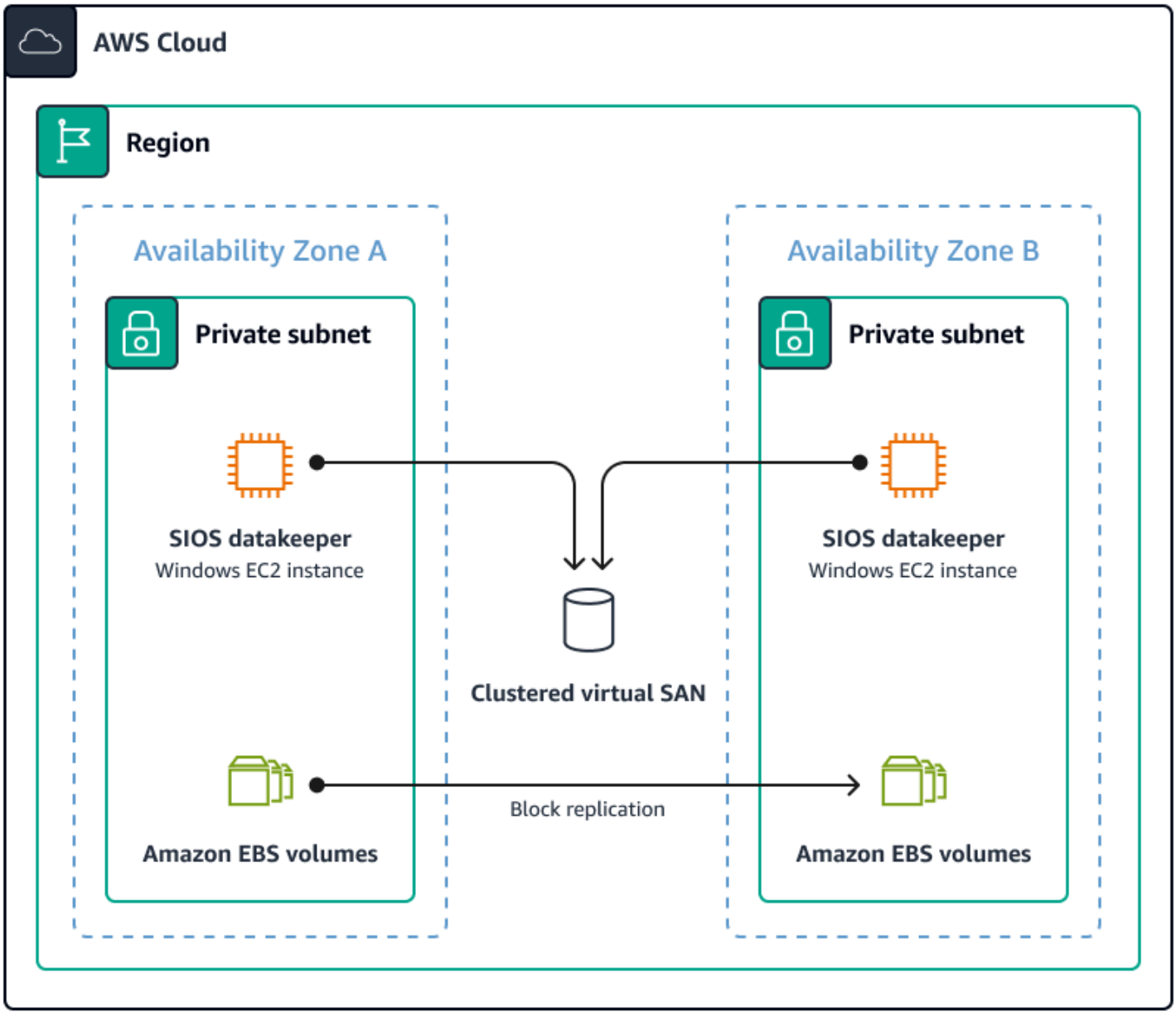
Puede utilizar SIOS DataKeeper para crear una FCI de SQL Server que cubra tanto las zonas de disponibilidad como las regiones y, al mismo tiempo, reduzca los costes entre un 58 y un 71 por ciento. SIOS DataKeeper puede ayudarlo a obtener los beneficios de alta disponibilidad de la FCI. Esto convierte a SIOS en DataKeeper una solución rentable y fiable para las organizaciones.

Tenga en cuenta las siguientes ventajas adicionales del uso de SIOS: DataKeeper

- SIOS DataKeeper crea una SAN virtual agrupada mediante volúmenes de EBS locales y utiliza la replicación sincrónica entre las zonas de disponibilidad para lograr una alta disponibilidad. Para la recuperación ante desastres, SIOS DataKeeper utiliza la replicación asíncrona entre regiones.
- SIOS DataKeeper proporciona funciones de agrupamiento en clústeres de clase empresarial mediante el uso de la edición SQL Server Standard. Esto reduce los costes de licencia de SQL Server entre un 65 y un 75 por ciento en comparación con la implementación de la alta disponibilidad con los grupos de disponibilidad de SQL Server Always On que utilizan la edición SQL Server Enterprise. Con SIOS DataKeeper, puede crear un entorno de SQL Server rentable, flexible y de alta disponibilidad que satisfaga las necesidades de su organización.

> **Note**
>
> Para obtener información adicional sobre las diferencias de costo entre las ediciones de SQL Server, consulte la sección Comparación de las ediciones de SQL Server de esta guía.

El siguiente diagrama muestra un ejemplo de arquitectura para una FCI de SQL Server que utiliza una solución de SAN virtual agrupada en clústeres.

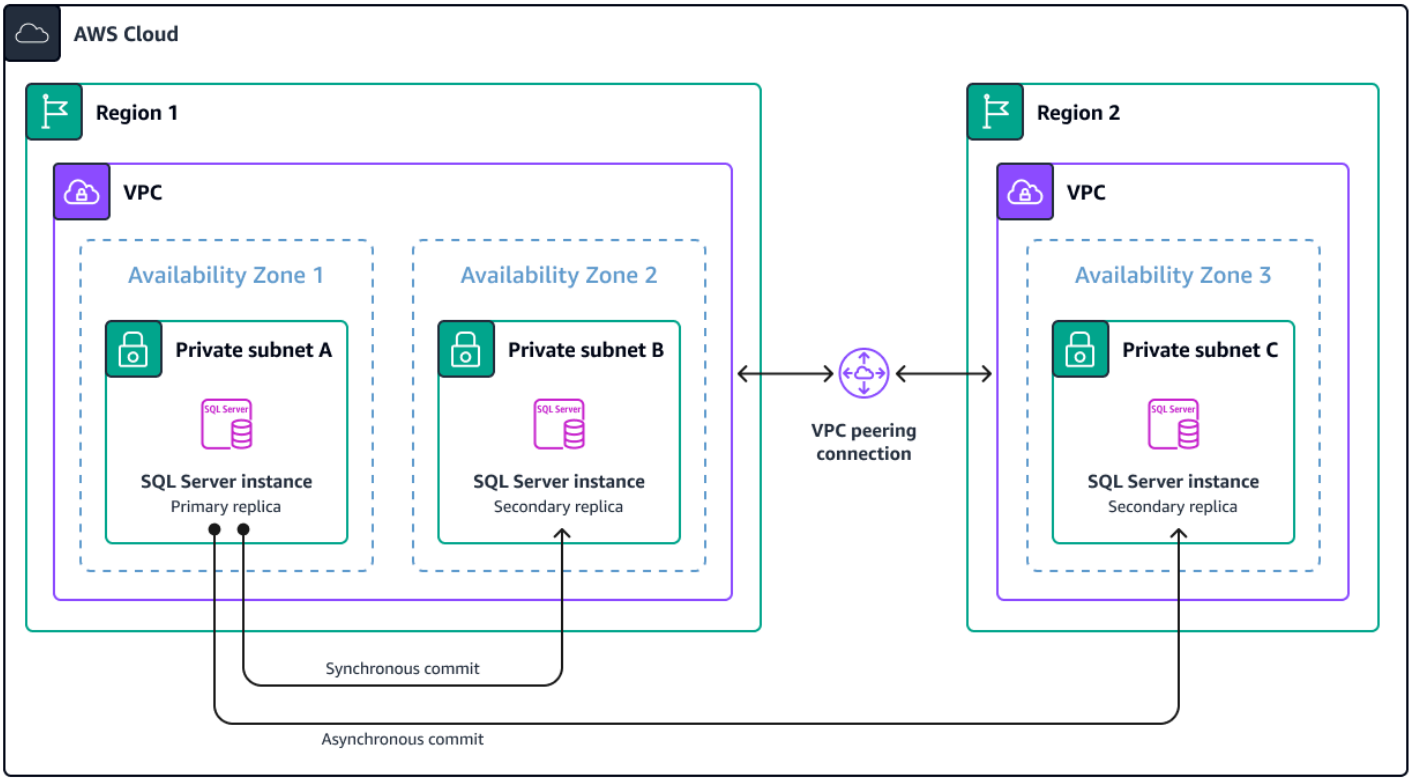


Grupos de disponibilidad AlwaysOn

Puede usar los grupos de disponibilidad Always On tanto para fines de alta disponibilidad como de recuperación ante desastres. Puede lograr una alta disponibilidad mediante la implementación de SQL Server en dos zonas de disponibilidad de una región. Puede lograr la recuperación ante desastres ampliando los grupos de disponibilidad entre las regiones.

El siguiente diagrama muestra un ejemplo de arquitectura para una solución basada en los grupos de disponibilidad de Always On. Las réplicas de la región 1 del diagrama utilizan una confirmación sincrónica, que proporciona una conmutación por error automática del grupo de disponibilidad. La

réplica de la región 2 utiliza una confirmación asíncrona, que requerirá una conmutación por error manual del grupo de disponibilidad.



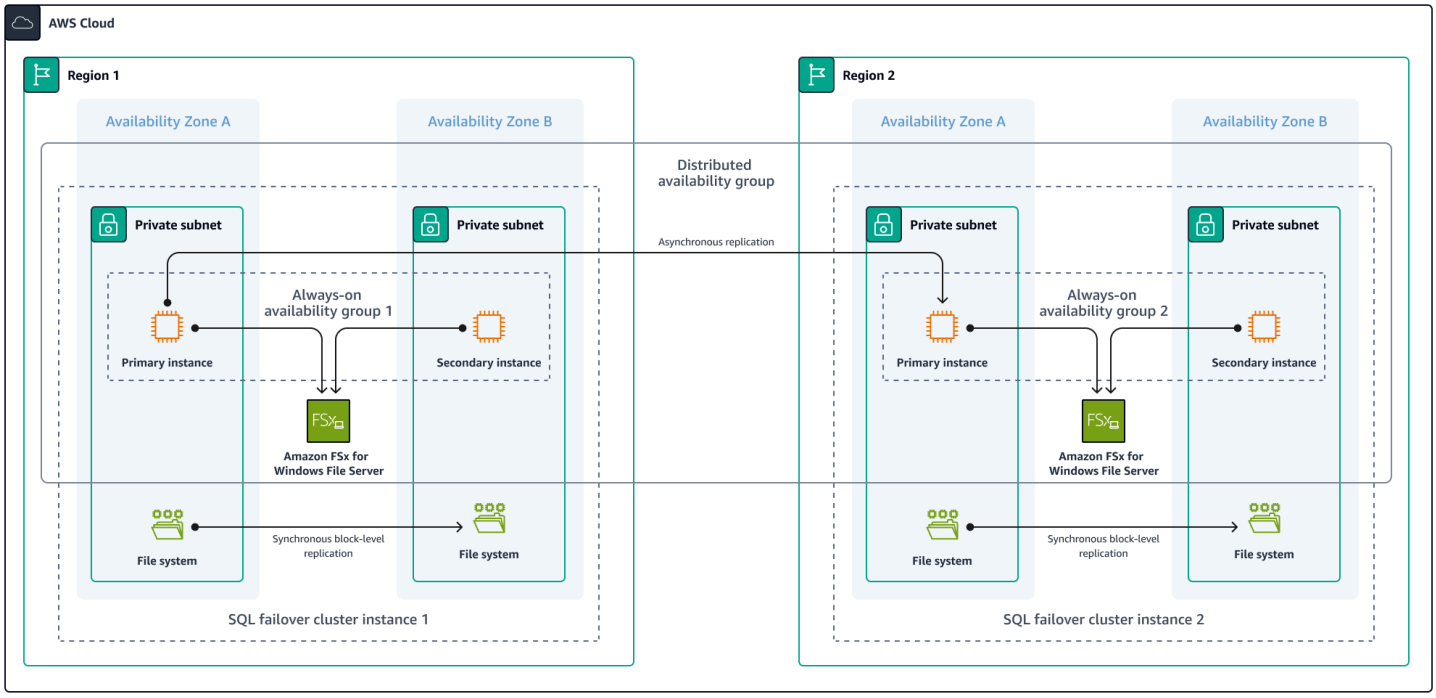
Grupos de disponibilidad distribuida

Para las implementaciones de SQL Server de misión crítica en las que no se puede comprometer la confiabilidad o la recuperación ante desastres, recomendamos un enfoque multirregional. La distribución de los grupos de disponibilidad en varias regiones es la solución más sólida para mantener la continuidad empresarial y minimizar el tiempo de inactividad.

Esta arquitectura aprovecha al máximo las capacidades de Amazon FSx for Windows File Server, incluido el almacenamiento compartido, la replicación sincrónica a nivel de bloque y SQL Server. FCIs Estas capacidades le permiten crear un entorno de SQL Server de alta disponibilidad que abarque varias zonas de disponibilidad. Al replicar esta configuración en otra región, obtiene un sistema totalmente redundante que puede gestionar incluso las interrupciones más graves. Lo que distingue a esta solución es el nivel de flexibilidad y seguridad que ofrece. La arquitectura independiente del dominio de los grupos de disponibilidad distribuida permite que los servidores de clúster de Windows subyacentes se unan a diferentes dominios de Active Directory, mientras que la autenticación basada en certificados garantiza la máxima protección para los entornos de SQL Server y proporciona altos requisitos de RTO y RPO para una estrategia de DR multirregional.

Para obtener información sobre cómo crear una arquitectura multirregional, consulte Notas de campo: Creación de una arquitectura multirregional para SQL Server mediante FCI y grupos de disponibilidad distribuida en el blog de arquitectura. AWS

El siguiente diagrama muestra un ejemplo de arquitectura para una solución multirregional que utiliza grupos de disponibilidad distribuidos.



Envío de registros

El envío de registros es un método comprobado, confiable y rentable para proteger sus bases de datos en todas las regiones en caso de una interrupción inesperada. Las organizaciones llevan décadas utilizando el transporte de registros para proteger sus datos.

Si implementa el envío de registros AWS, puede lograr el RPO y el RTO en cuestión de minutos, en función de la frecuencia de las transacciones y de las tareas de envío de registros. En el improbable caso de que una región quede inaccesible, el envío de registros mantiene sus datos seguros y recuperables.

Tenga en cuenta las siguientes ventajas adicionales del uso del transporte de troncos:

- Reduzca los costos y cumpla con los requisitos de su empresa mediante el uso del transporte de troncos para mejorar la resiliencia de la recuperación ante desastres en todas las regiones. El

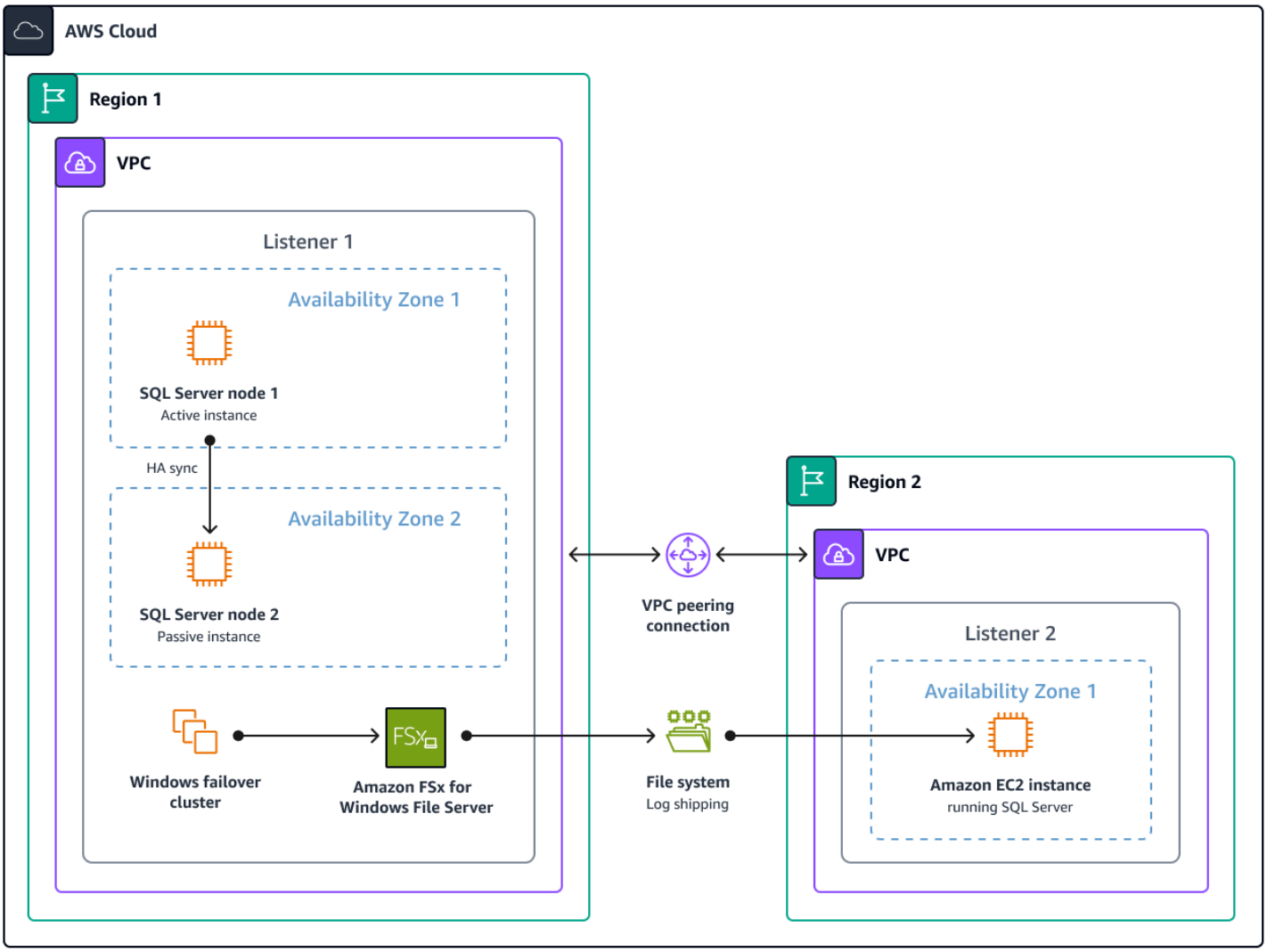
transporte de registros reduce el costo total de propiedad, ya que solo necesita licencias de SQL Server Standard o SQL Server Web Edition.

- Elimine los costos de licencia de un servidor pasivo o de recuperación ante desastres mediante el envío de registros con Software Assurance activo. Solo se necesita una licencia para el SQL Server principal o activo cuando se utiliza el envío de registros con Software Assurance.
- Reduzca los costos de licencias de SQL Server entre un 65 y un 75 por ciento al eliminar la necesidad de utilizar la edición SQL Server Enterprise para configurar grupos de disponibilidad distribuidos entre las regiones. Para ello, puede utilizar la edición Standard de SQL Server y SQL Server en FCIs combinación con el envío de registros para cumplir con sus requisitos de recuperación ante desastres.

> **Note**
>
> Para obtener información adicional sobre las diferencias de costo entre las ediciones de SQL Server, consulte la sección Comparación de las ediciones de SQL Server de esta guía.

Para obtener más información, consulte Ampliar la recuperación ante desastres de SQL Server mediante el envío de registros para la configuración FCI de SQL Server con Amazon FSx para Windows en el blog de AWS arquitectura.

El siguiente diagrama muestra un ejemplo de arquitectura para una solución de transporte de registros.

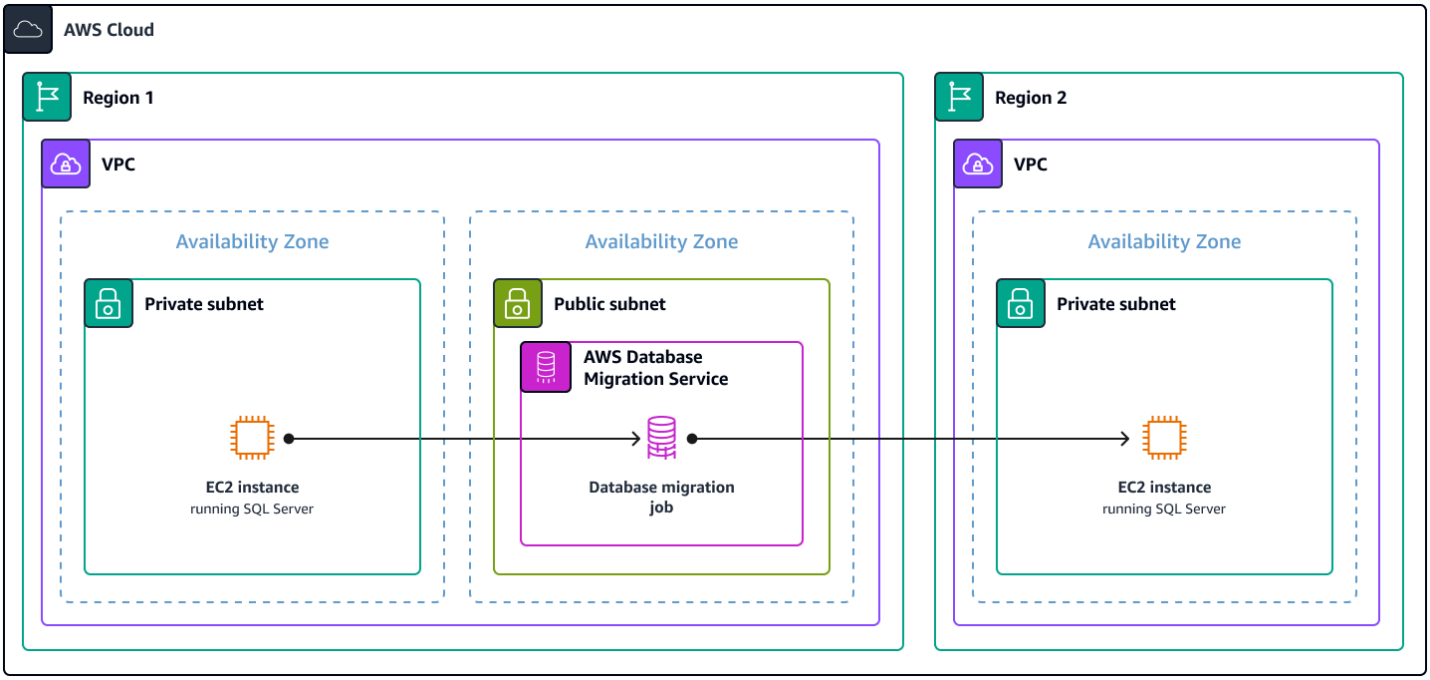


AWS Database Migration Service

Puede usar AWS Database Migration Service (AWS DMS) para diseñar una solución HA/DR en función de las necesidades de su aplicación. AWS DMS le permite copiar fácilmente los datos a una base de datos secundaria de SQL Server en la misma región (HA) o entre regiones (DR). Este enfoque es técnicamente sólido y le permite maximizar su inversión en AWS infraestructura y, al mismo tiempo, optimizar el uso de los recursos.

AWS DMS es un servicio rentable. Solo se le cobrarán los recursos de CPU utilizados durante el proceso de transferencia y cualquier almacenamiento de registros adicional. Esto significa que puede beneficiarse de esta solución sin incurrir en costes adicionales significativos. Puede utilizarla AWS DMS para garantizar que sus datos estén disponibles y accesibles y, al mismo tiempo, minimizar los costos asociados con las licencias y el uso de los recursos.

El siguiente diagrama muestra un ejemplo de arquitectura para una solución basada en AWS DMS.



AWS Elastic Disaster Recovery

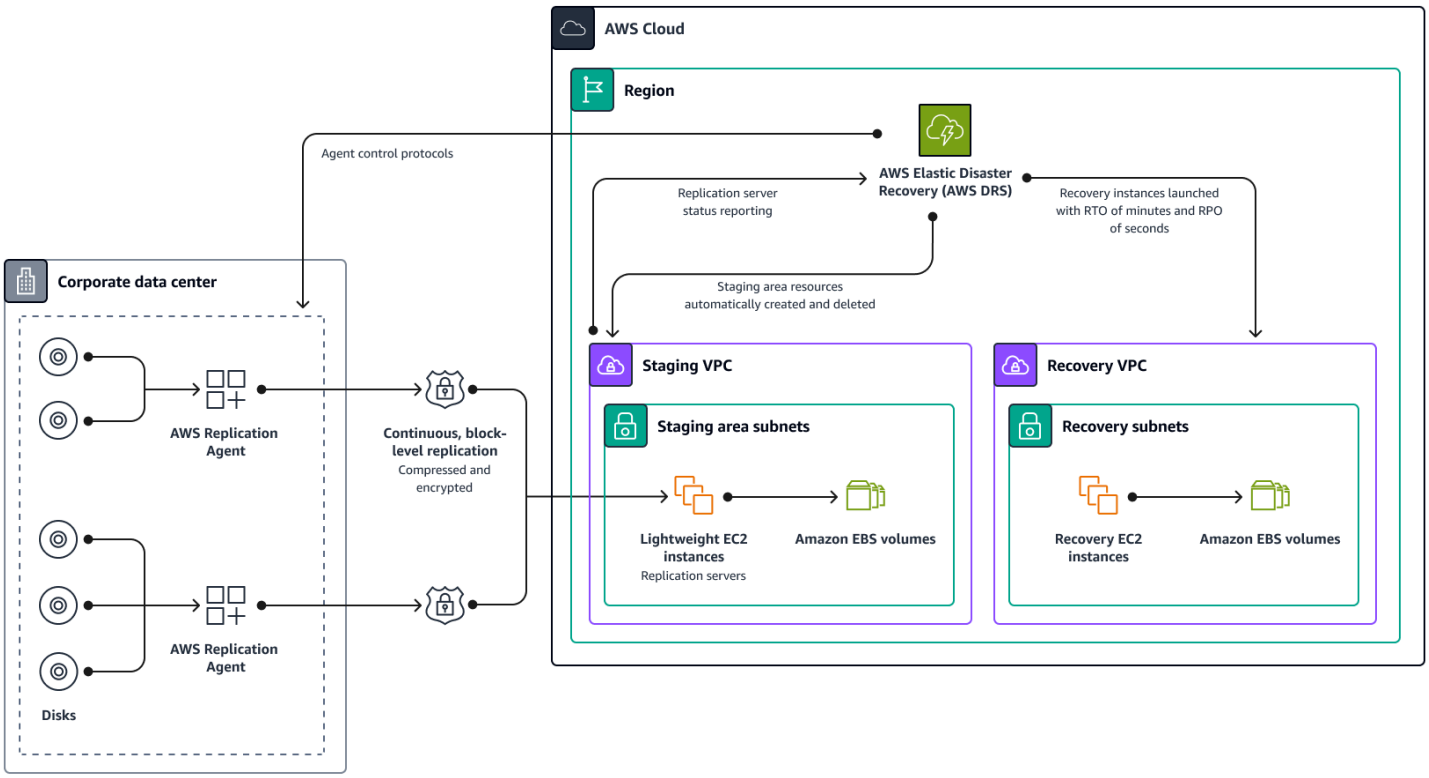
Algunas organizaciones deben asegurarse de que todas las aplicaciones empresariales críticas cuenten con un plan de recuperación ante desastres. En el pasado, muchas de estas organizaciones hacían grandes inversiones en las soluciones tradicionales de recuperación ante desastres, que requerían la creación previa y el mantenimiento de toda una infraestructura duplicada. Este enfoque es costoso, requiere mucho tiempo y es difícil de escalar.

Ahora puede usarlo AWS Elastic Disaster Recovery para eliminar la necesidad de crear previamente una infraestructura de recuperación ante desastres. Las máquinas de recuperación ante desastres no se inician en Elastic Disaster Recovery hasta que son necesarias, por lo que solo pagas por lo que usas cuando lo necesitas. Esto significa que puede reducir considerablemente los costos de licencias de software y de cómputo de alto rendimiento.

Además, el área de almacenamiento de la solución de recuperación ante desastres contiene volúmenes de bajo coste de Amazon Elastic Block Store (Amazon EBS). Los volúmenes de EBS reducen aún más el costo de aprovisionar recursos duplicados. Esto le permite reducir los costos generales de recuperación ante desastres y, al mismo tiempo, mantener una solución de recuperación ante desastres sólida y confiable que cumpla con los requisitos de su empresa. Puedes usar Elastic Disaster Recovery para centrarte en tus actividades comerciales principales y, al AWS mismo tiempo, cuidar la infraestructura subyacente de tu solución de recuperación ante desastres.

En el caso de SQL Server, puedes usar Elastic Disaster Recovery como una opción rentable de recuperación ante desastres. La licencia para el nodo pasivo en una arquitectura de SQL Server altamente disponible y tolerante a errores está cubierta si utiliza Software Assurance activo. Sin embargo, sigue pagando los costos de procesamiento para que el servidor pasivo esté en línea. Con Elastic Disaster Recovery, el servidor principal puede replicarse en el entorno de recuperación ante desastres sin necesidad de mantener activo Software Assurance y sin tener que pagar los costos de cómputo de la recuperación ante desastres. Esta combinación de ahorros puede reducir los costos de recuperación ante desastres de SQL Server en un 50 por ciento o más.

El siguiente diagrama muestra un ejemplo de arquitectura para una solución basada en Elastic Disaster Recovery.



Para obtener más información, consulte Cómo configurar la alta disponibilidad de SQL Server en un sitio de recuperación ante desastres que se restauró mediante AWS Elastic Disaster Recovery el AWS blog Microsoft Workloads on.

Comparación de costos

En la siguiente tabla se comparan los costes de las soluciones HA/DR incluidas en esta sección. A los efectos de esta comparación, se hacen las siguientes suposiciones:

- Tipo de instancia: r5d.xlarge
- Tipo de licencia: licencia incluida para Windows y SQL Server
- Región: `us-east-1`

Solución	Alta disponibilidad	Recuperación ante desastres	Enterprise Edition	Standard Edition	Costo
Envío de registros	No	Sí	Sí	Sí	Edición SQL Server Enterprise: 32.674,8$ (2 nodos) Edición SQL Server Standard: 14.804,4$ (2 nodos)
Grupos de disponibilidad AlwaysOn	Sí	Sí	Sí	Sí, pero grupos de disponibilidad básicos (2 nodos)	Edición SQL Server Enterprise: 32.674,8$ (2 nodos) Edición SQL Server Standard: 14.804,4$ (2 nodos)
Siempre encendido FCIs	Sí	No	Sí	Sí (2 nodos)	Edición estándar de SQL Server: 14.804,4$

Solución	Alta disponibilidad	Recuperación ante desastres	Enterprise Edition	Standard Edition	Costo
Grupos de disponibilidad distribuida	Sí	Sí	Sí	No	Edición SQL Server Enterprise: 65.349,6 dólares (4 nodos)

Solución	Alta disponibilidad	Recuperación ante desastres	Enterprise Edition	Standard Edition	Costo
Recuperación de desastres elástica	No	Sí	Sí	Sí	Aproximadamente 107,48$ al mes para la replicación de 1 instancia y 1 TB de almacenamiento Nota: Elastic Disaster Recovery se factura por hora y por servidor de replicación. El costo es el mismo, independientemente de la cantidad de discos, el tamaño del almacenamiento, la cantidad de operaciones de perforación o recuperación iniciadas o la región en la

Solución	Alta disponibilidad	Recuperación ante desastres	Enterprise Edition	Standard Edition	Costo
					que se vaya a replicar.
Administrador de datos de SIOS	Sí	Sí	Sí	Sí	Grupos de disponibilidad Always On con Software Assurance (2 nodos, 24 núcleos): 213.480$ Clúster de SQL Server de 2 nodos que se ejecuta en la edición SQL Server Standard con SIOS DataKeeper y Software Assurance: 61.530$ (2 nodos)

Solución	Alta disponibilidad	Recuperación ante desastres	Enterprise Edition	Standard Edition	Costo
AWS DMS	No	Sí	Sí	Sí	745,38$ al mes por una instancia r5.xlarge y 1 TB de almacenamiento

Recomendaciones de optimización de costos

Le recomendamos que siga los siguientes pasos para elegir una solución de alta disponibilidad y recuperación ante desastres que cumpla con los requisitos de su organización:

- Consulte la sección Seleccione la EC2 instancia correcta para las cargas de trabajo de SQL Server de esta guía.
- Determine los requisitos de IOPS y rendimiento de sus cargas de trabajo ejecutando contadores de rendimiento durante los picos de carga de trabajo:
  - IOPS = disco reads/sec + disk writes/sec
  - Rendimiento = lectura de disco bytes/sec + disk write bytes/sec
- Utilice los siguientes tipos de volúmenes de almacenamiento para obtener un mejor rendimiento y ahorrar costes:
  - NVMe almacenamiento de instancias `tempdb` y extensión del grupo de búferes
  - volúmenes io2 para archivos de bases de datos
- Úselo AWS Trusted Advisorpara obtener recomendaciones sobre la optimización de costos de SQL Server en Amazon EC2. No necesita instalar un agente Trusted Advisor para realizar las comprobaciones de optimización de SQL Server. Trusted Advisor inspecciona las configuraciones de instancias incluidas en la licencia de Amazon EC2 SQL Server, como virtual CPUs (vCPUs), versión y edición. A continuación, Trusted Advisor hace recomendaciones basadas en las mejores prácticas.
- Úselo AWS Compute Optimizer para las recomendaciones de tamaño correcto de las EC2 instancias de Amazon y Amazon EBS.

- Úselo Calculadora de precios de AWSpara diseñar su estrategia de alta disponibilidad y recuperación ante desastres para la estimación de costos.
- Para determinar si es posible pasar de la edición SQL Server Enterprise a la edición SQL Server Standard, utilice la vista de administración dinámica sys dm_db_persisted_sku_features para identificar las funciones específicas de la edición que están activas en la base de datos actual.

> **Note**
>
> Side-by-side Las migraciones son necesarias para realizar cambios en la edición de SQL EC2 Server cuando se utilizan instancias con licencia incluida.

- Realice simulacros de recuperación ante desastres semestrales o anuales para diseñar mejor un diseño que pueda recuperar la base de datos con un RTO y un RPO definidos. Esto también puede ayudarle a identificar cualquier punto débil de la arquitectura.

Recursos adicionales

- Simplifique las implementaciones de alta disponibilidad de Microsoft SQL Server con Amazon FSx for Windows File Server (blog sobre AWS almacenamiento)
- Notas de campo: Creación de una arquitectura multirregional para SQL Server mediante FCI y grupos de disponibilidad distribuida (AWS blog de arquitectura)
- Diseñe una recuperación ante desastres para SQL Server en AWS: Parte 1 (blog sobreAWS bases de datos)
- Alta disponibilidad de Microsoft SQL con Amazon FSx para Windows (YouTube)
- Cómo maximizar el rendimiento de Microsoft SQL Server con Amazon EBS (blog sobre AWS almacenamiento)
- Comparación de sus patrones de almacenamiento local con los servicios de AWS almacenamiento (AWS blog sobre almacenamiento)
- Planea sustituir el NAS de un centro de datos por Amazon FSx File Gateway (blog sobre AWS almacenamiento)
- Optimización de los costes de las implementaciones de SQL Server de alta disponibilidad en AWS(blog sobreAWS almacenamiento)
- Cómo configurar la recuperación ante desastres para los grupos de disponibilidad Always On de SQL Server mediante AWS Elastic Disaster Recovery (Microsoft Workloads activado AWS)

- Cómo configurar la alta disponibilidad de SQL Server en un sitio de recuperación ante desastres que se restauró con AWS Elastic Disaster Recovery (Microsoft Workloads activado AWS)

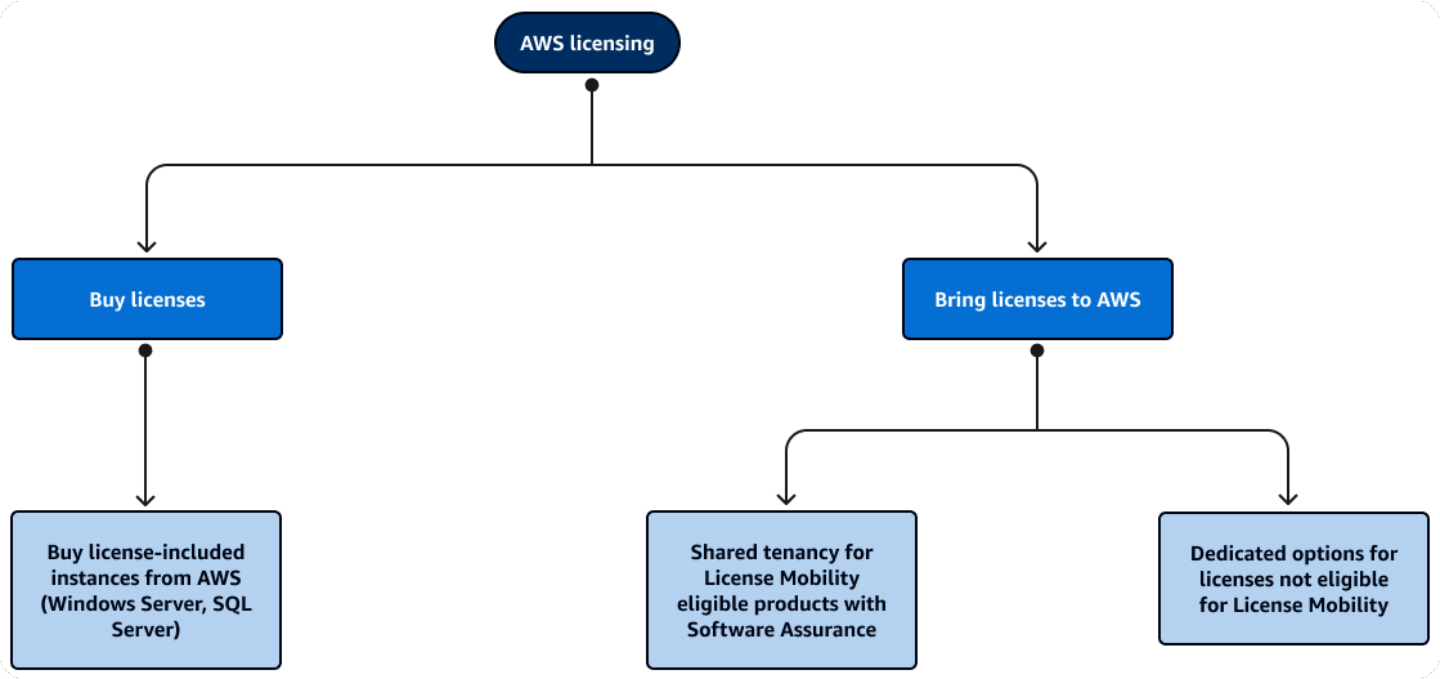
Comprenda las licencias de SQL Server

Descripción general

A medida que más y más empresas trasladan sus cargas de trabajo a la nube, la optimización de los costes en las plataformas en la nube se ha convertido en una prioridad fundamental. La concesión de licencias es uno de los costes más importantes asociados a la ejecución de las cargas de trabajo de Microsoft. AWS En esta sección se explica cómo optimizar los costes AWS mediante la optimización de las licencias de Microsoft para SQL Server.

AWS opciones de licenciamiento

AWS ofrece una gama de opciones flexibles de optimización de costes para la concesión de licencias. Estas opciones de licencia están diseñadas para ayudarlo a reducir los costos, mantener el cumplimiento y satisfacer las necesidades de su empresa.



AWS clasifica las licencias en tres tipos principales:

1. Licencia incluida: esta opción de licencia le permite comprar y utilizar licencias a pedido, pagando solo por lo que utilice. La opción con licencia incluida es ideal para situaciones en las que necesita flexibilidad en el uso de las licencias y desea evitar costes iniciales. Puede elegir entre una gama de productos de Windows Server, SQL Server y otros productos de Microsoft.
2. Productos con licencia propia (BYOL) y movilidad de licencias: esta opción de licencia está diseñada para situaciones en las que ya dispone de licencias y desea utilizarlas en la nube. AWS permite a los clientes llevar sus propias licencias a la nube a través del programa License Mobility de Microsoft. Puede utilizar productos con movilidad de licencias, como SQL Server with Software Assurance (SA), en régimen de arrendamiento compartido o dedicado para reducir los costes de las AWS instancias.
3. Productos BYOL sin movilidad de licencias: para los productos de Microsoft que no tienen movilidad de licencias, como Windows Server, AWS ofrece opciones específicas para usar estos productos en la nube. Además, los hosts dedicados ofrecen la oportunidad de conceder licencias a nivel básico físico. Esto puede ahorrarle un 50 por ciento o más en las licencias necesarias para ejecutar sus cargas de trabajo. Los hosts dedicados son una excelente opción para cargas de trabajo estables y predecibles que se ejecutan la mayor parte del tiempo.

El impacto en los costos de traer licencias

La incorporación de licencias puede tener un impacto significativo en el costo de ejecutar las cargas de trabajo de Microsoft. AWS Si traes tus propias licencias, no estás obligado a pagar costes de licencia adicionales por las instancias que se ejecutan en la nube. Esto puede suponer un importante ahorro de costes.

La siguiente comparación muestra el costo mensual bajo demanda que implica ejecutar una sola instancia de c5.xlarge las 24 horas del día, los 7 días de la semana:

- Edición Windows Server + SQL Server Enterprise: 1353$ al mes (licencia incluida)
- Edición estándar de Windows Server + SQL Server: 609$ al mes (licencia incluida)
- Solo para Windows Server: 259$ al mes (licencia incluida)
- Solo informática (Linux): 127$ al mes

En última instancia, traer sus propias licencias puede tener un impacto significativo en el costo de ejecutar las cargas de trabajo de AWS Microsoft. Si usa sus licencias actuales, puede reducir los costos de las licencias y ahorrar dinero en su AWS factura total.

Optimización de licencias

Una evaluación de AWS optimización y licencias (AWS OLA) puede ayudarlo a optimizar sus licencias al reducir los costos informáticos y de licencias. AWS La OLA está diseñada para evaluar sus requisitos de licencia para las cargas de trabajo que se estén ejecutando AWS o para las cargas de trabajo cuya migración esté planificada. AWS OLA ofrece recomendaciones para optimizar el uso de las licencias.

Una de las estrategias clave para optimizar el uso de las licencias es el tamaño correcto de las instancias. El tamaño correcto implica seleccionar el tipo de instancia adecuado para su carga de trabajo en función de sus requisitos de CPU, memoria y almacenamiento. Al elegir el tamaño de instancia adecuado, puede asegurarse de utilizar los recursos de forma rentable. Esto puede suponer un importante ahorro de costes.

Con las licencias de software de Microsoft, la cantidad de núcleos en los que se ejecuta el software es un factor fundamental para determinar los costos de las licencias. Por ejemplo, las licencias de Windows Server y SQL Server se suelen licenciar según el número de núcleos. Al dimensionar las instancias correctamente, puede reducir la cantidad de núcleos en los que se ejecuta el software de Microsoft y, a su vez, reducir tanto el costo de la instancia como la cantidad de licencias necesarias.

Recomendaciones de optimización de costos

La optimización de las licencias es un componente clave de la optimización de costes. AWS Al implementar las estrategias adecuadas, puede reducir los costos de licencias, mantener el cumplimiento y obtener el mejor valor posible de su inversión en licencias. En esta sección se describen varias estrategias para la optimización de las licencias.

Traiga sus licencias de Windows Server aptas

Incorporar sus propias licencias de Windows Server es una de las estrategias más eficaces para la optimización de las licencias. Esta estrategia le permite aprovechar sus inversiones actuales para reducir sus AWS gastos.

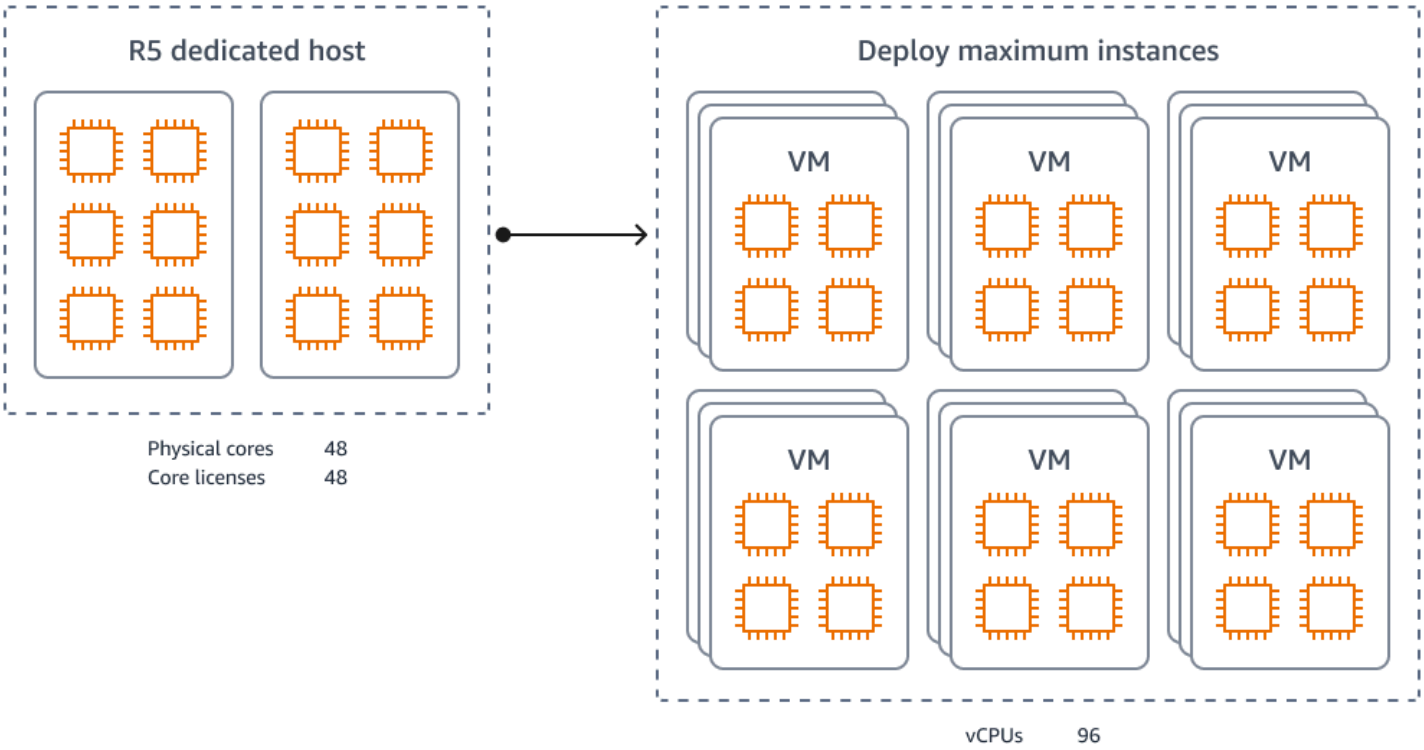
Por ejemplo, puede implementar Windows Server 2019 y versiones anteriores en Amazon EC2 Dedicated Hosts si compró las licencias antes del 1 de octubre de 2019 o las compró como ajustes en virtud de acuerdos empresariales activos firmados antes de esa fecha. Esta regla se basa en un cambio que Microsoft realizó en 2019 en sus términos y condiciones de licencia para productos sin movilidad de licencia, como Windows Server, cuando se implementaron en proveedores listados

(por ejemplo, Alibaba o Google Cloud). AWS Según las nuevas condiciones, no puedes traer tus propias licencias de Windows Server, AWS sino que debes usar instancias con licencia incluida. Sin embargo, si adquirió licencias perpetuas antes de esa fecha, aún puede implementar esas licencias de Windows Server en Amazon EC2 Dedicated Hosts.

Licencias de nivel físico

La concesión de licencias a nivel de núcleo físico le permite licenciar solo los núcleos físicos de un host, de modo que pueda implementar un número máximo de instancias sin que ello afecte al número de licencias necesarias. Por lo general, esto se hace con Windows Server Datacenter y SQL Server Enterprise Edition.

Como ejemplo, consideremos un host dedicado R5 con 48 núcleos, lo que se traduce en 96 v. CPUs Si utiliza la edición Windows Server Datacenter, solo necesitará 48 licencias. Esto le permite implementar una combinación de instancias de hasta 96 vCPUs, como se muestra en el siguiente diagrama.



Este enfoque puede resultar especialmente rentable si tiene cargas de trabajo suficientes para maximizar la cantidad de instancias que puede ejecutar en un host. Al licenciar a nivel de núcleo físico, puede evitar costes de licencia adicionales para cada instancia y obtener el mejor valor posible de su inversión en licencias.

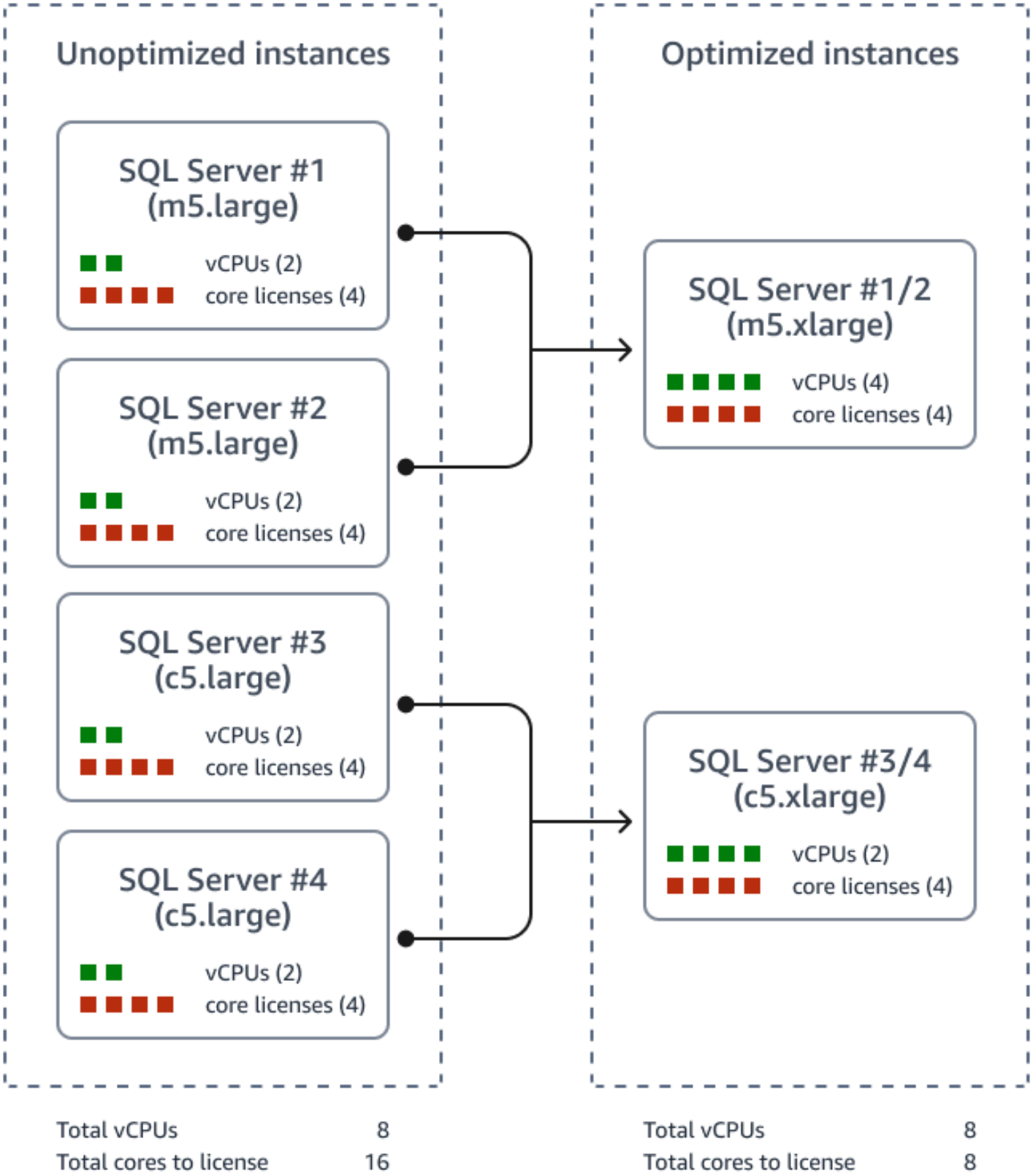
Obtenga una licencia a nivel de núcleo físico de SQL Server

En la tenencia compartida, las licencias de SQL Server se basan en la cantidad de v CPUs asignada a la instancia. Por el contrario, con los hosts dedicados, puede licenciar la edición SQL Server Enterprise a nivel de núcleo físico o a nivel de vCPU.

Al igual que en el ejemplo anterior del host dedicado R5, si licencia la edición SQL Server Enterprise a nivel de núcleo físico, solo necesitará 48 licencias de la edición SQL Server Enterprise para licenciar los hosts. Por el contrario, en el arrendamiento compartido, donde la única opción es licenciar mediante vCPU, debe tener 96 licencias de la edición Enterprise de SQL Server para la misma carga de trabajo. Por lo tanto, los hosts dedicados pueden ahorrarle hasta un 50 por ciento en los costos de licencias de SQL Server en comparación con el arrendamiento compartido. Esto se suma a ahorrar en los costes de las instancias al disponer de licencias de Windows aptas.

Consolide las instancias de SQL Server

La consolidación de SQL Server es el proceso de combinar varias instancias de SQL Server en un servidor. SQL Server requiere un mínimo de cuatro licencias principales por instancia, incluso si la instancia solo tiene dos vCPUs. Esto significa que ejecutar SQL Server en servidores con menos de cuatro núcleos puede provocar que se sobrelicencien estas instancias y se usen más licencias de las necesarias.



Total vCPUs	8
Total cores to license	16

Total vCPUs	8
Total cores to license	8

Por ejemplo, la consolidación de dos instancias con dos v CPUs cada una en una sola instancia con cuatro v CPUs puede reducir el requisito de licencias en un 50 por ciento. Esto se debe a que solo se requieren cuatro licencias principales en lugar de ocho.

Para obtener más información sobre la consolidación, consulte la sección de consolidación de SQL Server de esta guía.

Cambie a una versión anterior de las ediciones de SQL Server

Cambiar las ediciones de SQL Server puede ser una estrategia clave para optimizar el uso de las licencias y reducir los costes. La edición Enterprise de SQL Server es considerablemente más cara que la edición estándar, por lo que la actualización a una versión anterior puede suponer un importante ahorro de costes.

El cifrado de datos transparente (TDE) y los grupos de disponibilidad Always On son dos funciones populares de la edición Enterprise de SQL Server. Sin embargo, existen alternativas rentables a estas funciones que puede considerar si no necesita el conjunto completo de funciones de la edición SQL Server Enterprise. Por ejemplo, puede obtener el TDE en la edición Standard de SQL Server a partir de SQL Server 2019. En lugar de los grupos de disponibilidad Always On, puede utilizar clústeres de conmutación por error con almacenamiento compartido FSx para Windows File Server para obtener una alta disponibilidad con la edición SQL Server Standard.

Al cambiar de la edición SQL Server Enterprise a la edición SQL Server Standard, puede reducir considerablemente los costes de licencia. Para obtener más información, consulte la sección Cómo optimizar los costes de las implementaciones de alta disponibilidad de SQL Server, publicada en el AWS blog sobre almacenamiento. AWS

Además de reducir los costes de licencia, la degradación de las ediciones de SQL Server puede ayudarle a reducir sus gastos en Software Assurance y a evitar futuros ajustes. Si devuelve las licencias no utilizadas a la estantería, puede evitar costes adicionales de licencia y sacar el máximo partido posible de su inversión en licencias.

Es importante evaluar cuidadosamente las cargas de trabajo de SQL Server y determinar qué funciones son fundamentales para las necesidades de su empresa. Para obtener más información, consulte Evaluación del entorno en Guía AWS prescriptiva y determine si la base de datos de Microsoft SQL Server utiliza funciones específicas de la edición SQL Server Enterprise.

Si elige la edición correcta de SQL Server y utiliza alternativas a las funciones de la edición SQL Server Enterprise, puede lograr importantes ahorros de costos y, al mismo tiempo, mantener el

cumplimiento y satisfacer las necesidades de su empresa. Para obtener más información sobre las opciones de degradación, consulte la sección Comparación de las ediciones de SQL Server de esta guía.

Utilice la edición SQL Server Developer en entornos que no sean de producción

En entornos que no son de producción, puede implementar ediciones licenciables de SQL Server, como la edición Enterprise o Standard, mediante suscripciones a MSDN en entornos locales. Sin embargo, la suscripción a MSDN no incluye movilidad de licencias. Por lo tanto, si migras a AWS, no puedes traer esas licencias. En su lugar, debe usar la edición SQL Server Developer.

La edición SQL Server Developer es una edición con todas las funciones de SQL Server que está disponible de forma gratuita. Esta edición está disponible para las versiones 2016 y posteriores de SQL Server. Puede descargarlo del sitio web de Microsoft. La edición SQL Server Developer está diseñada para usarse en todos los entornos que no son de producción, como el desarrollo, las pruebas y la puesta en escena, siempre que no se conecte a datos de producción en vivo.

Si utiliza la edición SQL Server Developer en entornos que no son de producción, puede evitar costes de licencia adicionales. Para obtener más información, consulte la sección Evaluación de la edición SQL Server Developer de esta guía.

Optimice la CPU para las cargas de trabajo de SQL Server

En algunos casos, es posible que tengas que elegir un tipo de instancia con CPUs más instancias de las necesarias para tu carga de trabajo debido a otros factores, como los límites de RAM o de red. Sin embargo, AWS ofrece una solución que le ayuda a optimizar los costes de licencia en estas situaciones.

Al igual que la mayoría de los clientes que ofrecen licencias principales de SQL Server, puede deshabilitar el hiperprocesamiento o desactivar CPUs la EC2 instancia para limitar la cantidad de licencias disponibles CPUs para el host. Esta opción le permite aprovechar otras capacidades de la instancia, como la RAM, y, al mismo tiempo, ahorrar en el costo de adquirir licencias adicionales.

Por ejemplo, si implementa una instancia r5.4xlarge porque su carga de trabajo requiere 128 GB de memoria, pero solo necesita ocho núcleos de SQL Server, puede deshabilitar el hiperprocesamiento al lanzar una instancia con solo ocho activos. CPUs De este modo, puede ahorrar un 50 por ciento en las licencias de SQL Server necesarias, ya que solo necesita licenciar los ocho núcleos que se utilizan activamente.

Tipo de instancia	Total v CPUs	vCPU activa con función de optimizac ión CPUs	Ahorros en licencias de SQL Server
r5.4xlarge	16	8	50%
r5.12xlarge	48	8	83%

Si ajusta el tamaño adecuado a sus instancias, puede asegurarse de utilizar los tipos de instancias más rentables para sus cargas de trabajo. Al AWS introducir nuevos tipos de instancias, es importante evaluar si estas nuevas instancias pueden satisfacer los requisitos de carga de trabajo con menos núcleos.

Recursos adicionales

- Amazon Web Services y Microsoft: Preguntas frecuentes (AWS documentación)

Seleccione la EC2 instancia adecuada para las cargas de trabajo de SQL Server

> **Important**
>
> Antes de leer esta sección, le recomendamos que lea las secciones Comprenda las licencias de SQL Server y Seleccione el tipo de instancia correcto para las cargas de trabajo de Windows de esta guía.

Descripción general

Microsoft SQL Server se ha estado ejecutando en instancias de Amazon Elastic Compute Cloud (Amazon EC2) durante más de 15 años. AWS ha aprovechado esa experiencia y la ha utilizado para ayudar a desarrollar EC2 instancias de Amazon que se adapten a las cargas de trabajo de SQL Server, desde especificaciones mínimas hasta clústeres multirregionales de alto rendimiento.

La elección de la EC2 instancia correcta para SQL Server depende en gran medida de la carga de trabajo. Entender cómo se licencia SQL Server, cómo utiliza la memoria y cómo las funciones de

SQL Server se alinean con EC2 las ofertas de Amazon puede ayudarle a encontrar la mejor EC2 instancia para su aplicación.

En esta sección se abordan diversas cargas de trabajo de SQL Server y cómo pueden combinarse con determinadas EC2 instancias para reducir al mínimo los costes de licencias y procesamiento.

Comparación de costos

Amazon EC2 le permite traer su propia licencia (BYOL) o pagar por uso con las licencias de Windows Server y SQL Server. En el caso de las pay-as-you-go licencias, los costes de las licencias de Windows Server y SQL Server se incluyen en el coste por hora de la EC2 instancia. Por ejemplo, puede tener diferentes AMIs precios. El precio de la AMI depende de la edición de SQL Server en la que se ejecute la AMI.

Los precios de Windows Server y SQL Server no están detallados. No encontrará precios detallados en herramientas como la. Calculadora de precios de AWS Si selecciona diferentes combinaciones de ofertas con licencia incluida, se pueden deducir los costes de la licencia, tal y como se muestra en la siguiente tabla.

EC2 instancia	AMI	Calcule el precio	Precio de la licencia de Windows	Precio de la licencia de SQL	Precio total
r5.xlarge	Linux (precios de computación)	183,96$	-	-	183,96 DÓLARES
r5.xlarge	Desarrollador de Linux + SQL	183,96$	$0	$0	183,96 DÓLARES
r5.xlarge	Windows Server (LI)	183,96 DÓLARES	134,32 DÓLARES	-	318,28 DÓLARES
r5.xlarge	Desarrollador de Windows + SQL	183,96$	134,32 DÓLARES	$0	318,28 DÓLARES

EC2 instancia	AMI	Calcule el precio	Precio de la licencia de Windows	Precio de la licencia de SQL	Precio total
r5.xlarge	Windows y SQL Web (LI)	183,96$	134,32 DÓLARES	49,64 DÓLARES	367,92 DÓLARES
r5.xlarge	Windows + SQL Standard (LI)	183,96$	134,32 DÓLARES	350,4 DÓLARES	668,68 DÓLARES
r5.xlarge	Windows y SQL Enterprise (LI)	183,96 DÓLARES	134,32 DÓLARES	1095 DÓLARES	1413,28 DÓLARES

> **Note**
>
> Los precios de la tabla anterior se basan en los precios bajo demanda de la región. `us-east-1`

El método más rentable para ejecutar SQL Server es utilizar una edición de nivel inferior hasta que necesite una función de una edición de nivel superior. Para obtener más información, consulte la sección Comparación de las ediciones de SQL Server de esta guía. La actualización de la edición Web de SQL Server a la edición Standard de SQL Server cuesta más de siete veces el coste de licencia de SQL Server y más del triple del coste de pasar de la edición Standard a la edición Enterprise. La disparidad en los costos de las licencias es un factor importante a tener en cuenta y se analiza en el resto de esta sección.

Escenario de optimización de costes

Considere un ejemplo de escenario en el que una empresa de análisis que rastrea los vehículos de entrega busca mejorar el rendimiento de SQL Server. Una vez que un experto de MACO analiza los problemas de rendimiento de la empresa, la empresa pasa de las instancias x1e.2xlarge a las instancias x2iedn.xlarge. Si bien el tamaño de la instancia es menor, las mejoras introducidas en las instancias x2 mejoran el rendimiento y la optimización de SQL Server mediante el uso de

extensiones de grupos de búferes. Esto permitió a la empresa pasar de la edición Enterprise de SQL Server a la edición Standard y reducir sus licencias de SQL Server de 8 CPUs a 4 versiones. CPUs

Antes de la optimización:

Server	EC2 instancia	Edición de SQL Server	Costo mensual
Prod DB1	x1e.2xlarge	Enterprise	3.918,64 DÓLARES
Prod DB2	x1e.2xlarge	Enterprise	3.918,64 DÓLARES
Total			7.837,28 DÓLARES

Tras la optimización:

Server	EC2 instancia	Edición de SQL Server	Costo mensual
Prod DB1	x2iedn.xlarge	Estándar	1.215,00$
Orgulloso DB2	x2iedn.xlarge	Estándar	1.215,00$
Total			2.430,00 DÓLARES

Los cambios combinados de las instancias x1e.2xlarge a las instancias x2iedn.xlarge permitieron al cliente del ejemplo ahorrar 5.407$ al mes en sus servidores de bases de datos de producción. Esto redujo el costo total de la carga de trabajo en un 69 por ciento.

> **Note**
>
> Los precios de la tabla anterior se basan en los precios bajo demanda de la `us-east-1` región.

Recomendaciones de optimización de costos

instancias optimizadas para memoria

Uno de los aspectos más importantes de SQL Server es comprender su dependencia de la memoria. SQL Server intenta utilizar toda la RAM disponible que no esté siendo utilizada por el sistema operativo (hasta 2 TB para una instalación predeterminada). Lo hace por motivos de rendimiento. Trabajar con datos en la memoria es mucho más eficaz que tener que extraer datos del disco constantemente, realizar cambios y volver a escribirlos en el disco. En su lugar, SQL Server intenta cargar tantos datos de las bases de datos adjuntas como sea posible y los guarda en la RAM. Los cambios realizados en los datos se producen en la memoria y, posteriormente, se almacenan en el disco.

> **Note**
>
> Para obtener una explicación detallada de cómo SQL Server escribe los cambios, consulte Cómo escribir páginas en la documentación de Microsoft.

Dado que SQL Server funciona mejor con grandes cantidades de RAM, normalmente recomendamos empezar con los tipos de instancias optimizadas para EC2 memoria de Amazon. Las instancias con memoria optimizada son versátiles y ofrecen una variedad de opciones diferentes. La familia R tiene una vCPU-to-RAM proporción de 1 a 8 y cuenta con opciones para procesadores Intel, procesadores AMD, redes mejoradas, rendimiento EBS mejorado, almacenamiento de instancias y velocidad de procesador mejorada. Para cargas de trabajo con un uso intensivo de memoria, también hay una familia X que combina muchas de las mismas opciones y amplía la proporción de 1 a 32. vCPU-to-RAM Gracias a la versatilidad de las instancias optimizadas para memoria, puede aplicarlas a cargas de trabajo de SQL Server de todas las formas y tamaños.

Cargas de trabajo por debajo del mínimo de recursos (menos de 4 v) CPUs

Si bien algunos casos de uso funcionan bien con las instancias de fragmentación (T3), le recomendamos que, en general, evite el uso de instancias de fragmentación para las cargas de trabajo de SQL Server. La licencia de SQL Server se basa en el número de v CPUs asignado a una instancia. Si SQL Server está inactivo la mayor parte del día y está adquiriendo créditos por ráfaga, tendrá que pagar por las licencias de SQL que no utilice en su totalidad. Además, SQL Server requiere una licencia mínima de 4 núcleos por servidor. Esto significa que si tiene una carga de trabajo de SQL Server que no requiere una potencia de cómputo equivalente a 4 VCPUs , está

pagando una licencia de SQL Server que no está utilizando. En estos escenarios, lo mejor sería consolidar varias instancias de SQL Server en un servidor más grande.

Cargas de trabajo que utilizan recursos mínimos (menos de 64 GB de RAM)

Muchas cargas de trabajo de SQL Server de menos de 64 GB de RAM no dan prioridad al alto rendimiento o la alta disponibilidad. Para este tipo de cargas de trabajo, la edición web de SQL Server podría ser una buena opción si la aplicación está cubierta por las restricciones de licencia de Microsoft.

> **Important**
>
> La edición web de SQL Server tiene un caso de uso restringido basado en las condiciones de licencia de Microsoft. SQL Server Web Edition solo se puede usar para admitir páginas web, sitios web, aplicaciones web y servicios web públicos y accesibles a través de Internet. No se puede usar para dar soporte a line-of-business aplicaciones (por ejemplo, aplicaciones de administración de relaciones con los clientes, administración de recursos empresariales y otras aplicaciones similares).

La edición Web de SQL Server se amplía hasta 32 V CPUs y 64 GB de RAM y es un 86 por ciento más económica que la edición SQL Server Standard. Para cargas de trabajo con pocos recursos, utilizar una instancia optimizada para memoria de AMD como la r6a, que tiene un precio de procesamiento un 10 por ciento más bajo que su homóloga de Intel, también es una buena forma de reducir al mínimo los costos de procesamiento y licencias de SQL.

Cargas de trabajo con recursos medios (menos de 128 GB de RAM)

La edición SQL Server Standard se utiliza en la mayoría de las cargas de trabajo de SQL Server de hasta 128 GB de RAM. La edición SQL Server Standard es entre un 65 y un 75 por ciento más económica que la edición SQL Server Enterprise y puede ampliarse hasta 48 V CPUs y 128 GB de RAM. Dado que el límite de 128 GB de RAM suele alcanzarse antes que el límite de 48 vCPU, es el objetivo de la mayoría de los clientes que desean evitar la actualización a la edición SQL Server Enterprise.

SQL Server tiene una función llamada extensión de grupo de búferes. Esta función permite a SQL Server utilizar una parte de un disco como extensión de la RAM. La extensión del grupo de búferes funciona bien cuando se combina con un almacenamiento ultrarrápido, como el que NVMe SSDs

se usa en el almacenamiento de EC2 instancias de Amazon. Las EC2 instancias de Amazon que contienen almacenamiento de instancias se indican con una «d» en el nombre de la instancia (por ejemplo, r5d, r6id y x2iedn).

Las extensiones del grupo de búferes no sustituyen a la RAM normal. Sin embargo, si necesitas más de 128 GB de RAM, puedes usar extensiones de búfer con EC2 instancias como la r6id.4xlarge y la x2iedn.xlarge para retrasar la actualización a las licencias de la edición Enterprise.

Cargas de trabajo de alto rendimiento (más de 128 GB de RAM)

Las cargas de trabajo de SQL Server que requieren un alto rendimiento representan un desafío para la optimización de costos debido a que dependen de una gran cantidad de recursos. Sin embargo, comprender las diferencias entre las EC2 instancias puede evitar que tome una decisión equivocada.

En la siguiente tabla, se muestran diversas EC2 instancias con memoria optimizada y sus límites de rendimiento.

	r5b	r6idn	r7iz	x2iedn	x2iezn
Procesador	3.1 GHz Procesador Intel Xeon de segunda generación	3.5 GHz Procesador Intel Xeon de tercera generación	3.9 GHz Procesador escalable Intel Xeon de cuarta generación	3.5 GHz Procesador Intel Xeon de tercera generación	4.5 GHz Procesador Intel Xeon de segunda generación
Relación CPU:RAM	1:8	1:8	1:8	1:32	1:32
vCPU máxima	96	128	128	128	48
RAM máxima	768 GB	1.024 GB	1.024 GB	4.096 GB	1.536 GB
Almacenamiento de instancias	–	NVMe SSD (4 x 1900 GB)	–	NVMe SSD (2 x 1900 GB)	–

	r5b	r6idn	r7iz	x2iedn	x2iezn
io2 Block Express	Soportado	Soportado	Soportado	Compatible	–
IOPS máximas de EBS	260 000	350.000	160 000	260 000	80 000
Rendimiento máximo de EBS	60 Gbps	80 Gbps	40 Gbps	80 Gbps	19 Gbps
Ancho de banda máximo de red	25 Gbps	200 Gbps	50 Gbps	100 Gbps	100 Gbps

Cada instancia se usa para un propósito diferente. Comprender la carga de trabajo de SQL Server puede ayudarle a elegir el tipo de instancia que mejor se adapte a sus necesidades.

Detalles sobre los atributos:

- r5b: el atributo «b» de r5b significa que este tipo de instancia se centra en un alto rendimiento de EBS. En la quinta generación de instancias optimizadas para memoria, la r5b era la opción preferida. Fue el primer tipo de instancia que utilizó los volúmenes de io2 Block Express y alcanzó un máximo de 260.000 IOPS de almacenamiento. El tipo de instancia r5b sigue siendo una alternativa rentable para las necesidades de alto rendimiento de EBS.
- r6idn: la sexta generación de instancias optimizadas para memoria ofreció mejoras considerables con respecto a la generación anterior. Las mejoras de rendimiento de EBS derivadas de la r5b van un paso más allá con la r6idn, que eleva el máximo de IOPS a 350 000. El r6idn también tiene un volumen de almacenamiento de instancias para las extensiones tempdb y de grupos de búferes a fin de aumentar aún más el rendimiento de SQL Server.
- x2iedn: el x2iedn es similar al r6idn. Ofrece niveles similares de EBS mejorado, redes mejoradas y almacenamiento de instancias NVMe SSD mejorados, pero con una vCPU-to-RAM proporción de 1:32 para cargas de trabajo con gran cantidad de memoria y baja cantidad de CPU (menores costos de licencia de SQL Server).

- x2iezn: el atributo «z» de x2iezn indica que este tipo de instancia se centra en un alto rendimiento del procesador. El procesador Cascade Lake tiene una frecuencia turbo integral de hasta 4,5. GHz Le recomendamos que utilice esta EC2 instancia, junto con una vCPU-to-RAM proporción de 1:32, en un escenario en el que desee mantener baja la cantidad de vCPU. Esto, a su vez, puede mantener bajos los costos de licencia de SQL Server.
- r7iz: el atributo «z» de r7iz indica que este tipo de instancia se centra en un alto rendimiento del procesador. El procesador Sapphire Rapids tiene una frecuencia turbo integral de hasta 3.9. GHz Al igual que las instancias x2iezn, el r7iz prioriza el rendimiento del procesador de alta frecuencia, pero con una relación de 1:8. vCPU-to-RAM

Recursos adicionales

- EC2Instancias de Amazon de uso general (AWS documentación)
- Herramienta de comparación (Vantage)
- Licencias: SQL Server (AWS documentación)

Consolide las instancias

Esta sección se centra en la técnica de optimización de costes que consiste en combinar varias instancias de SQL Server en el mismo servidor para minimizar los costes de licencia y maximizar la utilización de los recursos.

Descripción general

La creación de una instancia forma parte del proceso de instalación del motor de base de datos de SQL Server. La instancia de SQL Server es una instalación completa que contiene sus propios archivos de servidor, inicios de sesión de seguridad y bases de datos del sistema (master, model, msdb y tempdb). Como una instancia tiene todos sus propios archivos y servicios, puede instalar varias instancias de SQL Server en el mismo sistema operativo sin que las instancias interfieran entre sí. Sin embargo, dado que todas las instancias están instaladas en el mismo servidor, todas comparten los mismos recursos de hardware, como el procesamiento, la memoria y las redes.

Lo habitual es utilizar solo una instancia de SQL Server por servidor en los entornos de producción, de modo que una instancia «ocupada» no utilice en exceso los recursos de hardware compartidos. Ofrecer a cada instancia de SQL Server su propio sistema operativo, con sus propios recursos, es un límite mejor que confiar en la gobernanza de los recursos. Esto es especialmente cierto en el caso

de las cargas de trabajo de SQL Server de alto rendimiento que requieren grandes cantidades de recursos de RAM y CPU.

Sin embargo, no todas las cargas de trabajo de SQL Server utilizan una gran cantidad de recursos. Por ejemplo, algunas organizaciones asignan a cada uno de sus clientes su propia instancia de SQL Server dedicada por motivos de cumplimiento o seguridad. En el caso de clientes más pequeños o clientes que no suelen estar activos, eso significa ejecutar las instancias de SQL Server con un mínimo de recursos.

Como se indica en la guía de licencias de Microsoft SQL Server 2019, cada servidor que ejecute SQL Server debe tener un mínimo de cuatro licencias de CPU. Esto significa que incluso si ejecuta un servidor con solo dos vCPUs, debe seguir licenciando SQL Server para cuatro vCPUs. Según el precio público de SQL Server de Microsoft, esto supone una diferencia de 3.945$ si utilizas la edición estándar de SQL Server. Para las organizaciones que utilizan varios servidores con una única instancia de SQL Server y utilizan un mínimo de recursos, el coste combinado de tener que licenciar los recursos no utilizados puede ser considerable.

Escenario de optimización de costes

En esta sección se analiza un ejemplo de escenario en el que se compara la diferencia entre ejecutar cuatro servidores de Windows Server, cada uno con una única instancia de SQL Server, con un único servidor de Windows Server de mayor tamaño que ejecuta varias instancias de SQL Server simultáneamente.

Si cada instancia de SQL Server solo necesita dos v CPUs y 8 GB de RAM, el coste total por servidor es de 7.890$ para la licencia de SQL Server, además de un coste informático por hora de 0,096$.

EC2 instancia	v CPUs	RAM	Precio	v CPUs para licenciar	Coste total de las licencias de SQL Server
m6i.large	2	8	0,096	4	7.890 DÓLARES

Si se amplía esta cifra a cuatro servidores, el coste total de la licencia de SQL Server es de 31 560 dólares, con un coste informático por hora de 0,384 dólares.

EC2 instancia	v CPUs	RAM	Precio	v CPUs para licenciar	Coste total de las licencias de SQL Server
4 veces 6 mi.large	2	32	0,384	16	31.560 DÓLARES

Si combina las cuatro instancias de SQL Server en una sola EC2 instancia, la cantidad total de recursos de cómputo y de cómputo permanece igual. Sin embargo, al eliminar los costos innecesarios de licencias de SQL Server, puede reducir el costo total de ejecutar la carga de trabajo en 15 780$.

EC2 instancia	v CPUs	RAM	Precio	v CPUs para licenciar	Coste total de las licencias de SQL Server
m6i.2xlarge	8	32	0,384	8	15.780 DÓLARES

> **Note**
>
> En las tablas anteriores, los costos de procesamiento muestran los precios por hora y bajo demanda para EC2 los servidores de Amazon que ejecutan Windows Server en la `us-east-1` región. Los costos de licencia de la edición estándar de SQL Server se refieren a los precios públicos de SQL Server de Microsoft.

Recomendaciones de optimización de costos

Si está pensando en consolidar las instancias de SQL Server, lo que más le preocupa es el consumo de recursos de cada una de las instancias que desee consolidar. Es importante obtener métricas de rendimiento durante períodos prolongados para comprender mejor los patrones de carga de trabajo de cada servidor. Algunas herramientas comunes para la supervisión del consumo de recursos son

Amazon CloudWatch, Windows Performance Monitor (perfmon) y las herramientas de supervisión nativas de SQL Server.

Le recomendamos que tenga en cuenta las siguientes preguntas al analizar si sus cargas de trabajo de SQL Server se pueden combinar para utilizar los mismos recursos del servidor sin que interfieran entre sí:

- ¿Qué recursos (CPU, memoria y ancho de banda de red) se consumen durante su estado estable?
- ¿Qué recursos (CPU, memoria y ancho de banda de red) se consumen durante los picos?
- ¿Con qué frecuencia se producen picos? ¿Los picos son consistentes?
- ¿Los picos de recursos de un servidor coinciden con los picos de recursos de otro servidor?
- ¿Cuáles son las IOPS de almacenamiento y el rendimiento que utiliza SQL Server?

Si desea seguir adelante con un plan para combinar instancias de SQL Server, consulte la publicación Ejecutar varias instancias de SQL Server en una EC2 instancia de Amazon en el blog AWS Cloud Operations & Migrations. En esta publicación, se proporcionan instrucciones sobre cómo realizar los cambios de configuración en SQL Server para añadir instancias adicionales. Antes de empezar, tenga en cuenta las pequeñas diferencias que se producen cuando se instalan varias instancias en el mismo servidor:

- La instancia de base de datos de SQL Server predeterminada tiene un nombre `MSSQLSERVER` y utiliza el puerto 1433.
- Cada instancia adicional instalada en el mismo servidor es una instancia de base de datos «con nombre».
- Cada instancia con nombre tiene un nombre de instancia y un puerto únicos.
- El navegador de SQL Server debe ejecutarse para coordinar el tráfico a las instancias nombradas.
- Cada instancia puede usar ubicaciones distintas para los archivos de datos de la base de datos y los inicios de sesión independientes.
- La configuración máxima de memoria del servidor SQL Server debe configurarse de acuerdo con las necesidades de rendimiento de cada instancia, y su total combinado también debe dejar suficiente memoria para el sistema operativo subyacente.
- Puede utilizar las capacidades de backup y restauración nativas de SQL Server o AWS DMSpara la migración o la consolidación.

Recursos adicionales

- Hoja de datos de licencias de SQL Server (blog sobre operaciones y migraciones en la AWS nube)
- Entrada del blog sobre la configuración de instancias múltiples de SQL Server (blog sobre operaciones y migraciones en la AWS nube)
- Guía de prácticas recomendadas de SQL Server (documentación de orientación AWS prescriptiva)

Compare las ediciones de SQL Server

Descripción general

Las licencias de Microsoft SQL Server son uno de los mayores gastos de un entorno de carga de trabajo de Windows. Los costos de licencia de SQL Server pueden extenderse fácilmente más allá de los costos de cómputo necesarios para ejecutar la carga de trabajo. Si eliges la edición equivocada, podrías pagar por funciones que no utilizas o que ni siquiera necesitas. En esta sección se comparan las siguientes ediciones de SQL Server, incluidas sus características y costes relativos:

- Enterprise: la edición SQL Server Enterprise proporciona capacidades de centro de datos con alto rendimiento, virtualización ilimitada y varias herramientas de inteligencia empresarial (BI).
- Estándar: la edición estándar de SQL Server proporciona administración de datos básica e inteligencia empresarial para organizaciones y departamentos más pequeños.
- Web: la edición web de SQL Server es adecuada para empresas que ofrecen alojamiento web o proveedores de valor agregado web (VAPs). Esta edición ofrece un coste total de propiedad bajo y proporciona capacidades de escalabilidad y capacidad de administración para propiedades web de pequeña y gran escala.

  > **Important**
  >
  > Puede usar la edición web de SQL Server para admitir únicamente páginas web, sitios web, aplicaciones web y servicios web públicos y accesibles desde Internet. No puede usar la edición web de SQL Server para admitir line-of-business aplicaciones (como las aplicaciones de administración de relaciones con los clientes o de administración de recursos empresariales).

- Desarrollador: la edición SQL Server Developer incluye todas las funciones de la edición Enterprise, pero está destinada únicamente a fines de desarrollo.

- Express: la edición SQL Server Express es una base de datos gratuita y se puede utilizar para aprender o para crear aplicaciones de escritorio. Puede actualizar la edición Express a otras ediciones.

> **Note**
> La edición SQL Server Evaluation está disponible durante un período de prueba de 180 días.

Impacto del costo

Puede comprar licencias de SQL Server en un distribuidor de Microsoft y llevárselas AWS con Software Assurance. Como alternativa, puede utilizar las licencias de SQL Server con un pay-as-you-go modelo que incluya la licencia de Amazon. EC2 AMIs

Si compra licencias de SQL Server a distribuidores de Microsoft, las licencias principales se venden en paquetes de dos y debe licenciar un mínimo de cuatro núcleos por servidor. La siguiente tabla muestra una comparación de costos entre las ediciones Enterprise y Standard.

Versión	Edición SQL Server Enterprise (paquete de 2 núcleos)	Edición SQL Server Standard (paquete de 2 núcleos)	Ahorro
2022	15.123 DÓLARES	3.945 DÓLARES	74%
2019	13.748 DÓLARES	3.586 DÓLARES	74%

> **Note**
> Los precios de la tabla anterior se basan en los precios públicos de Microsoft para SQL Server 2022 y SQL Server 2019.

La siguiente comparación de costos muestra el alojamiento de diferentes ediciones de SQL Server con Amazon con licencia incluida. EC2 AMIs En esta comparación, SQL Server está alojado en r6i.xlarge (4 vCPU) de la región. `us-east-1`

instancia	Costo de cómputo	Coste de la licencia de Windows	Coste de la licencia de SQL Server	Total
R6i.xlarge (Linux)	183,96$	–	–	183,96 DÓLARES
R6i.XLarge + Windows	183,96$	134,32 DÓLARES	–	318,28 DÓLARES
R6I.xlarge + edición web de SQL Server	183,96$	134,32 DÓLARES	49,35 DÓLARES	367,63 DÓLARES
R6i.xlarge + edición estándar de SQL Server	183,96$	134,32 DÓLARES	350,4 DÓLARES	668,68 DÓLARES
Edición R6I.xlarge + SQL Enterprise	183,96$	134,32 DÓLARES	1.095 DÓLARES	1.413,28 DÓLARES

Puede ahorrar hasta un 95 por ciento en los costos de licencia de SQL Server si selecciona la edición de SQL Server adecuada para su carga de trabajo. En la siguiente tabla se compara el costo de las licencias de SQL Server en las instancias r6i.xlarge.

Edición	% de ahorro
Estándar en comparación con Enterprise	68%
Web comparada con Standard	86%
Web comparada con Enterprise	95%

En la mayoría de los escenarios, las organizaciones cambian de la edición Enterprise a la Standard, pero hay algunos casos en los que es posible cambiar de la edición Standard o Enterprise a la edición Web.

Recomendaciones de optimización de costos

Puede elegir la mejor edición para su carga de trabajo en función de los límites de escalabilidad, la alta disponibilidad, el rendimiento y la seguridad. En la siguiente tabla se muestran las funciones compatibles con todas las ediciones de SQL Server. Esto puede ayudarle a decidir qué edición usar. Esta comparación se aplica a SQL Server 2016 SP1 y versiones posteriores.

Límites de escalado

En la siguiente tabla se comparan los límites de escalado de las distintas ediciones de SQL Server.

Característica	Enterprise Edition	Standard Edition	Edición web	Edición exprés
Capacidad de cómputo máxima utilizada por una sola instancia de SQL Server Database Engine, SQL Server Analysis Services (SSAS) o SQL Server Reporting Services (SSRS)	Sistema operativo máximo	Limitado a un mínimo de 4 sockets o 24 núcleos	Limitado a un mínimo de 4 sockets o 16 núcleos	Limitado a un número menor de 4 sockets o 4 núcleos
Memoria máxima para el grupo de búferes por instancia del motor de base de datos de SQL Server	Sistema operativo máximo	128 GB	64 GB	1410 MB
Capacidad máxima de extensión del	Memoria configurada 32	Memoria configurada 4	N/A	N/A

Característica	Enterprise Edition	Standard Edition	Edición web	Edición exprés
grupo de búferes por instancia del motor de base de datos de SQL Server	veces como máximo	veces como máximo		
Tamaño máximo de la base de datos relacional	524 PB	524 PB	524 PB	10 GB
Memoria máxima para cachés de almacenamiento de columnas o datos optimizados para memoria	Sistema operativo máximo	32 GB	16 GB	352 MB

Si su aplicación requiere menos de 16 núcleos (32 vCPUs) y 64 GB de RAM, puede empezar a evaluarla desde la edición web de SQL Server. Si su carga de trabajo requiere más de 64 GB de memoria u otras opciones de alta disponibilidad, debe actualizar a la edición SQL Server Standard.

Puede usar la edición web de SQL Server para admitir páginas web, sitios web, aplicaciones web y servicios web públicos y accesibles a través de Internet, pero no puede usar la edición web de SQL Server para admitir aplicaciones de línea empresarial. Para obtener más información acerca de los casos de uso de SQL Server Web Edition, póngase en contacto con Microsoft Licensing Support o con su distribuidor de Microsoft.

Puede usar la edición SQL Server Standard para cargas de trabajo de hasta 24 núcleos (48 vCPUs) y 128 GB de memoria. Sin embargo, puede usar extensiones de grupos de búferes para permitir que la edición SQL Server Standard utilice el almacenamiento de instancias local, como los presentes en las instancias r6id EC2 . Esto amplía la memoria hasta un tamaño cuatro veces superior al máximo de la configuración de memoria. Esta combinación de características puede retrasar la actualización de un servidor a la edición Enterprise cuando los requisitos de memoria comienzan a aumentar.

Para identificar el uso de la memoria, busque las páginas de las bases de datos en el conjunto de búferes y los contadores de esperanza de vida útil de las páginas. La esperanza de vida de la página indica cuánto tiempo permanece la página en la memoria antes de volver a guardarla en el disco. El valor predeterminado de este contador es 300. Si las páginas permanecen en la memoria durante horas o días, existe la posibilidad de que se reduzca la memoria asignada.

Alta disponibilidad

En la siguiente tabla se comparan las capacidades de alta disponibilidad de las distintas ediciones de SQL Server.

Característica	Enterprise Edition	Standard Edition	Edición web	Edición exprés
Soporte básico para servidores 1	Sí	Sí	Sí	Sí
Envío de registros	Sí	Sí	Sí	No
Duplicación de bases de datos	Sí	Modo de seguridad COMPLETO	Solo como testigo	Solo como testigo
Compresión de Backup	Sí	Sí	No	No
Instancias de clúster de conmutación por error de Always On	16 nodos	2 nodos	No	No
Grupos de disponibilidad AlwaysOn	Hasta 8 réplicas secundari as, incluidas 2 réplicas	No	No	No

Característica	Enterprise Edition	Standard Edition	Edición web	Edición exprés
	secundarias sincrónicas			
Grupos de disponibilidad básicos	No	2 nodos	No	No
Restauración de páginas y archivos en línea	Sí	No	No	No
Indexación en línea	Sí	No	No	No
Cambio de esquema online	Sí	No	No	No
Recuperación rápida	Sí	No	No	No
Copias de seguridad duplicadas	Sí	No	No	No
No añadir memoria ni CPU	Sí	No	No	No
Copia de seguridad cifrada	Sí	Sí	No	No

Característica	Enterprise Edition	Standard Edition	Edición web	Edición exprés
Copia de seguridad híbrida en Microsoft Azure (copia de seguridad en URL)	Sí	Sí	No	No
Servidor de conmutación por error para recuperación ante desastres	Sí	Sí	No	No
Servidores de conmutación por error para una alta disponibi lidad	Sí	Sí	No	No

Otras características comunes

En la siguiente tabla se comparan las características más comunes de las distintas ediciones de SQL Server. Para obtener una lista exhaustiva de funciones, consulte Ediciones y funciones compatibles de SQL Server 2019 en la documentación de Microsoft.

Característica	Enterprise Edition	Standard Edition	Edición web	Edición exprés
(Rendimiento) Regulador de recursos	Sí	No	No	No

Característica	Enterprise Edition	Standard Edition	Edición web	Edición exprés
(Seguridad) Cifrado transparente de bases de datos (TDE)	Sí	Sí	No	No
(Seguridad) Administración de claves ampliable (EKM)	Sí	No	No	No
(Replicación) Publicación de Oracle	Sí	No	No	No
(Replicación) Replicación transaccional punto a punto	Sí	No	No	No
Captura de datos de cambios	Sí	Sí	No	No

Edición SQL Server Developer

Todas las cargas de trabajo ajenas a la producción, como los entornos de desarrollo, control de calidad, pruebas, puesta en escena y UAT, pueden utilizar la edición SQL Server Developer para ahorrar un 100 por ciento en los costos de licencia de SQL Server. Después de descargar SQL Server, puede instalar la edición SQL Server Developer en las EC2 instancias mediante la tenencia compartida. La edición SQL Server Developer no requiere una infraestructura dedicada. Para obtener más información, consulte la recomendación de esta guía para la edición SQL Server Developer.

Cambiar de edición

En el caso de las cargas de trabajo existentes, el cambio de una edición a otra requiere pruebas exhaustivas. Se recomienda comprobar las cargas de trabajo que se ejecutan en las ediciones Enterprise o Standard para comprobar si se utilizan funciones específicas de cada edición y si hay alguna solución alternativa para esas funciones. Por ejemplo, si desea comprobar si sus bases de datos utilizan alguna función de nivel empresarial, puede ejecutar vistas de administración dinámica (DMV) en todas las bases de datos, tal y como se muestra en el siguiente comando de ejemplo.

```
SELECT feature_name FROM sys.dm_db_persisted_sku_features; GO
```

Hay algunas funciones de la edición Enterprise que no se pueden capturar en T-SQL, como la reindexación en línea como parte de las tareas de mantenimiento de SQL. Estas deben verificarse manualmente.

Consideraciones sobre la migración

La forma en que otorgue la licencia a SQL Server determinará sus opciones para cambiar de edición. AMIs, incluido SQL Server AMIs, incluyen el costo de la licencia en el precio de la EC2 instancia; el costo de la licencia está vinculado a la AMI. Puede utilizar los códigos AWS de facturación para comprobar la versión de SQL Server incluida en la AMI. En el caso AWS de las instancias con licencia incluida, cambiar la edición de SQL Server en el sistema operativo no cambiará la facturación asociada a la AMI. Debe migrar las bases de datos a una nueva EC2 instancia con una AMI que ejecute la nueva edición de SQL Server.

Si trae su propia licencia, dispondrá de más flexibilidad. Por lo general, se recomienda migrar a otra EC2 instancia que ejecute la nueva versión. Esto permite una recuperación rápida si algo no sale según lo planeado. Sin embargo, si debe utilizar el servidor existente, puede realizar una side-by-side instalación de SQL Server y migrar las bases de datos de una instancia a otra. Para obtener información más detallada sobre las side-by-side versiones anteriores a una versión anterior, consulte Actualización y degradación de ediciones en SQL Server en el MSSQLTips sitio web.

Recursos adicionales

- Ediciones y funciones compatibles de SQL Server 2022 (Microsoft Learn)
- sys.dm_db_persisted_sku_features (Transact-SQL) (Microsoft Learn)
- ¿Qué versión de SQL Server debe utilizar? (Brent Ozar Unlimited)
- Calculadora de precios de AWS (AWS)

Evalúe la edición SQL Server Developer

Descripción general

La edición SQL Server Developer es una edición gratuita de SQL Server que contiene todas las funciones de la edición Enterprise y se puede utilizar en cualquier entorno que no sea de producción. En la nube, donde no se pueden usar las licencias de Microsoft Developer Network (MSDN), la edición SQL Server Developer es una buena forma de ahorrar costos sin tener que proporcionar licencias para las cargas de trabajo de desarrollo y prueba. Esto es especialmente cierto para los equipos que gestionan grandes entornos de desarrollo y pruebas y buscan reducir costes innecesarios.

Un entorno de producción se define como un entorno al que acceden los usuarios finales de una aplicación (como un sitio web de Internet) y que se utiliza para algo más que recopilar comentarios o realizar pruebas de aceptación de esa aplicación. Otros escenarios que constituyen entornos de producción incluyen:

- Entornos que se conectan a una base de datos de producción
- Entornos que admiten la recuperación ante desastres o el respaldo para un entorno de producción
- Entornos que se utilizan para la producción al menos una parte del tiempo, como un servidor que pasa a estar en producción durante los períodos de mayor actividad

Para obtener más información sobre licencias, consulte Amazon Web Services y Microsoft: Preguntas frecuentes en la AWS documentación.

Impacto del costo

Si utiliza la edición SQL Server Developer para cargas de trabajo que no son de producción, puede ahorrar el 100 por ciento de los costes actuales de licencias de SQL Server para entornos de desarrollo y pruebas.

Versión de SQL Server	Edición SQL Server Enterprise (paquete de 2 núcleos)	Edición SQL Server Standard (paquete de 2 núcleos)	Edición SQL Server Developer
2022	15.123 DÓLARES	3.945 DÓLARES	Free

Versión de SQL Server	Edición SQL Server Enterprise (paquete de 2 núcleos)	Edición SQL Server Standard (paquete de 2 núcleos)	Edición SQL Server Developer
2019	13.748 DÓLARES	3.586 DÓLARES	Free

> **Note**
>
> Los precios de la tabla anterior se basan en los precios públicos de Microsoft para SQL Server 2022 y SQL Server 2019.

En la siguiente tabla se compara el costo de las diferentes ediciones de SQL Server que funcionan con 4 v CPUs y utilizan precios bajo demanda en la `us-east-2` región. Esto se aplica a los escenarios que dependen de instancias con licencia incluida de. AWS

EC2 instancia	AMI	Calcule el precio	Precio de la licencia de Windows	Precio de la licencia de SQL Server	Precio total
r5.xlarge	Linux (precios de computación)	183,96$	–	–	183,96 DÓLARES
r5.xlarge	Edición Linux + SQL Server Developer	183,96$	$0	$0	183,96 DÓLARES
r5.xlarge	Windows Server (LI)	183,96 DÓLARES	134,32 DÓLARES	–	318,28 DÓLARES
r5.xlarge	Edición para desarroll adores de Windows + SQL Server	183,96$	134,32 DÓLARES	$0	318,28 DÓLARES

EC2 instancia	AMI	Calcule el precio	Precio de la licencia de Windows	Precio de la licencia de SQL Server	Precio total
r5.xlarge	Edición web (LI) de Windows y SQL Server	183,96$	134,32 DÓLARES	49,64 DÓLARES	367,92 DÓLARES
r5.xlarge	Windows + SQL Server Standard Edition (LI)	183,96$	134,32 DÓLARES	350,4 DÓLARES	668,68 DÓLARES
r5.xlarge	Windows y SQL Server Enterprise Edition (LI)	183,96$	134,32 DÓLARES	1095 DÓLARES	1413,28 DÓLARES

Escenario de optimización de costes

Después de que una empresa de integridad de datos realizara una nueva adquisición, quiso migrar la carga de trabajo recién adquirida de su ubicación actual a un proveedor de alojamiento gestionado para consolidarla con las demás cargas de trabajo del Nube de AWS. Los precios iniciales mostraron que la carga de trabajo de SQL Server de la empresa costaría un 60 por ciento más de funcionamiento AWS que con el proveedor de servicios gestionados actual. Una pyme de MACO evaluó la estimación y descubrió que, en realidad, el cliente estaba pagando las licencias de SQL Server al proveedor de alojamiento gestionado para sus entornos de desarrollo y pruebas. Al cambiar las cargas de trabajo no relacionadas con la producción a la edición SQL Server Developer durante la migración, la empresa redujo sus licencias de SQL Server en un 40 por ciento.

Licencia de SQL Server incluida en Amazon EC2

Si tienes SQL Server en EC2 instancias que utilizan la licencia incluida AMIs, no es posible realizar una conversión directa de la edición Enterprise a la edición Developer. Los costos de licencia de las instancias con licencia incluida están sujetos a la AMI. Aunque SQL Server se desinstale desde el sistema operativo, se seguirán cobrando los costes de licencia de la EC2 instancia.

Para convertirla a la edición para desarrolladores, debe descargar la edición para desarrolladores de SQL Server, instalarla en una EC2 instancia nueva y, a continuación, migrar las bases de datos. Puede migrar bases de datos de SQL Server entre EC2 instancias mediante diversos métodos. Para obtener más información, consulte Métodos de migración de bases de datos de SQL Server en la guía Migración de bases de datos de Microsoft SQL Server a la Nube de AWS guía. También puede usar la solución Automated SQL Server Developer para preparar la nueva instancia a la que planea migrar.

BYOL de SQL Server en Amazon EC2

Si tiene instancias de SQL Server que utilizan un BYOL, puede elegir entre las siguientes opciones de conversión o side-by-side degradación in situ:

- Descargue la edición SQL Server Developer desde el sitio web de Microsoft. Para obtener instrucciones de instalación manual o automática, consulte la publicación del blog sobre cómo automatizar las implementaciones de SQL Server Developer. AWS
- Utilice el backup y la restauración nativos de SQL Server para migrar bases de datos o separar/adjuntar bases de datos de una instancia de SQL a otra.
- Utilice una herramienta de automatización para las implementaciones masivas.

> **Note**
>
> La edición SQL Server Developer es solo para entornos que no son de producción.

Recursos adicionales

- Automatizar las implementaciones de SQL Server Developer para implementar SQL Server Developer Edition en EC2 (Blog)AWS
- Precios de SQL 2022 (Microsoft)
- Precios de SQL 2019 (Microsoft)
- Opciones de licencia (SQL Server en Amazon EC2)
- Calculadora de precios de AWS(EC2 documentación de SQL Server en Amazon)
- Guía de licencias de Microsoft SQL Server 2019 (descarga de Microsoft)
- Edición para desarrolladores de SQL Server 2022 (descarga desde Microsoft)

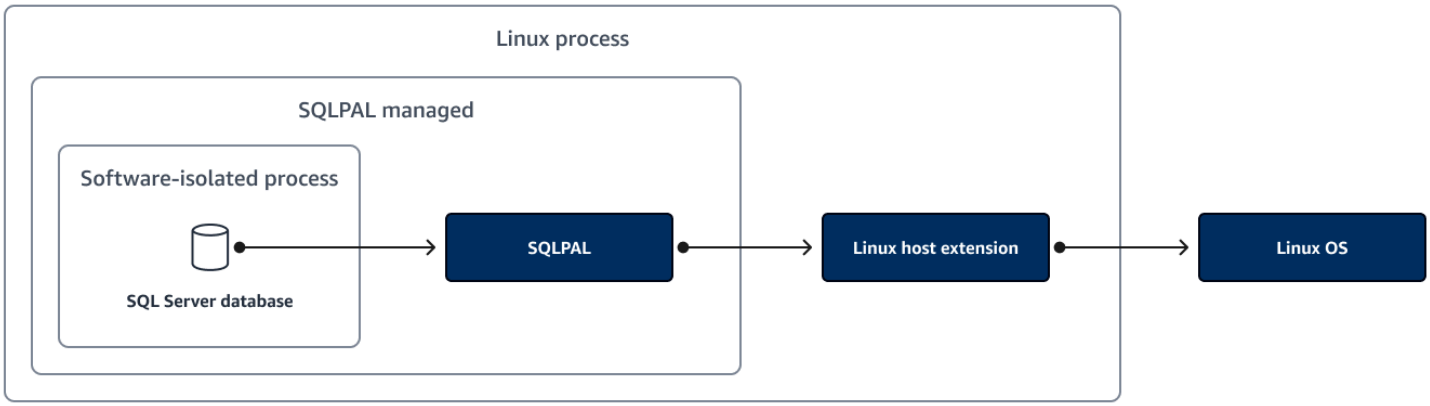
Evalúe SQL Server en Linux

Descripción general

Desde SQL Server 2017, es posible instalar SQL Server en sistemas operativos Linux. SQL Server para Linux está preparado para empresas y ofrece flexibilidad, alto rendimiento, funciones de seguridad, reducción del coste total de propiedad, funciones de alta disponibilidad y recuperación ante desastres y una excelente experiencia de usuario. Puede cambiar de SQL Server en Windows Server a SQL Server en Linux para ahorrar en los costes de licencia de Windows Server.

Para Linux, SQL Server está disponible para su implementación en Red Hat Enterprise Linux (RHEL), SUSE Linux Enterprise Server (SLES), Ubuntu y Amazon Linux 2. El motor de base de datos de SQL Server funciona de la misma manera tanto en Windows Server como en Linux, pero hay algunos cambios fundamentales en determinadas tareas cuando se utiliza Linux. Una diferencia clave entre la ejecución de aplicaciones de SQL Server Always On en Linux y Windows está relacionada con la agrupación en clústeres de conmutación por error. Si implementa grupos de disponibilidad Always On en un host de Windows Server, puede aprovechar los clústeres de conmutación por error de Windows Server (WSFC) y Active Directory como funciones integradas que admiten la agrupación en clústeres de conmutación por error. Sin embargo, ni WSFC ni Active Directory están disponibles para admitir la agrupación en clústeres de conmutación por error en Linux. Si desea lanzar clústeres de conmutación por error para SQL Server en Linux, puede utilizar Pacemaker AWS Launch Wizardpara simplificar la configuración del clúster y la instalación de SQL en instancias de Linux. ClusterLabs

SQL Server en Windows y Linux comparten una base de código común. Es decir, el motor principal de SQL Server no se ha modificado en absoluto para que funcione en Linux. SQL Server introdujo una capa de abstracción de plataforma (SQLPAL), como se muestra en el siguiente diagrama.



SQLPAL es responsable de la abstracción de las llamadas y la comunicación entre SQL Server y el sistema operativo subyacente. La extensión host es simplemente una aplicación nativa de Linux. Las funciones de bajo nivel del sistema operativo son llamadas nativas para optimizar la E/S, la memoria y el uso de la CPU. Cuando se inicia la extensión de host, carga e inicializa SQLPAL, que luego abre SQL Server. SQLPAL lanza procesos de software aislados que proporcionan la traducción necesaria para el resto del código. La adición de esta nueva capa a la arquitectura de SQL Server significa que las mismas funciones y beneficios principales de nivel empresarial que han hecho que SQL Server sea tan potente en Windows están disponibles independientemente del sistema operativo.

Impacto del costo

En el caso de las instancias r5.2 x large, la reducción del costo de las licencias de Windows Server es de unos 268$ en cada escenario. La reducción supone un porcentaje mayor del coste total del servidor en comparación con el uso de ediciones más baratas de SQL Server. En la siguiente tabla se muestran los ahorros de costos.

instancia	Edición	Coste mensual de SQL Server en Windows	Coste mensual de SQL Server en Linux	Ahorro
r5.2xlarge	Web	735 DÓLARES	466 DÓLARES	37%
r5.2xlarge	Estándar	1.337 DÓLARES	1.068 DÓLARES	20%
r5.2xlarge	Enterprise	2.826 DÓLARES	2.558 DÓLARES	10%

> **Note**
>
> La estimación de precios de la tabla anterior se basa en los precios bajo demanda de la `us-east-1` región y se puede consultar directamente en la. Calculadora de precios de AWS

Considere un ejemplo de escenario en el que un cliente de ISV del segmento de pequeñas y medianas empresas busca ahorrar costos en su entorno de desarrollo. Ya utilizan la edición SQL Server Developer en un conjunto de servidores Windows. Al cambiar de Windows con la edición SQL Server Developer a Linux con la edición SQL Server Developer, el cliente de ISV puede ahorrar un

33 por ciento en su carga de trabajo de desarrollo. En la siguiente tabla se muestran los siguientes costos estimados para este escenario.

Estimación	Costo mensual
Windows + SQL Server	9.307,72 DÓLARES
Linux + SQL Server	6.218,36 DÓLARES
Ahorro de costos estimado	3.089,36$ (33%)

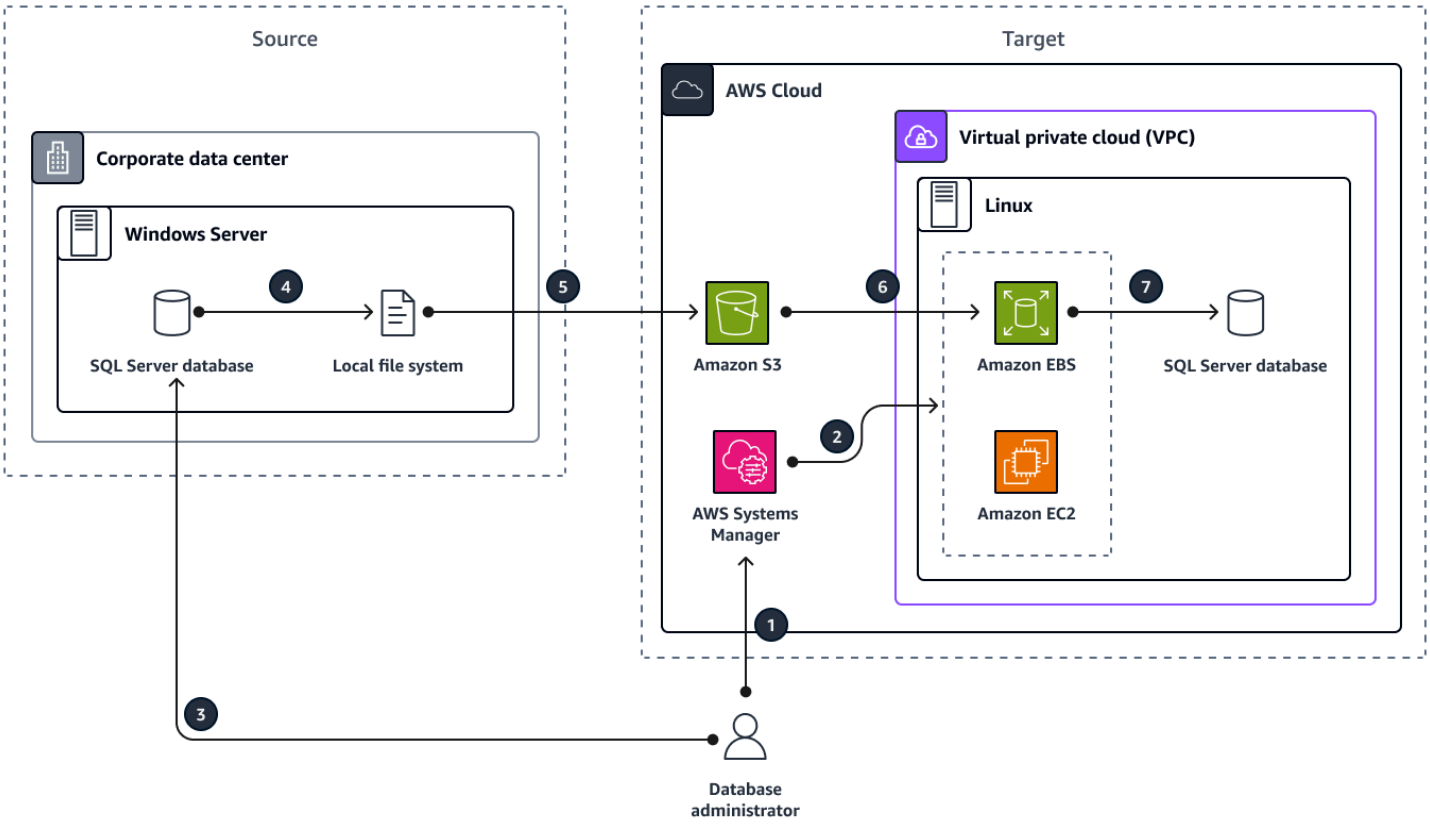
En otro escenario de ejemplo, una empresa migra instancias de SQL Server EC2 con licencia incluida de Windows a Linux. La empresa ahorra un total de 300 000 dólares al año en costos de licencias de Windows Server, aproximadamente el 20 por ciento de su factura total. AWS

Recomendaciones de optimización de costos

Le recomendamos que tenga en cuenta lo siguiente:

- A partir de SQL Server 2017, se admite SQL Server en Linux.
- Para facilitar el cambio, puede utilizar el asistente de cambio de plataforma de Windows a Linux para las bases de datos de Microsoft SQL Server. El asistente de cambio de plataforma es una herramienta de creación de scripts que puede ayudarle a mover las cargas de trabajo de SQL Server existentes de los sistemas operativos Windows a Linux comprobando las incompatibilidades habituales, exportando las bases de datos desde el host de Windows y, a continuación, importándolas a una instancia que EC2 ejecute Microsoft SQL Server 2017 en Ubuntu 16.04.
- También puede utilizar las funciones de copia de seguridad y restauración de SQL Server para cambiar de SQL Server en Windows a Linux.
- Puede realizar una implementación fácil y rápida en SQL Server en Linux o Ubuntu mediante el AWS Launch Wizard. El Launch Wizard puede implementar SQL Server en Linux o Ubuntu tanto en escenarios independientes como de alta disponibilidad en función de las necesidades de la aplicación. Para obtener más información, consulte la publicación Implementación en SQL Server Always on Linux AWS Launch Wizard with en el AWS blog Microsoft Workloads on.

El siguiente diagrama muestra la arquitectura de una solución que utiliza el asistente de cambio de plataforma de Windows a Linux para las bases de datos de Microsoft SQL Server.



Recursos adicionales

- Descripción general de SQL Server en Linux (Microsoft Learn)
- Guía de instalación de SQL Server en Linux (Microsoft Learn)
- Implementación en SQL Server Always en Linux con AWS Launch Wizard (Microsoft Workloads en el AWS blog)
- SQL Server de alta disponibilidad en Linux (blog de código AWS abierto)

Optimice las estrategias de respaldo de SQL Server

Descripción general

La mayoría de las organizaciones buscan la solución adecuada para proteger sus datos en SQL Server en Amazon EC2 a fin de cumplir con sus requisitos actuales de objetivo de punto de recuperación (RPO), el tiempo máximo aceptable desde la última copia de seguridad, y objetivo de tiempo de recuperación (RTO), el máximo retraso aceptable entre la interrupción del servicio y la restauración del servicio. Si ejecuta SQL Server en EC2 instancias, tiene varias opciones para crear

copias de seguridad de sus datos y restaurarlos. Las estrategias de Backup para proteger los datos de SQL Server en Amazon EC2 incluyen las siguientes:

- Copia de seguridad a nivel del servidor mediante instantáneas de Amazon Elastic Block Store (Amazon EBS) habilitadas para Windows Volume Shadow Copy Service (VSS) o AWS Backup
- Respaldo a nivel de base de datos mediante backup y restauración nativos en SQL Server

Dispone de las siguientes opciones de almacenamiento para las copias de seguridad nativas a nivel de base de datos:

- Una copia de seguridad local con un volumen de Amazon EBS
- Una copia de seguridad del sistema de archivos en red con Amazon FSx para Windows File Server o Amazon FSx para NetApp ONTAP
- Una copia de seguridad en red para Amazon Simple Storage Service (Amazon S3) mediante AWS Storage Gateway
- Respaldo directo a Amazon S3 para SQL Server 2022

En esta sección se hace lo siguiente:

- Destaca las funciones que le ayudan a ahorrar espacio de almacenamiento
- Compara los costos entre las diferentes opciones de almacenamiento de back-end
- Proporciona enlaces a documentación detallada para ayudar a implementar estas recomendaciones

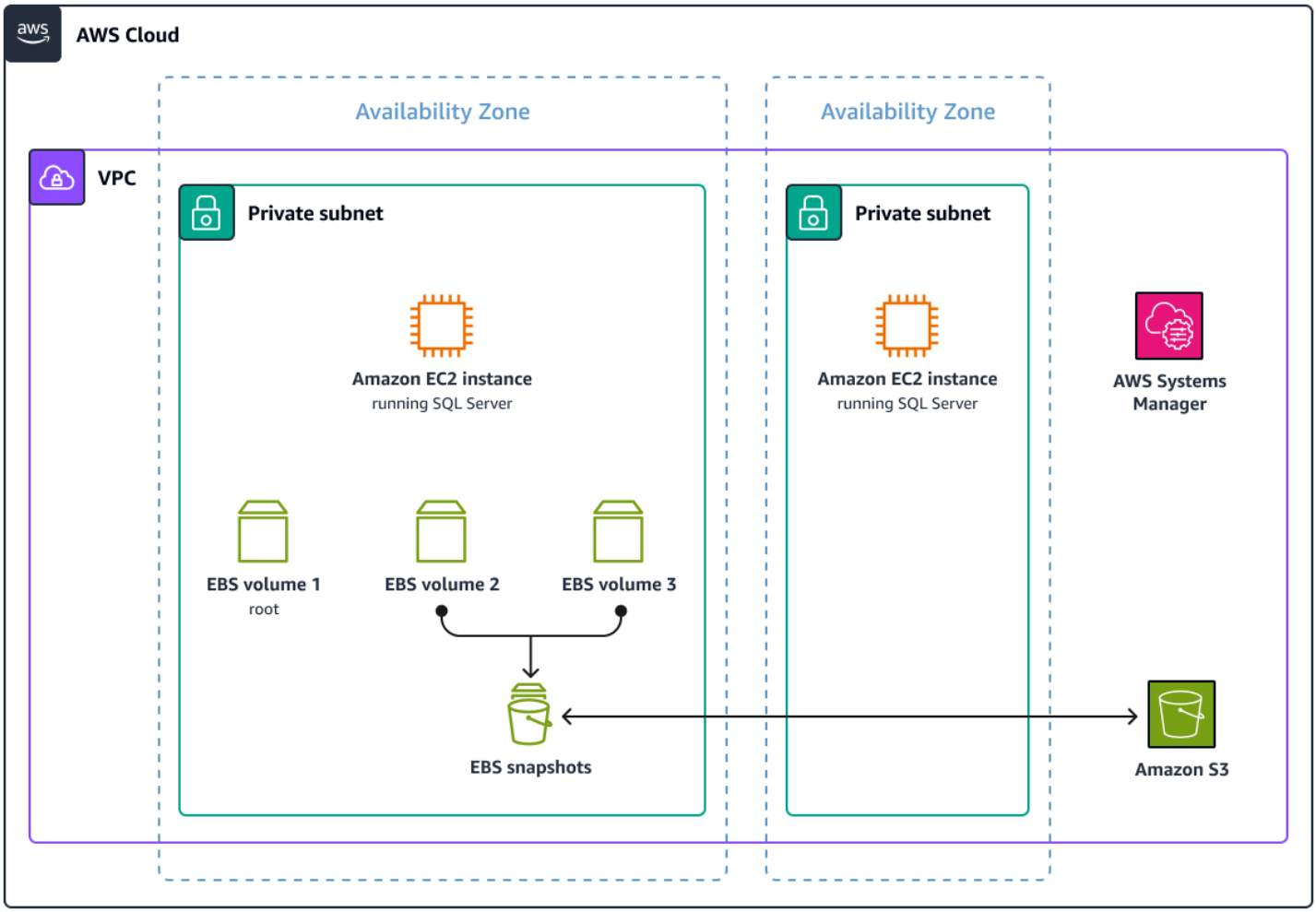
Respaldo a nivel de servidor mediante instantáneas habilitadas para VSS

Una arquitectura de instantáneas compatible con VSS utiliza el comando AWS Systems Manager Run para instalar el agente VSS en las instancias de SQL Server. También puede usar el comando Run para invocar todo el flujo de trabajo consistente en vaciar los búferes del sistema operativo y las aplicaciones en el disco, pausar las operaciones de E/S, tomar una point-in-time instantánea de los volúmenes de EBS y, a continuación, reanudar la E/S.

Run Command crea instantáneas automatizadas de todos los volúmenes de EBS conectados a una instancia de destino. También tiene la opción de excluir el volumen raíz, ya que los archivos de bases de datos de usuarios suelen almacenarse en otros volúmenes. Si divide varios volúmenes de EBS para crear un único sistema de archivos para los archivos de SQL Server, Amazon EBS

también admite instantáneas de varios volúmenes coherentes ante bloqueos mediante un único comando de API. Para obtener más información sobre las instantáneas de EBS compatibles con VSS compatibles con las aplicaciones, consulte Crear una instantánea coherente con las aplicaciones con VSS en la documentación de Amazon. EC2

El siguiente diagrama muestra una arquitectura para realizar copias de seguridad a nivel de servidor mediante instantáneas habilitadas para VSS.



Tenga en cuenta las siguientes ventajas del uso de instantáneas habilitadas para VSS:

- La primera instantánea de una instancia de base de datos contiene los datos de la instantánea de base de datos completa. Las instantáneas posteriores de la misma instancia de base de datos son graduales, lo que significa que solo se guardan los datos que han cambiado después de guardar la última instantánea.
- Las instantáneas de EBS proporcionan recuperación. point-in-time

- Puede restaurar en una nueva EC2 instancia de SQL Server a partir de una instantánea.
- Si una instancia se cifra con Amazon EBS o si se cifra una base de datos en la instancia con TDE, esa instancia o base de datos se restaura automáticamente con el mismo cifrado.
- Puede copiar copias de seguridad automatizadas entre regiones.
- Al restaurar un volumen de EBS a partir de una instantánea, queda inmediatamente disponible para que las aplicaciones accedan a él. Esto significa que puede poner SQL Server en línea de forma inmediata después de restaurar uno o más de sus volúmenes de EBS subyacentes a partir de instantáneas.
- De forma predeterminada, los volúmenes restaurados obtienen bloques subyacentes de Amazon S3 la primera vez que una aplicación intenta leerlos. Esto significa que puede haber un retraso en el rendimiento después de restaurar un volumen de EBS a partir de una instantánea. Con el tiempo, el volumen alcanza el rendimiento nominal. Sin embargo, puede evitar ese retraso mediante la restauración rápida de instantáneas (FSR).
- Puede utilizar la administración del ciclo de vida de las instantáneas de EBS.

Tenga en cuenta las siguientes limitaciones del uso de instantáneas compatibles con VSS:

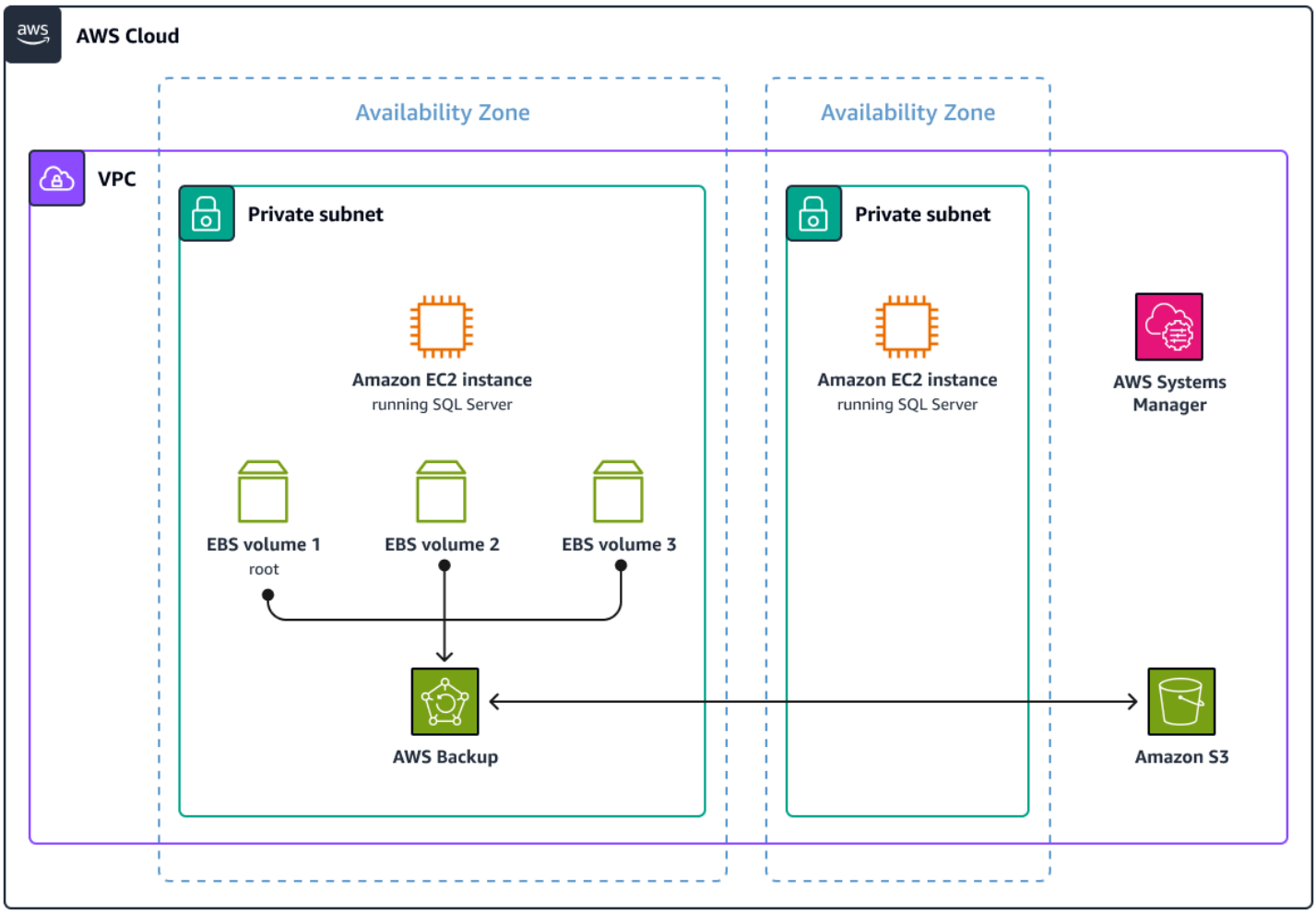
- No puede realizar una point-in-time recuperación entre regiones con una instantánea cifrada para una instancia de SQL Server.
- No puedes crear una instantánea cifrada de una instancia no cifrada.
- No puede restaurar una base de datos individual porque la instantánea se toma a nivel de volumen de EBS.
- No puede restaurar la instancia a sí misma.
- Una instantánea de la instancia de base de datos debe cifrarse con la misma clave AWS Key Management Service (AWS KMS) que la instancia de base de datos.
- La E/S del almacenamiento se suspende durante una fracción de segundo (aproximadamente 10 milisegundos) durante el proceso de copia de seguridad de la instantánea.

Realice una copia de seguridad de SQL Server mediante AWS Backup

Se puede utilizar AWS Backuppara centralizar y automatizar la protección de datos en todas Servicios de AWS partes. AWS Backup ofrece una solución rentable, totalmente gestionada y basada en políticas que simplifica la protección de datos a gran escala. AWS Backup también le ayuda a cumplir con sus obligaciones de conformidad normativa y a cumplir sus objetivos de

continuidad empresarial. Además AWS Organizations, le AWS Backup permite implementar de forma centralizada políticas de protección de datos (respaldo) para configurar, administrar y gobernar su actividad de respaldo en todos los recursos Cuentas de AWS y la organización.

El siguiente diagrama muestra la arquitectura de una solución de respaldo y restauración para SQL Server EC2 mediante el uso AWS Backup.



Tenga en cuenta las siguientes ventajas de realizar copias de seguridad de SQL Server mediante AWS Backup:

- Puede automatizar la programación de copias de seguridad, la administración de la retención y la administración del ciclo de vida.
- Puede centralizar su estrategia de respaldo en toda su organización, abarcando varias cuentas y. Regiones de AWS
- Puede centralizar la supervisión de la actividad de backup y las alertas en todas partes. Servicios de AWS

- Puede implementar copias de seguridad entre regiones para planificar la recuperación de desastres.
- La solución es compatible con las copias de seguridad entre cuentas.
- Puede realizar copias de seguridad seguras con un cifrado de copias de seguridad secundario.
- Todas las copias de seguridad admiten el cifrado mediante claves de AWS KMS cifrado.
- La solución funciona con TDE.
- Puede restaurar a un punto de recuperación específico desde la consola de AWS Backup .
- Puede hacer una copia de seguridad de una instancia completa de SQL Server, que incluye todas las bases de datos de SQL Server.

Copia de seguridad a nivel de la base de datos

Estos enfoques utilizan la funcionalidad de copia de seguridad nativa de Microsoft SQL Server. Puede realizar copias de seguridad de bases de datos individuales en la instancia de SQL Server y restaurar una base de datos individual.

Cada una de estas opciones de copia de seguridad y restauración nativas de SQL Server también admite lo siguiente:

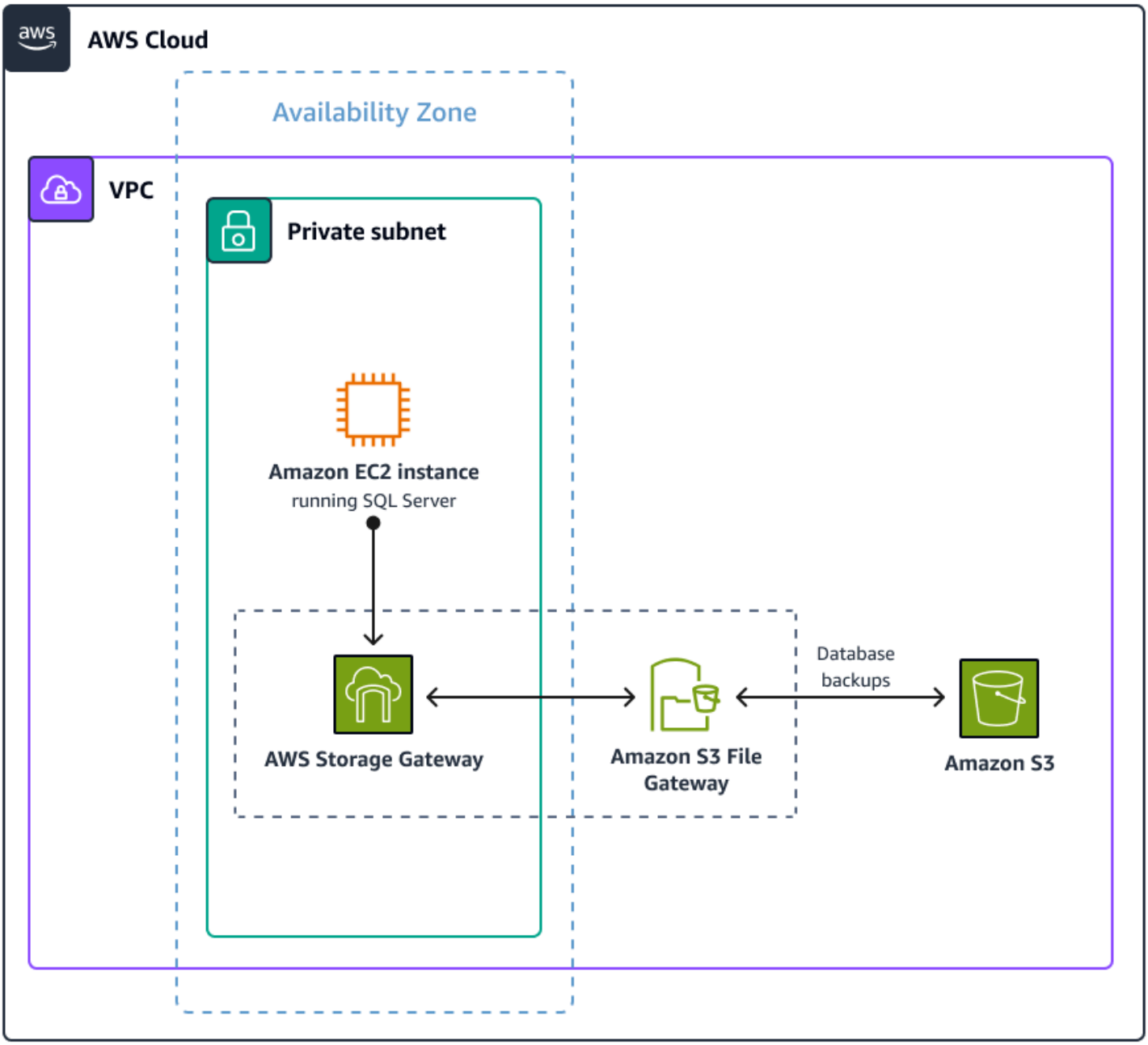
- Compresión y copia de seguridad de varios archivos
- Copias de seguridad completas, diferenciales y de registros T
- Bases de datos cifradas con TDE

Respaldo y restauración nativos de SQL Server en Amazon S3

SQL Server en Amazon EC2 admite la copia de seguridad y la restauración nativas de las bases de datos de SQL Server. Puede realizar una copia de seguridad de la base de datos de SQL Server y, a continuación, restaurar el archivo de copia de seguridad en una base de datos existente o en una nueva EC2 instancia de SQL Server, Amazon RDS for SQL Server o un servidor local.

Storage Gateway es un servicio de almacenamiento de nube híbrida que brinda acceso a las aplicaciones en las instalaciones a un almacenamiento en la nube prácticamente ilimitado. Puede usar Storage Gateway para hacer copias de seguridad de sus bases de datos de Microsoft SQL Server directamente en Amazon S3, lo que reduce el espacio de almacenamiento local y utiliza Amazon S3 para un almacenamiento duradero, escalable y rentable.

El siguiente diagrama muestra la arquitectura de una solución nativa de respaldo y restauración que usa Storage Gateway y Amazon S3.



Tenga en cuenta las siguientes ventajas de utilizar copias de seguridad nativas de SQL Server con Storage Gateway:

- Puede asignar una pasarela de almacenamiento como un recurso compartido de archivos del bloque de mensajes del servidor (SMB) en la EC2 instancia y enviar la copia de seguridad a Amazon S3.

- La copia de seguridad va directamente al bucket de S3 o a través de la caché de archivos de Storage Gateway.
- Se admiten copias de seguridad de varios archivos.

Tenga en cuenta las siguientes limitaciones del backup nativo mediante Storage Gateway:

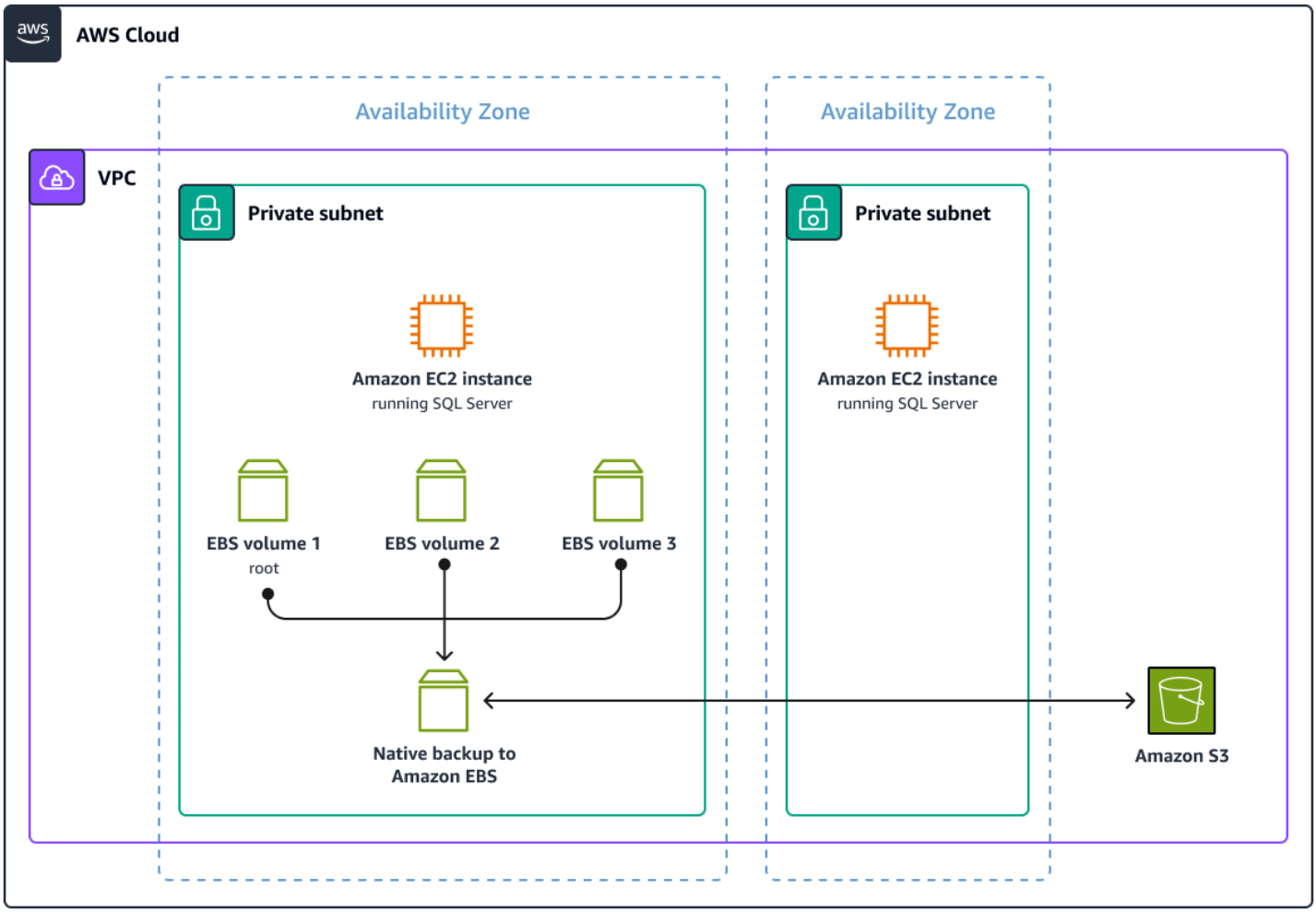
- Debe configurar la copia de seguridad y la restauración para cada base de datos individual.
- Debe administrar la Política de ciclo de vida de Amazon S3 para los archivos de copia de seguridad.

Para obtener más información sobre cómo configurar Storage Gateway, consulte la AWS Storage Gateway publicación del AWS blog sobre cómo almacenar copias de seguridad de SQL Server en Amazon S3.

Copias de seguridad nativas de SQL Server en volúmenes de EBS

Puede realizar una copia de seguridad nativa de la base de datos de SQL Server y almacenar el archivo en un volumen de Amazon EBS. Amazon EBS es un servicio de almacenamiento en bloques de alto rendimiento. Los volúmenes de EBS son elásticos, lo que permite el cifrado. Se pueden separar y adjuntar a una EC2 instancia. Puede hacer una copia de seguridad de SQL Server en una EC2 instancia del mismo tipo de volumen de EBS o en un tipo de volumen de EBS diferente. Una ventaja de realizar copias de seguridad en un volumen de EBS diferente es el ahorro de costos.

En el siguiente diagrama, se muestra la arquitectura de una copia de seguridad nativa en un volumen de EBS.



Tenga en cuenta las siguientes ventajas de utilizar el backup nativo de SQL Server en volúmenes de EBS:

- Puede realizar copias de seguridad de bases de datos individuales en una EC2 instancia de SQL Server y restaurar una base de datos individual en lugar de tener que restaurar la instancia completa.
- Se admiten copias de seguridad de varios archivos.
- Puede programar trabajos de copia de seguridad mediante el agente de SQL Server y el motor de trabajos de SQL Server.
- Puede obtener beneficios de rendimiento a través de las opciones de equipos. Por ejemplo, puede utilizar los volúmenes de almacenamiento st1 para lograr un mayor rendimiento.

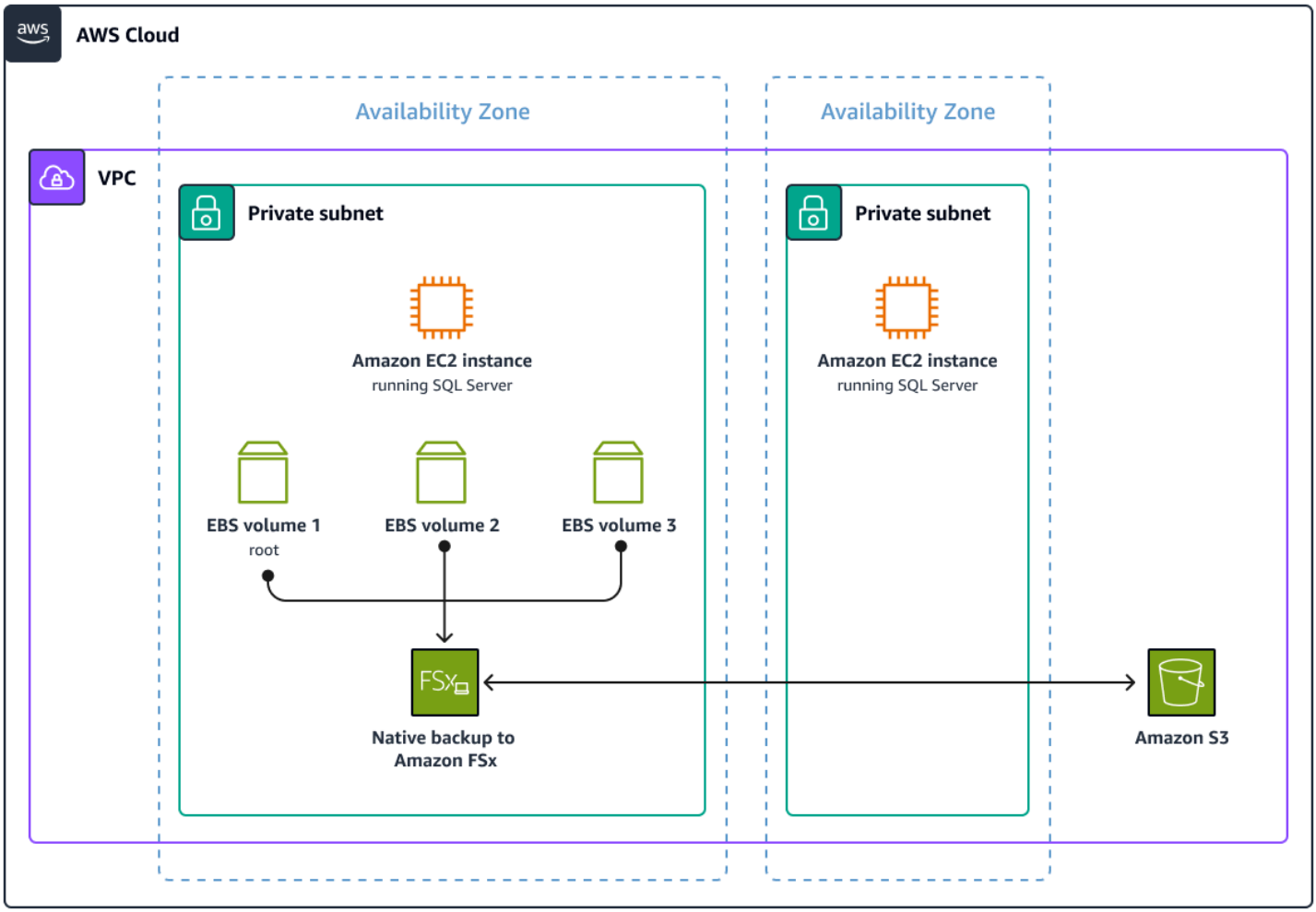
Tenga en cuenta las siguientes limitaciones del uso de copias de seguridad nativas en los volúmenes de EBS:

- Debe mover de forma manual las copias de seguridad a Amazon S3 desde el volumen de EBS.
- Para copias de seguridad de gran tamaño, debes administrar el espacio en disco en Amazon EC2.
- En este EC2 caso, el rendimiento de Amazon EBS puede ser un cuello de botella.
- Se requiere almacenamiento adicional para almacenar copias de seguridad en Amazon EBS.

Backup nativo de SQL Server en Amazon FSx para Windows File Server

Amazon FSx for Windows File Server es un sistema de archivos de Windows nativo totalmente gestionado que ofrece hasta 64 TB de almacenamiento y está diseñado para ofrecer un rendimiento rápido, predecible y uniforme. AWS introdujo el soporte nativo para las implementaciones de sistemas de archivos Multi-AZ en FSx Windows File Server. El soporte nativo facilita la implementación del almacenamiento de archivos de Windows AWS con alta disponibilidad y redundancia en múltiples zonas de disponibilidad. AWS también introdujo la compatibilidad con los recursos compartidos de archivos SMB de disponibilidad continua (CA). Puede utilizar el FSx servidor de archivos de Windows como almacenamiento de respaldo para una base de datos de SQL Server.

El siguiente diagrama muestra la arquitectura de una copia de seguridad nativa de SQL Server FSx para Windows File Server.



Tenga en cuenta las siguientes ventajas de utilizar copias de seguridad nativas de SQL Server FSx para Windows File Server:

- Puede hacer una copia de seguridad de la base de datos de SQL Server en un recurso compartido de FSx archivos de Amazon.
- Puede realizar copias de seguridad de bases de datos individuales en una instancia de SQL Server y restaurar una base de datos individual, en lugar de tener que restaurar la instancia completa.
- Se admiten copias de seguridad de varias partes.
- Puede programar trabajos de copia de seguridad mediante el agente de SQL Server y el motor de trabajos.
- Las instancias tienen un ancho de banda de red superior en comparación con Amazon EBS.

Tenga en cuenta las siguientes limitaciones a la hora de utilizar copias de seguridad nativas de SQL Server FSx para Windows File Server:

- Debe mover manualmente las copias de seguridad a Amazon S3 desde Amazon FSx mediante AWS Backup o AWS DataSync.
- Las copias de seguridad de gran tamaño pueden requerir una sobrecarga adicional para la administración del espacio en disco en Amazon FSx.
- EC2 el rendimiento de la red de instancias puede ser un cuello de botella.
- Se necesita almacenamiento adicional para almacenar las copias de seguridad en el servidor de archivos FSx de Windows.

Copia de seguridad de SQL Server en Amazon FSx para NetApp ONTAP

FSx Para ONTAP, las instantáneas son siempre consistentes con los fallos, pero es necesario poner en modo inactivo (o pausar la E/S) de la base de datos para poder crear una instantánea coherente con la aplicación. Puede utilizar ONTAP NetApp SnapCenter (una herramienta de organización con complementos para aplicaciones específicas, como SQL Server) para crear instantáneas coherentes con las aplicaciones y FSx proteger, replicar y clonar sus bases de datos sin coste adicional.

NetApp SnapCenter

NetApp SnapCenter es una plataforma unificada para una protección de datos coherente con las aplicaciones. SnapCenter se refiere a las instantáneas como copias de seguridad. Esta guía adopta la misma convención de nomenclatura. SnapCenter proporciona un panel único para administrar copias de seguridad, restauraciones y clones coherentes con las aplicaciones. Agregue un SnapCenter complemento para su aplicación de base de datos específica para crear copias de seguridad coherentes con las aplicaciones. El SnapCenter complemento para SQL Server proporciona las siguientes funciones que simplifican el flujo de trabajo de protección de datos.

- Opciones de backup y restauración con granularidad para copias de seguridad completas y de registros
- Restauración in situ y restauración en una ubicación alternativa

Para obtener más información al respecto SnapCenter, consulte la publicación Proteja sus cargas de trabajo de SQL Server NetApp SnapCenter con Amazon FSx for NetApp ONTAP en el blog sobre AWS almacenamiento.

Optimización de costes para las copias de seguridad

Las siguientes opciones pueden ayudarle a reducir el costo de almacenar las copias de seguridad de SQL Server en ellas AWS.

- Active la compresión de SQL Server durante la creación del archivo de copia de seguridad y envíe el archivo más pequeño posible al almacenamiento. Por ejemplo, una relación de compresión de 3:1 indica que está ahorrando alrededor del 66 por ciento de espacio en disco. Para realizar consultas en estas columnas, puede usar la siguiente instrucción de Transact-SQL: `SELECT backup_size/compressed_backup_size FROM msdb..backupset;`
- Para las copias de seguridad que se destinan a depósitos S3, habilite la clase de almacenamiento Amazon S3 Intelligent-Tiering para reducir los costos de almacenamiento en un 30 por ciento.
- En el caso de las copias de seguridad FSx destinadas a Windows File Server o FSx a ONTAP, utilice una única zona de disponibilidad para ahorrar un 50 por ciento en los costes (en comparación con el uso de varias zonas de disponibilidad). Para obtener información sobre precios, consulte los precios de Amazon FSx para Windows File Server y Amazon FSx para NetApp ONTAP.
- La opción más eficaz para SQL Server 2022 es la copia de seguridad directa en Amazon S3. Puede ahorrar costes adicionales si evita Storage Gateway.

Resultados de pruebas comparativas para copias de seguridad

En esta sección, se comparan las siguientes opciones desde el punto de vista del costo y el rendimiento de un ejemplo de base de datos de 1 TB, según los resultados de las pruebas comparativas de rendimiento de las soluciones de respaldo incluidas en esta guía.

- EC2 especificación de instancia: r5d.8xlarge con Windows Server 2019 y SQL Server 2019 Developer Edition
- Especificación de base de datos: 1 TB de tamaño con el TDE desactivado

Las pruebas se realizaron con una instancia r5d.8xlarge y una base de datos de SQL Server de 1 TB como fuente. El sistema de origen se configuró de acuerdo con las mejores prácticas y la base de datos de origen contenía cuatro archivos de datos (250 GB cada uno) y un archivo de registro (50 GB) repartidos en volúmenes gp3 independientes. El `BACKUP` comando nativo de SQL Server incluye escribir en 10 archivos de copia de seguridad y utilizar la compresión para optimizar el rendimiento

de la copia de seguridad y reducir la cantidad de datos que se envían a través de la red y se escriben en el destino. En todos los casos de prueba, el rendimiento del almacenamiento fue el obstáculo.

Existe una variedad casi infinita de configuraciones posibles para este tipo de pruebas. Esta prueba se centró en optimizar el rendimiento, el costo, la escalabilidad y los casos de uso reales. En la siguiente tabla se muestran las métricas de rendimiento que se capturaron para las opciones de backup objetivo.

Opciones de copia de seguridad	Nivel	Duración de la ejecución (aproxima damente)	Velocidad de la copia de seguridad	Coste en USD al mes*
Respaldo nativo en un disco duro EBS st1 local de 2 TB	Base de datos	00:30:46 min	554,7 Mbps	92,16 DÓLARES
Copia de seguridad nativa en el SSD gp3 de EBS local, de 2 TB	Base de datos	00:22:00 min	512 Mbps	193,84 DÓLARES
Respaldo nativo FSx para disco duro Windows File Server, 2 TB @512 Mbps de rendimiento	Base de datos	00:20:58 min	814,0 Mbps	1.146 DÓLARES
Respaldo nativo FSx para SSD Windows File Server, 2 TB @512 Mbps de rendimiento	Base de datos	00:20:00 min	814,0 Mbps	1.326 DÓLARES

Opciones de copia de seguridad	Nivel	Duración de la ejecución (aproxima damente)	Velocidad de la copia de seguridad	Coste en USD al mes*
Respaldo nativo en S3 File Gateway m6i.4xlarge (16 vCPU, 64 GB) con gp3 de 2 TB	Base de datos	00:23:20 min	731,5 Mbps	470,42 DÓLARES
Instantánea de EBS VSS	Volumen de EBS	00:00:02 segundos 00:00:53 segundos	N/A instantánea	51$
AWS Backup (Respaldo de AMI)	AMI	00:00:04 segundos 00:08:00 min	N/A instantánea	75 DÓLARES
Respaldo nativo de SQL Server directamente en Amazon S3 (SQL Server 2022)	Base de datos	00:12:00 min	731,5 Mbps	Los primeros 50 TB al mes, 0,023$ por GB 23,55$ al mes
Respaldo nativo para ONTAP (FSx mediante) SnapCenter	Base de datos	–	–	440,20$

En la tabla anterior se presupone lo siguiente:

- Los costes de transferencia de datos y Amazon S3 no están incluidos.

- El precio del almacenamiento está incluido en el precio de la instancia.
- Los costes se calculan en función de la `us-east-1` región.
- El rendimiento y las IOPS aumentan un 10 por ciento, y las copias de seguridad múltiples tienen una tasa general de cambio del 10 por ciento a lo largo del mes.

Los resultados de las pruebas muestran que la opción más rápida es realizar una copia de seguridad nativa de la base de datos de SQL Server FSx para Windows File Server. Una copia de seguridad en Storage Gateway y en volúmenes EBS conectados localmente es la opción más rentable, pero tiene un rendimiento más lento. Para los respaldos a nivel de servidor (AMI), recomendamos utilizarlos AWS Backup para obtener un rendimiento, un costo y una capacidad de administración óptimos.

Recomendaciones de optimización de costos

Comprender las posibles soluciones para realizar copias de seguridad de SQL Server en Amazon EC2 es clave para proteger sus datos, garantizar que cumple con sus necesidades de copia de seguridad y establecer un plan para recuperarse de eventos críticos. Las diferentes formas de realizar copias de seguridad y restaurar las instancias y bases de datos de SQL Server que se analizan en esta sección pueden ayudarle a diseñar una estrategia de copia de seguridad y restauración que proteja sus datos y cumpla con los requisitos de su organización.

En esta sección se describen las siguientes opciones de copia de seguridad:

- Compresión
- Amazon S3 Intelligent-Tiering
- Una sola zona de disponibilidad
- Backup en URL

La orientación proporcionada para cada una de estas opciones es de alto nivel. Si deseas implementar alguna de estas recomendaciones en tu organización, te recomendamos que te pongas en contacto con tu equipo de cuentas. A continuación, el equipo puede ponerse en contacto con un especialista de Microsoft SA para que dirija la conversación. También puedes ponerte en contacto con nosotros enviando un correo electrónico a optimize-microsoft@amazon.com.

En resumen, te recomendamos lo siguiente:

- Si utiliza SQL Server 2022, la opción más rentable es realizar copias de seguridad en Amazon S3.

- Si utiliza SQL Server 2019 y ediciones anteriores de SQL Server, considere la posibilidad de realizar copias de seguridad en Storage Gateway respaldado por Amazon S3 como la opción más rentable.

Compresión

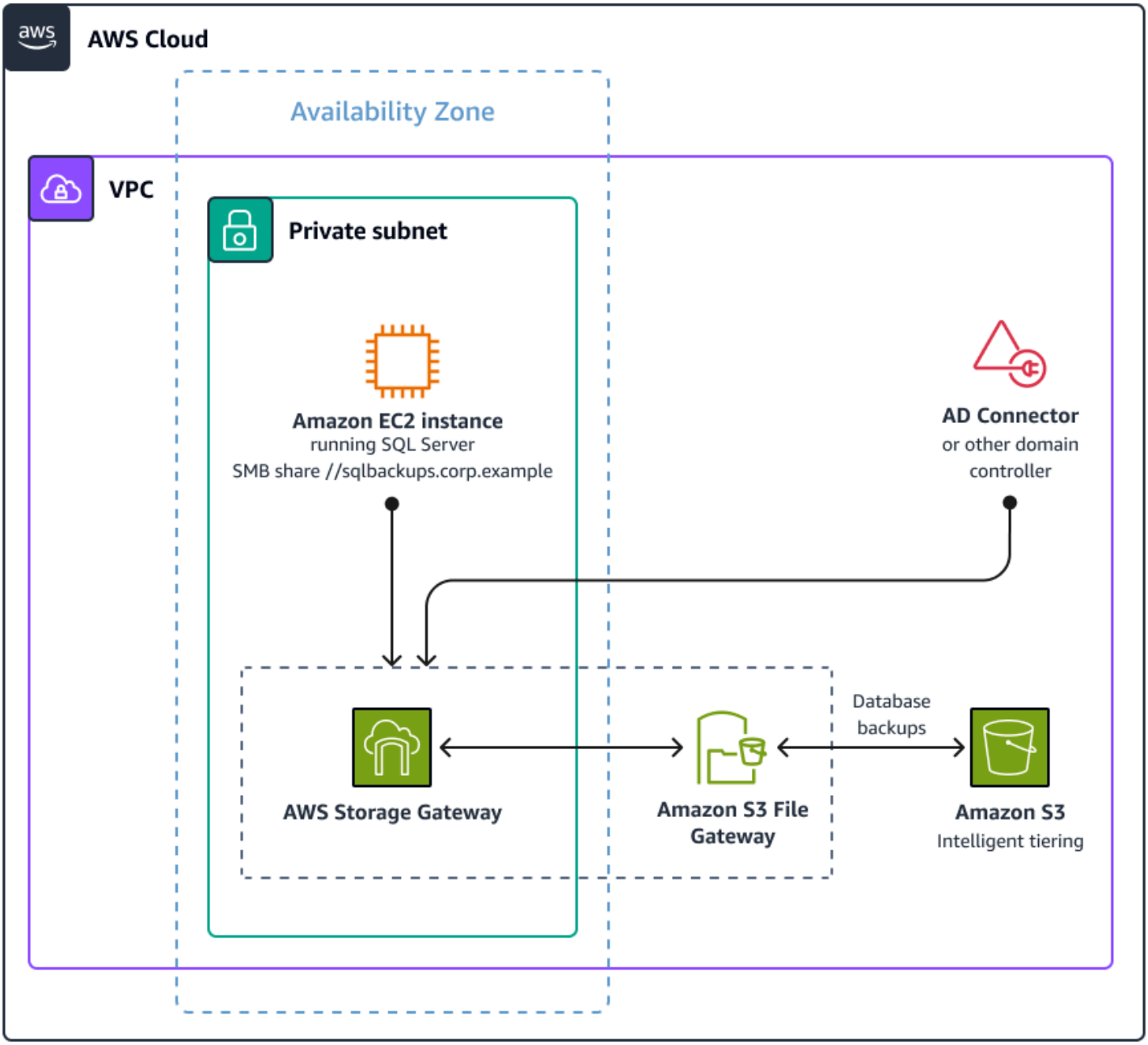
El objetivo de la compresión es que cada copia de seguridad consuma menos espacio de almacenamiento, lo que resulta beneficioso para varias opciones de almacenamiento. Debe habilitar la compresión de una copia de seguridad de SQL Server en el nivel de la instancia de SQL Server. El siguiente ejemplo muestra cómo agregar la palabra clave de compresión a una base de datos de respaldo:

```
BACKUP DATABASE <database_name> TO DISK WITH COMPRESSION (ALGORITHM =
QAT_DEFLATE)
```

Amazon S3 Intelligent-Tiering

Para las copias de seguridad destinadas a depósitos de Amazon S3, puede habilitar Amazon S3 Intelligent-Tiering como su clase de almacenamiento de Amazon S3 File Gateway. Esto puede reducir los costos de almacenamiento hasta en un 30 por ciento. A continuación, instale S3 File Gateway en sus servidores SQL mediante un recurso compartido de archivos SMB que se pueda integrar con su dominio de Active Directory. Esto le proporciona el control de acceso para lo que comparte, la posibilidad de aprovechar las cuentas de servicio existentes y el acceso a Amazon S3 mediante un protocolo de archivos común centrado en Microsoft. En el caso de las cuentas que quizás no tengan conectividad directa con un controlador de dominio, puede utilizar el conector de Active Directory para facilitar la comunicación con Active Directory de forma local o en la nube. Para configurar los ajustes de Active Directory en la puerta de enlace, debe especificar el conector de Active Directory IPs para que el controlador de dominio envíe las solicitudes por proxy a Active Directory.

El siguiente diagrama muestra la arquitectura de una solución basada en S3 Intelligent-Tiering.



De forma predeterminada, los archivos de respaldo escritos en el bucket de S3 utilizan el nivel estándar. Para convertir los archivos de respaldo del nivel estándar al nivel S3 Intelligent-Tiering, debe crear una regla de ciclo de vida. También puede utilizarla para activar S3 AWS Management ConsoleIntelligent-Tiering. Para obtener más información, consulte Cómo empezar a utilizar Amazon S3 Intelligent-Tiering en la documentación. AWS

Una sola zona de disponibilidad

Para crear un sistema de archivos de zona de disponibilidad única, elija la opción Single-AZ al crear un sistema de archivos FSx para Windows File Server. Amazon FSx también realiza copias de seguridad muy duraderas (almacenadas en Amazon S3) de su sistema de archivos a diario mediante el servicio Windows Volume Shadow Copy y le permite realizar copias de seguridad adicionales en cualquier momento. Tenga en cuenta algunos de los problemas que conlleva el uso de una zona de disponibilidad única. Por ejemplo, el recurso compartido de archivos SMB deja de estar accesible si una zona de disponibilidad afectada en la que se aprovisiona el sistema de archivos deja de funcionar durante horas seguidas. Si necesita acceder a los datos, debe restaurarlos a partir de las copias de seguridad de una zona de disponibilidad disponible dentro de la región de origen. Para obtener más información, consulte la sección Usar una única zona de disponibilidad de esta guía.

Backup en URL

Para SQL Server 2022, la función de copia de seguridad en URL permite realizar copias de seguridad directas en Amazon S3. Este es el enfoque de respaldo ideal para ejecutar SQL Server 2022, AWS ya que se obtiene el conjunto completo de funciones de Amazon S3 en la capa de almacenamiento y se elimina el costo del AWS Storage Gateway dispositivo necesario en las versiones anteriores para facilitar esta funcionalidad. Al implementar esta función, se deben tener en cuenta dos costos principales: los costos de transferencia de datos y la clase de almacenamiento S3 elegida. Si desea las capacidades de recuperación ante desastres nativas de Amazon S3, debe tener en cuenta que la replicación entre regiones genera costos de salida de datos entre regiones. Para obtener más información sobre cómo configurar esta opción, consulte la publicación Backup SQL Server databases to Amazon S3 en el AWS blog Microsoft Workloads on.

Recursos adicionales

- Opciones de backup y restauración para SQL Server en Amazon EC2 (guía AWS prescriptiva)
- Point-in-time recuperación y copia de seguridad continua para Amazon RDS con AWS Backup (AWS Storage Blog)
- Proteja sus cargas de trabajo de SQL Server NetApp SnapCenter con Amazon FSx for NetApp ONTAP (blog sobreAWS almacenamiento)
- Introducción al uso de Amazon S3 Intelligent-Tiering (Centro de recursos de AWS introducción)
- Estrategias de backup y restauración de Amazon RDS for SQL ServerAWS (blog sobre bases de datos)

- Migrar una base de datos local de Microsoft SQL Server a Amazon EC2 (orientación AWS prescriptiva)
- Prácticas recomendadas para implementar Microsoft SQL Server en Amazon EC2 (AWS documento técnico)

Modernice las bases de datos de SQL Server

Descripción general

Si está empezando a modernizar las bases de datos antiguas para mejorar la escalabilidad, el rendimiento y la optimización de costos, es posible que se enfrente a desafíos con bases de datos comerciales como SQL Server. Las bases de datos comerciales son caras, bloquean a los clientes y ofrecen condiciones de licencia punitivas. En esta sección se proporciona una descripción general de alto nivel de las opciones para migrar y modernizar bases de datos de SQL Server a bases de datos de código abierto, así como información sobre cómo elegir la mejor opción para su carga de trabajo.

Puede refactorizar sus bases de datos de SQL Server en bases de datos de código abierto, como Amazon Aurora PostgreSQL, para ahorrar en los costes de licencias de Windows y SQL Server. Las bases de datos modernas nativas de la nube, como Aurora, combinan la flexibilidad y el bajo costo de las bases de datos de código abierto con las sólidas funciones de nivel empresarial de las bases de datos comerciales. Si tiene cargas de trabajo variables o cargas de trabajo de varios inquilinos, también puede migrar a Aurora Serverless V2. Esto puede reducir los costos hasta un 90 por ciento, según las características de la carga de trabajo. Además, AWS ofrece capacidades como Babelfish para Aurora PostgreSQL, herramientas como AWS Schema Conversion Tool (AWS SCT) y servicios como AWS Database Migration Service (AWS DMS) para simplificar la migración y la modernización de las bases de datos de SQL Server en. AWS

Ofertas de bases de datos

La migración de SQL Server en Windows a bases de datos de código abierto como Amazon Aurora, Amazon RDS for MySQL o Amazon RDS for PostgreSQL puede ofrecer importantes ahorros de costes sin comprometer el rendimiento ni las funciones. Considere lo siguiente:

- El cambio de la edición SQL Server Enterprise en Amazon EC2 a Amazon RDS for PostgreSQL o Amazon RDS for MySQL puede suponer un ahorro de costes de hasta un 80 por ciento.

- El cambio de la edición Enterprise de SQL Server en Amazon EC2 a la edición compatible con PostgreSQL de Amazon Aurora o la edición compatible con Amazon Aurora MySQL puede suponer un ahorro de costes de hasta un 70 por ciento.

Para las cargas de trabajo de bases de datos tradicionales, Amazon RDS for PostgreSQL y Amazon RDS for MySQL abordan los requisitos y proporcionan una solución rentable para las bases de datos relacionales. Aurora agrega numerosas funciones de disponibilidad y rendimiento que antes estaban limitadas a proveedores comerciales caros. Las funciones de resiliencia de Aurora tienen un costo adicional. Sin embargo, en comparación con funciones similares de otros proveedores comerciales, los costos de resiliencia de Aurora siguen siendo más baratos que los que cobra el software comercial por el mismo tipo de funciones. La arquitectura Aurora está optimizada para ofrecer mejoras significativas en el rendimiento en comparación con las implementaciones estándar de MySQL y PostgreSQL.

Como Aurora es compatible con bases de datos PostgreSQL y MySQL de código abierto, existe la ventaja adicional de la portabilidad. Si la mejor opción es Amazon RDS for PostgreSQL, Amazon RDS for MySQL o Aurora, hay que entender los requisitos empresariales y asignar las características necesarias a la mejor opción.

Comparación entre Amazon RDS y Aurora

En la siguiente tabla se resumen las principales diferencias entre Amazon RDS y Amazon Aurora.

Categoría	Amazon RDS para PostgreSQL o Amazon RDS para MySQL	Aurora PostgreSQL o Aurora MySQL
Rendimiento	Buen rendimiento	Rendimiento 3 veces o mejor
Conmutación por error	Normalmente de 60 a 120 segundos*	Normalmente 30 segundos
Escalabilidad	Réplica de hasta 5 lecturas Retraso en segundos	Hasta 15 réplicas de lectura Retraso en milisegundos
Almacenamiento	Hasta 64 TB	Hasta 128 TB

Categoría	Amazon RDS para PostgreSQL o Amazon RDS para MySQL	Aurora PostgreSQL o Aurora MySQL
HA de almacenamiento	Multi-AZ con uno o dos en espera, cada uno con una copia de la base de datos	De forma predeterminada, 6 copias de datos en 3 zonas de disponibilidad
Copia de seguridad	Copias de seguridad diarias de instantáneas y registros	Respaldo continuo y asíncrono en Amazon S3
Innovaciones con Aurora	N/D	100 GB Clonación rápida de bases de datos
	Réplicas de lectura con escalado automático	
	Administración de planes de consultas	
	Aurora Serverless	
	Réplicas entre regiones con base de datos global	
	Administración de caché de clúster**	
	Consulta paralela	
	Secuencias de actividades de la base de datos	

*Las transacciones grandes pueden aumentar los tiempos de conmutación por error

**Disponible en Aurora PostgreSQL

La siguiente tabla muestra el costo mensual estimado de los diferentes servicios de bases de datos que se tratan en esta sección.

Servicio de base de datos	Coste en USD al mes*	Calculadora de precios de AWS (requiere Cuenta de AWS)
Edición Amazon RDS para SQL Server Enterprise	3.750 DÓLARES	Estimación
Amazon RDS para SQL Server edición estándar	2.318$	Estimación
Edición SQL Server Enterprise en Amazon EC2	2.835 DÓLARES	Estimación
Edición estándar de SQL Server en Amazon EC2	1.345$	Estimación
Amazon RDS para PostgreSQL	742 DÓLARES	Estimación
Amazon RDS para MySQL	712 DÓLARES	Estimación
Aurora PostgreSQL	1.032 DÓLARES	Estimación
Aurora MySQL	1.031 DÓLARES	Estimación

* El precio del almacenamiento está incluido en el precio de la instancia. Los costos se basan en la `us-east-1` región. El rendimiento y las IOPS son suposiciones. Los cálculos son para las instancias r6i.2xlarge y r6g.2xlarge.

Recomendaciones de optimización de costos

Las migraciones de bases de datos heterogéneas suelen requerir convertir el esquema de la base de datos del motor de base de datos de origen al motor de base de datos de destino y migrar los datos de la base de datos de origen a la base de datos de destino. El primer paso hacia la migración es evaluar y convertir el esquema del servidor SQL y los objetos de código en el motor de base de datos de destino.

Puede utilizar el AWS Schema Conversion Tool (AWS SCT) para evaluar y evaluar la compatibilidad de la base de datos con varias opciones de bases de datos de código abierto de destino, como

Amazon RDS for MySQL o Amazon RDS for PostgreSQL, Aurora MySQL y PostgreSQL. También puede utilizar la herramienta Babelfish Compass para evaluar la compatibilidad con Babelfish para Aurora PostgreSQL. Esto convierte a Compass AWS SCT y a Compass en herramientas poderosas para comprender el trabajo inicial que implica antes de decidir una estrategia de migración. Si decide continuar, AWS SCT automatiza los cambios necesarios en el esquema. La filosofía central de Babelfish Compass es permitir que la base de datos SQL se traslade a Aurora sin modificaciones o con muy pocas modificaciones. Compass evaluará la base de datos SQL existente para determinar si esto se puede lograr. De esta forma, se conoce el resultado antes de que se dedique ningún esfuerzo a migrar datos de SQL Server a Aurora.

AWS SCT automatiza la conversión y migración del esquema y el código de la base de datos al motor de base de datos de destino. Puede utilizar Babelfish para Aurora PostgreSQL para migrar la base de datos y la aplicación de SQL Server a Aurora PostgreSQL con cambios mínimos o nulos en el esquema. Esto puede acelerar sus migraciones.

Una vez migrado el esquema, puede utilizarlo AWS DMS para migrar los datos. AWS DMS puede realizar una carga completa de datos y replicar los cambios para realizar la migración con un tiempo de inactividad mínimo.

En esta sección se analizan las siguientes herramientas con más detalle:

- AWS Schema Conversion Tool
- Babelfish para Aurora PostgreSQL
- Brújula Babelfish
- AWS Database Migration Service

AWS Schema Conversion Tool

Puede utilizarlas AWS SCT para evaluar sus bases de datos de SQL Server existentes y evaluar la compatibilidad con Amazon RDS o Aurora. Para simplificar el proceso de migración, también puede convertir el esquema de un motor de base de datos a otro en una migración de base de datos heterogénea. AWS SCT Se puede utilizar AWS SCT para evaluar la aplicación y convertir el código de la aplicación incrustado para aplicaciones escritas en C#, C++, Java y otros lenguajes. Para obtener más información, consulte Convertir el código SQL de una aplicación mediante AWS SCT la AWS SCT documentación.

AWS SCT es una AWS herramienta gratuita que admite muchas fuentes de bases de datos. Para utilizarla AWS SCT, hay que dirigirla a la base de datos de origen y, a continuación, realizar una

evaluación. A continuación, AWS SCTevalúa el esquema y genera el informe de evaluación. Los informes de evaluación incluyen un resumen ejecutivo, la complejidad y el esfuerzo de migración, los motores de bases de datos de destino adecuados y recomendaciones para la conversión. Para descargarlo AWS SCT, consulte Instalación, verificación y actualización AWS SCT en la AWS SCT documentación.

La siguiente tabla muestra un ejemplo de resumen ejecutivo generado por AWS SCT para mostrar la complejidad que implica cambiar la base de datos a diferentes plataformas de destino.

Plataforma de destino	Cambios automáticos o mínimos			Acciones complejas			
	Objetos de almacenamiento	Objetos de código	Acciones de conversión	Objetos de almacenamiento		Objetos de código	
Amazon RDS para MySQL	60 (98%)	8 (35%)	42	(12%)	1	15 (65%)	56
Amazon Aurora MySQL-Compatible Edition	60 (98%)	8 (35%)	42	(12%)	1	15 (65%)	56
Amazon RDS para PostgreSQL	60 (98%)	12 (52%)	54	(12%)	1	11 (48%)	26
Edición de	60 (98%)	12 (52%)	54	(12%)	1	11 (48%)	26

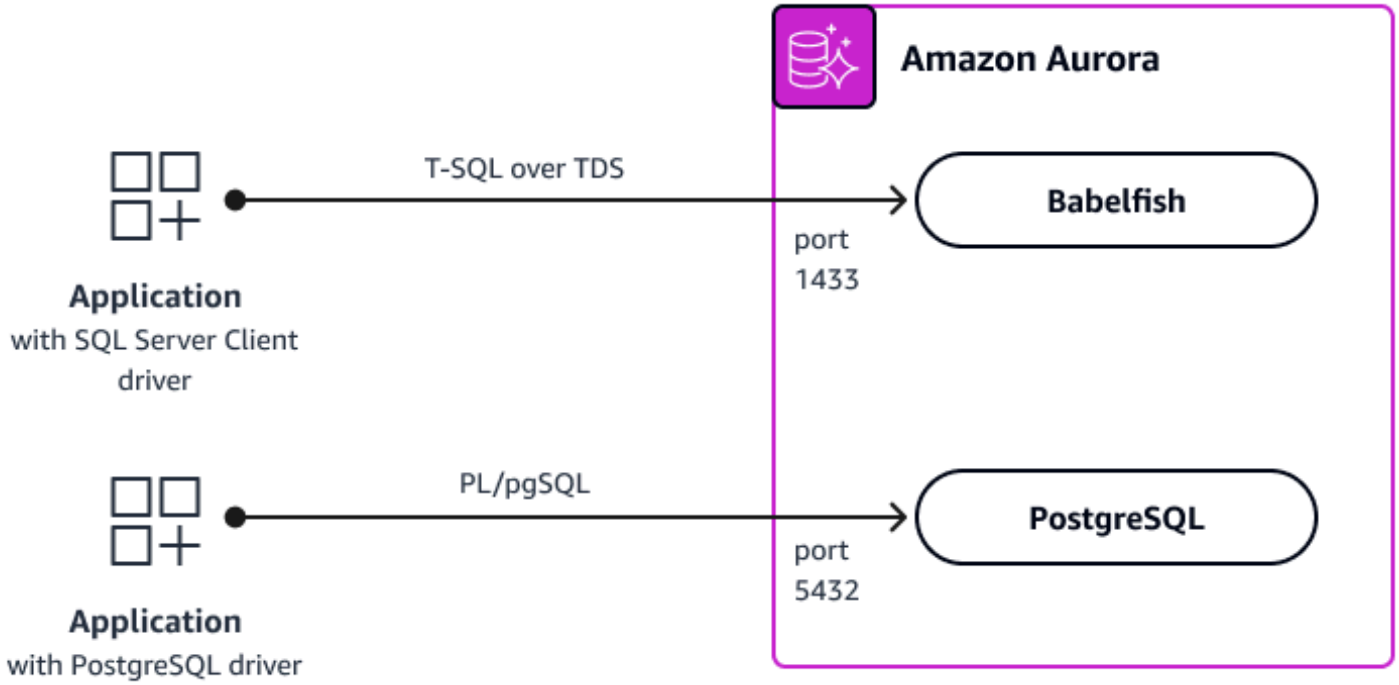
Amazon Aurora compatible con PostgreSQL							
Amazon RDS para MariaDB	60 (98%)	7 (30%)	42	(12%)	1	16 (70%)	58
Amazon Redshift	61 (100%)	9 (39%)	124	0 (0%)	0	14 (61%)	25
AWS Glue	0 (0%)	17 (100%)	0	0 (0%)	0	0 (0%)	0
Babelfish	59 (97%)	10 (45%)	20	(23%)	2	12 (55%)	30

Un AWS SCT informe también proporciona detalles sobre los elementos del esquema que no se pueden convertir automáticamente. Puede cerrar las brechas AWS SCT de conversión y optimizar los esquemas de destino consultando los manuales de AWS migración. Existen muchos manuales de migración de bases de datos para facilitar las migraciones heterogéneas.

Babelfish para Aurora PostgreSQL

Babelfish for Aurora PostgreSQL amplía Aurora PostgreSQL con la capacidad de aceptar conexiones a bases de datos desde clientes de SQL Server. Babelfish permite que las aplicaciones que se crearon originalmente para SQL Server funcionen directamente con Aurora PostgreSQL, con pocos cambios de código y sin cambiar los controladores de la base de datos. Babelfish hace que Aurora PostgreSQL sea bilingüe para que Aurora PostgreSQL pueda funcionar con los lenguajes T-SQL y PL/pgSQL. Babelfish minimiza los esfuerzos de migración de SQL Server a Aurora PostgreSQL. Esto acelera las migraciones, minimiza el riesgo y reduce los costos de migración de manera significativa. Puede seguir utilizando las migraciones de T-SQL después de las migraciones, pero también existe la opción de utilizar las herramientas nativas de PostgreSQL para el desarrollo.

El siguiente diagrama ilustra cómo una aplicación que usa T-SQL se conecta al puerto predeterminado 1433 de SQL Server y usa el traductor Babelfish para comunicarse con la base de datos Aurora PostgreSQL, mientras que una aplicación que usa PL/pgSQL puede conectarse directa y simultáneamente a la base de datos Aurora PostgreSQL mediante el puerto predeterminado 5432 en Aurora PostgreSQL.



Babelfish no admite determinadas funciones de T-SQL de SQL Server. Por este motivo, Amazon proporciona herramientas de evaluación para line-by-line analizar tus sentencias SQL y determinar si alguna de ellas no es compatible con Babelfish.

Hay dos opciones para las evaluaciones de Babelfish. AWS SCT puede evaluar la compatibilidad de su base de datos de SQL Server con Babelfish. Otra opción es la herramienta Babelfish Compass, que es una solución recomendada porque la herramienta Compass se actualiza en función de las nuevas versiones de Babelfish para Aurora PostgreSQL.

Babelfish Compass

Babelfish Compass es una herramienta de descarga gratuita que se ajusta a la última versión de Babelfish para Aurora PostgreSQL. Por el contrario, AWS SCT admitirá las versiones más nuevas de Babelfish después de algún tiempo. Babelfish Compass se ejecuta en el esquema de base de datos de SQL Server. También puede extraer el esquema de base de datos de SQL Server de origen mediante herramientas como SQL Server Management Studio (SSMS). A continuación, puede ejecutar el esquema a través de Babelfish Compass. Esto genera el informe que detalla la

compatibilidad del esquema de SQL Server con Babelfish y si es necesario realizar algún cambio antes de la migración. La herramienta Babelfish Compass también puede automatizar muchos de estos cambios y, en última instancia, acelerar las migraciones.

Una vez finalizados la evaluación y los cambios, puede migrar el esquema a Aurora PostgreSQL mediante herramientas nativas de SQL Server, como SSMS o sqlcmd. Para obtener instrucciones, consulte la publicación Migrar de SQL Server a Amazon Aurora con Babelfish en el blog de AWS bases de datos.

AWS Database Migration Service

Una vez migrado el esquema, puede usar AWS Database Migration Service (AWS DMS) para migrar los datos AWS con un tiempo de inactividad mínimo. AWS DMS no solo carga todos los datos, sino que también replica los cambios del origen al destino mientras el sistema de origen está en funcionamiento. Una vez sincronizadas las bases de datos de origen y destino, la actividad de transición puede tener lugar en la que la aplicación apunta a la base de datos de destino que finaliza la migración. AWS DMS actualmente solo realiza una carga de datos completa con Babelfish para un destino de Aurora PostgreSQL y no replica los cambios. Para obtener más información, consulte Uso de Babelfish como objetivo en la documentación. AWS Database Migration Service AWS DMS

AWS DMS puede realizar migraciones homogéneas (en el mismo motor de base de datos) y heterogéneas (en diferentes motores de base de datos). AWS DMS admite muchos motores de bases de datos de origen y destino. Para obtener más información, consulte la publicación Migración de la base de datos de SQL Server a Amazon RDS for SQL Server AWS DMS mediante AWS el blog de bases de datos.

Recursos adicionales

- Adiós Microsoft SQL Server, hola Babelfish (blog de AWS noticias)
- Convierta esquemas de bases de datos y SQL de aplicaciones mediante la AWS Schema Conversion Tool CLI (blog de AWS bases de datos)
- Migre SQL Server a Amazon Aurora PostgreSQL con las prácticas recomendadas y las lecciones aprendidas sobre el terreno AWS (blog sobre bases de datos)
- Valide los objetos de la base de datos después de la migración de Microsoft SQL Server a Amazon RDS para PostgreSQL y Amazon Aurora PostgreSQL (blog sobre bases de datos)AWS

Optimice el almacenamiento para SQL Server

Descripción general

Esta sección se centra en las optimizaciones de costes del almacenamiento SSD de Amazon Elastic Block Store (Amazon EBS) para SQL Server en cargas de trabajo. EC2

Dispone de una amplia variedad de opciones de almacenamiento para implementar y ejecutar cargas de trabajo de SQL Server. AWS La selección del almacenamiento adecuado debe basarse en el propósito, la arquitectura, la durabilidad, el rendimiento, la capacidad y el costo. AWS los clientes que ejecutan cargas de trabajo de SQL Server suelen utilizar una combinación de almacenamiento de Amazon EBS FSx, NVMe Amazon y Amazon Simple Storage Service (Amazon S3).

Amazon EBS es un almacenamiento conectado a la red conectado a instancias EC2 informáticas y que se utiliza para almacenar y procesar archivos generales del sistema operativo, las aplicaciones, las bases de datos y las copias de seguridad. El almacenamiento en unidades de estado sólido (SSD) de Amazon EBS incluye SSD de uso general (gp2 y gp3) y SSD de IOPS aprovisionadas (io1, io2 e io2bx). Considere lo siguiente:

- Algunas EC2 instancias, como la r5d, tienen una conexión física local a la instancia host. NVMe SSDs Estos volúmenes proporcionan un almacenamiento a nivel de bloque que se suele utilizar para la extensión tempdb o el grupo de búferes de SQL Server.
- Amazon FSx for Windows File Server es un servicio de almacenamiento de archivos totalmente gestionado, mientras que Amazon FSx for NetApp ONTAP es un almacenamiento compartido totalmente gestionado basado en el popular sistema NetApp de archivos ONTAP. Amazon FSx se utiliza con frecuencia para ejecutar cargas de trabajo de SQL Server en una configuración de instancia agrupada de conmutación por error (FCI) de SQL Server de alta disponibilidad. Esta solución aloja archivos de registro y datos de SQL Server, lo que reduce los requisitos de rendimiento de EBS en las instancias. EC2
- Amazon S3 es un servicio de almacenamiento de objetos que ofrece escalabilidad, disponibilidad de datos, seguridad y rendimiento líderes del sector. Puede almacenar archivos de backup nativos de SQL Server AMIs, instantáneas de EBS, registros de aplicaciones y mucho más en Amazon S3.

Tipos de almacenamiento SSD, rendimiento y costo para Amazon EBS

Los costes de almacenamiento en SSD de Amazon EBS suelen aumentar a medida que aumentan la durabilidad y el rendimiento. Actualmente, el almacenamiento viene en cinco tipos de volúmenes,

cada uno con sus propias métricas de rendimiento únicas. Para obtener un resumen de los casos de uso y las características de los volúmenes respaldados por SSD, consulte la tabla de la sección de volúmenes de unidades de estado sólido (SSD) de la documentación de Amazon EBS.

Puedes usar Amazon CloudWatch para monitorear el rendimiento de las SSD, capturar datos de tendencias y configurar alarmas cuando se alcancen ciertos umbrales. Si está ejecutando cargas de trabajo de SQL Server AWS, considere la posibilidad de habilitar una supervisión detallada y de implementar métricas CloudWatch personalizadas para capturar métricas detalladas de rendimiento del volumen, como la latencia del disco, las IOPS, el rendimiento, la longitud de la cola de discos, la capacidad utilizada frente a la capacidad libre, etc. Puede utilizar estas métricas de CloudWatch rendimiento para identificar el almacenamiento subaprovisionado y sobreaprovisionado y proporcionar puntos de datos históricos para definir con precisión los requisitos de almacenamiento.

Los costes de almacenamiento en SSD de Amazon EBS también varían en función de la capacidad asignada. En la siguiente tabla se muestra una comparación de los distintos tipos de volumen. Todos los tipos de volumen tienen 1 TB de capacidad y configuraciones de rendimiento similares.

Tipo de volumen	IOPS máximos (E/S de 16 KiB)	Rendimiento máximo (E/S de 128 KiB)	Precio por 1 TB	Porcentaje de ahorro de costes
gp2	3000	250	102,40$	
gp3	3000	250	86,92 DÓLARES	15%
io1	16,000	500	1.168 DÓLARES	
io2	16,000	500	1.168 DÓLARES	
gp3	16,000	500	146,92 DÓLARES	87%
io2bx	16,000	4.000	1.168 DÓLARES	
gp3	16,000	1 000	181,92 DÓLARES	84%

> **Note**
>
> Las métricas de rendimiento y coste de la tabla anterior son por volumen y se basan en una estimación de Calculadora de precios de AWS. Cuenta de AWS Se requiere un número para acceder a la estimación de Calculadora de precios de AWS.

Los volúmenes gp3 de SSD de Amazon EBS ofrecen un rendimiento excelente a un bajo coste. Puede ahorrar hasta un 87 por ciento si elige un volumen gp3 en lugar de los volúmenes io1 o io2 para cargas de trabajo que requieren menos de 16 000 IOPS y un rendimiento de 500. MiBps

Los volúmenes io2 Block Express (io2Bx) ofrecen un rendimiento superior al de los volúmenes io2 normales. Con 16 000 IOPS, los volúmenes io1 o io2 solo tienen un rendimiento de 500, mientras que los volúmenes de io2Bx se pueden configurar con MiBps un rendimiento de hasta 4000. MiBps En comparación con los volúmenes io1 e io2, los volúmenes io2Bx ofrecen más de cuatro veces más rendimiento, entre 16 000 y 64 000 IOPS, exactamente al mismo precio. Los volúmenes io2 normales se pueden convertir en volúmenes io2bx adjuntándolos a instancias compatibles con io2bx. EC2 Para obtener una lista de las EC2 instancias compatibles con IO2BX, consulte los volúmenes SSD de IOPS aprovisionados en la documentación de Amazon EBS. Antes de implementar un nuevo almacenamiento, puede usarlo Calculadora de precios de AWSpara estimar su costo mensual y comprender el impacto en el costo en función de las compensaciones entre durabilidad, rendimiento y capacidad.

Optimización general de costos de SSD para Amazon EBS

Le recomendamos que evalúe lo que está almacenando y que se asegure de utilizar el tipo y la clase de almacenamiento correctos. Por ejemplo, Amazon S3 ofrece un precio excelente, políticas de ciclo de vida integradas y opciones de replicación ideales para las copias de seguridad de SQL Server. SQL Server 2022 tiene la capacidad de realizar copias de seguridad directamente en Amazon S3, mientras que las versiones anteriores de SQL Server se basan en copias de seguridad locales nativas. Si utiliza versiones anteriores de SQL Server, considere la posibilidad de realizar copias de seguridad en volúmenes de disco duro de Amazon EBS y, a continuación, copiar la copia de seguridad en Amazon S3. Esta solución puede ahorrar un 53 por ciento en comparación con el uso de volúmenes gp3 para las copias de seguridad.

La siguiente tabla muestra la diferencia de precio de 1 TB de almacenamiento en Amazon EBS gp3, Amazon EBS HDD st1 y Amazon S3.

Tipo de almacenamiento	Capacidad	Precio: pm
EBS gp3 500 MiBps	1 TB	96,92 DÓLARES
EBS st1 Burst 500 MiBps		46,08 DÓLARES
S3 Standard		23,55 DÓLARES
Estándar S3 (acceso poco frecuente)		12,80$
S3 Glacier Deep Archive		1,03$

> **Note**
>
> Las métricas de costos de la tabla anterior se basan en una estimación de Calculadora de precios de AWS. Cuenta de AWS Se requiere un número para acceder a la estimación de Calculadora de precios de AWS.

Le recomendamos que tenga en cuenta lo siguiente:

- Habilite una supervisión detallada e implemente métricas CloudWatch personalizadas para capturar con precisión sus requisitos de rendimiento de almacenamiento.
- Actualice el almacenamiento de Amazon EBS de gp2 a gp3 para reducir costos, aumentar la flexibilidad y mejorar el rendimiento.
- Actualice el almacenamiento de Amazon EBS de io1 a io2 para aumentar la durabilidad y la flexibilidad del rendimiento.
- Utilice io2Bx en lugar de io1 o io2 cuando sea posible para aumentar la durabilidad y el rendimiento.
- Al elegir el almacenamiento, considere un mix-and-match enfoque que ayude a reducir los requisitos de capacidad y el costo de los volúmenes de alto rendimiento. Por ejemplo, podría utilizar volúmenes gp3 de bajo coste para el volumen raíz (sistema operativo), la instalación de SQL Server, las bases de datos del sistema (excepto tempdb) y las bases de datos de usuarios de menor rendimiento. Esto podría ayudar a reducir la capacidad y el coste de un volumen io2, que puede dedicarse a bases de datos de usuarios de alto rendimiento.

- Si aloja bases de datos de SQL Server AWS, le recomendamos que utilice varios archivos de datos de SQL Server por base de datos. Esto brinda la oportunidad de distribuir las cargas de trabajo de lectura y escritura en varios volúmenes, lo que reduce los requisitos de rendimiento y capacidad por volumen y, en consecuencia, reduce los costos.
- Incluso si las cargas de trabajo de producción requieren un almacenamiento de mayor rendimiento, como io1 o io2/io2Bx, considere los volúmenes gp3 para las cargas de trabajo que no son de producción para ayudar a reducir los costos.
- Realice un seguimiento y establezca tendencias en el uso del almacenamiento a lo largo del tiempo para identificar fácilmente los picos de uso y los costes inesperados.
- Úselo AWS Compute Optimizerpara obtener recomendaciones sobre cómo aumentar o reducir los volúmenes de EBS en función de la utilización real.
- Utilice la elasticidad de AWS para ajustar las necesidades de rendimiento y capacidad de sus volúmenes de SSD para Amazon EBS. A diferencia de los entornos locales, no es necesario aprovisionar en exceso el rendimiento y la capacidad del almacenamiento para futuras cargas de trabajo. Puede migrar sus cargas de trabajo actuales de SQL Server AWS y ajustar el rendimiento o la capacidad según sea necesario y, al mismo tiempo, mantener sus bases de datos en línea.

Recursos adicionales

- Tipos de volúmenes de Amazon EBS (documentación de Amazon EBS)
- Amazon Elastic Block Store (Amazon EBS) (documentación de Amazon EBS)
- Volúmenes SSD de IOPS aprovisionados (documentación de Amazon EBS)
- Volúmenes de almacenes de instancias SSD (EC2 documentación de Amazon)
- CloudWatch Métricas de Amazon para Amazon EBS (documentación de Amazon EBS)
- Especificaciones de las instancias optimizadas para EC2 almacenamiento de Amazon (EC2documentación de Amazon)
- Proteja sus cargas de trabajo de SQL Server NetApp SnapCenter con Amazon FSx for NetApp ONTAP (blog sobreAWS almacenamiento)
- EC2 Preguntas frecuentes de Amazon (página AWS del producto)

Optimice las licencias de SQL Server mediante Compute Optimizer

Guía sobre cómo optimizar las licencias de SQL Server mediante el uso AWS Compute Optimizer.

Descripción general

AWS Compute Optimizerpuede recomendar oportunidades de optimización de licencias para cargas de trabajo de Microsoft SQL Server en Amazon Elastic Compute Cloud (Amazon EC2). Compute Optimizer puede proporcionar recomendaciones automatizadas para reducir los costos de licencia. Las recomendaciones de Compute Optimizer aparecen junto a cada una de tus EC2 instancias con licencias de Microsoft SQL Server. La información que se proporciona incluye las oportunidades de ahorro recomendadas, EC2 por ejemplo, los precios bajo demanda y los precios de compra tu propia licencia (BYOL) por hora. Esta información puede ayudarle a decidir si debe actualizar su edición de licencia a una versión inferior.

Compute Optimizer descubre automáticamente tus instancias de SQL Server en Amazon EC2 según el tipo de carga de trabajo inferida. Para ver las recomendaciones de licencias, puedes seleccionar las instancias de SQL Server en Compute Optimizer y, a continuación, autenticarte con Amazon CloudWatch Application Insights con tus credenciales de base de datos de solo lectura. Compute Optimizer analiza si utilizas alguna de las funciones de la edición SQL Server Enterprise. Si no se utilizan funciones de la edición Enterprise, Compute Optimizer recomienda que cambies a la edición Standard para reducir los costos de licencia.

También puedes usar Compute Optimizer para hacer recomendaciones de tamaño para tus EC2 instancias de Amazon que ejecutan cargas de trabajo de SQL Server. Para obtener más información, consulta Optimizar el tamaño de SQL Server mediante Compute Optimizer en esta guía.

Recomendaciones de optimización de costos

Las recomendaciones de licencia de Compute Optimizer pueden ayudarte a evaluar las funciones que utilizas en Microsoft SQL Server y a elegir la edición más rentable para tus cargas de trabajo. La edición SQL Server Enterprise es significativamente más cara que la edición Standard. Para obtener más información, consulte Comparar las ediciones de SQL Server en esta guía y consulte los precios de SQL Server 2022 en el sitio web de Microsoft. Invertir tiempo en configurar Compute Optimizer para evaluar tu flota de SQL Server y ofrecer recomendaciones puede reducir drásticamente los costos de licencias.

La página de detalles de la licencia proporciona la siguiente información:

- Usa la tabla para comparar la configuración de licencia actual (como la edición, el modelo y la cantidad de núcleos de instancia) con las recomendaciones de Compute Optimizer.
- Usa los gráficos de uso para revisar la cantidad de funciones de la edición Enterprise que se usaron durante el período de análisis.

Para obtener más información, consulta Cómo ver los detalles de una recomendación de licencia de software comercial en la documentación de Compute Optimizer.

Configurar Compute Optimizer

Compute Optimizer analiza las licencias de software comercial mediante la `mssql_enterprise_features_used` métrica. Para obtener más información sobre esta métrica, consulta Métricas para licencias de software comercial.

1. Asegúrate de tener los permisos adecuados para suscribirte a Compute Optimizer. Para obtener más información, consulte los siguientes temas:
   - Política de suscripción a Compute Optimizer
   - Políticas para conceder acceso a Compute Optimizer de forma independiente Cuentas de AWS
   - Políticas para conceder acceso a Compute Optimizer para una cuenta de administración de una organización
2. Adjunta las funciones y la política de instancia requeridas para CloudWatch Application Insights. Para obtener instrucciones, consulte las Políticas para habilitar las recomendaciones de licencias de software comercial.
3. Habilite CloudWatch Application Insights mediante sus credenciales de base de datos de Microsoft SQL Server. Para obtener instrucciones, consulte Configurar la aplicación para la supervisión en la CloudWatch documentación.

   > **Note**
   >
   > Para generar recomendaciones de licencias de software comercial, se requieren al menos 30 horas consecutivas de datos CloudWatch métricos. Para obtener más información, consulte los requisitos de CloudWatch métricas.

4. Utilice la siguiente consulta SQL para configurar el acceso con privilegios mínimos para CloudWatch Application Insights.

   ```
   GRANT VIEW SERVER STATE TO [LOGIN];
   GRANT VIEW ANY DEFINITION TO [LOGIN];
   ```

   Esto habilita un nuevo servicio, SQL. PrometheusSqlExporter
5. Desde la cuenta de administración de destino Cuenta de AWS o de la organización, activa Compute Optimizer. Para obtener instrucciones, consulta Cómo optar por una cuenta.

> **Note**
>
> Una vez haya realizado la activación, los resultados y las recomendaciones de optimización pueden tardar hasta 24 horas en generarse.

6. En la consola de Compute Optimizer, selecciona Licencias en el panel de navegación.
7. En la columna Hallazgos, busca cualquier instancia en la que se indique que las métricas son insuficientes. Compute Optimizer devuelve este resultado si detecta que CloudWatch Application Insights no está habilitada o tiene permisos insuficientes. Para obtener más información, consulta Cómo encontrar motivos. Haga lo siguiente para resolver estos hallazgos:
   a. Seleccione la instancia.
   b. Agrega un secreto.
   c. Confirma que la función y la política de la instancia estén asociadas.
   d. Seleccione Habilitar las recomendaciones de licencia.
8. En la columna Hallazgos, busca cualquier instancia que tenga el resultado No optimizado. Compute Optimizer devuelve este resultado si detecta que tu EC2 infraestructura de Amazon no utiliza ninguna de las funciones de licencia de Microsoft SQL Server por las que estás pagando. Para obtener más información, consulta Cómo encontrar motivos. Haga lo siguiente para resolver estos hallazgos:
   a. Seleccione la instancia.
   b. Compare la edición de licencia actual con la edición recomendada.
   c. Revise el gráfico de uso de licencias actual.
   d. Si quiere bajar la licencia a una versión inferior, elija Implementar la recomendación.
   e. Revise los requisitos y siga las instrucciones para cambiar la licencia a una versión inferior. Si desea automatizar el proceso, consulte Cambiar la versión Enterprise de SQL Server mediante AWS Systems Manager Document para reducir los costes (AWS blog).

Recursos adicionales

- Reduzca los costos de licencias de Microsoft SQL Server con AWS Compute Optimizer (AWS Blog)
- ¿Qué es AWS Compute Optimizer? (AWS documentación)
- Visualización de las recomendaciones de licencias de software comercial (AWS documentación)

- Cambie a una versión anterior de su edición de Microsoft SQL Server (AWS documentación)
- Microsoft SQL Server en AWS (AWS)
- Licencias de Microsoft en AWS (AWS)
- Precios de Microsoft SQL Server 2019 (Microsoft)
- Precios de Microsoft SQL Server 2022 (Microsoft)

Optimice el tamaño de SQL Server mediante Compute Optimizer

Descripción general

AWS Compute Optimizerayuda a los administradores de bases de datos (DBAs) a descubrir las cargas de trabajo de Microsoft SQL Server en Amazon Elastic Compute Cloud (Amazon EC2) y a dimensionar EC2 las instancias correctamente para reducir los costos de licencia hasta un 25%. La función de tipo de carga de trabajo inferida de Compute Optimizer utiliza el aprendizaje automático (ML) y detecta automáticamente las aplicaciones que podrían estar ejecutándose en AWS tus recursos. Compute Optimizer incluye soporte para SQL Server como un tipo de carga de trabajo inferida. Al utilizar la función de tipo de carga de trabajo inferido, puede identificar oportunidades de ahorro de costes en función de la carga de trabajo específica que se ejecute en sus instancias de Amazon. EC2

Con esta función, puede clasificar las oportunidades de ahorro de costes según los tipos de carga de trabajo inferida compatibles, como SQL Server. Compute Optimizer puede detectar automáticamente las EC2 instancias de SQL Server que están sobreaprovisionadas. Puedes cambiar a la EC2 consola para reducir el tamaño de la instancia, lo que ayuda a reducir los costos de licencia e infraestructura.

También puedes usar Compute Optimizer para hacer recomendaciones de licencias de SQL Server. Para obtener más información, consulta Optimizar las licencias de SQL Server mediante Compute Optimizer en esta guía.

Configurar Compute Optimizer

Para obtener instrucciones sobre cómo usar Compute Optimizer con cargas de trabajo inferidas de SQL Server, consulte Optimización del rendimiento y reducción de los costos de licencia: aprovechamiento de las instancias de AWS Compute Optimizer Amazon EC2 SQL Server (blog).AWS Puedes optar por cuentas independientes, cuentas que sean miembros de una organización y cuentas de administración de una organización. En el caso de las cuentas independientes y de miembros, la opción habilita Compute Optimizer únicamente para esa cuenta.

En el caso de una cuenta de administración de la organización, puedes elegir si deseas habilitar Compute Optimizer solo en esa cuenta o para todas las cuentas de los miembros de la organización.

El proceso de suscripción de Compute Optimizer crea automáticamente un rol vinculado al servicio AWS Identity and Access Management (IAM). Para obtener más información, consulte Uso de roles vinculados a servicios para AWS Compute Optimizer.

Compute Optimizer analiza los recursos en función de CloudWatch las métricas de Amazon, como la CPU, las E/S, la red y el uso de Amazon Elastic Block Store (Amazon EBS). Para generar recomendaciones, se requieren al menos 30 horas consecutivas de datos de CloudWatch métricas en los últimos 14 días. Si habilita la función de métricas de infraestructura mejoradas, las métricas de uso se ampliarán a 93 días. Para obtener más información, consulta los requisitos de CloudWatch métricas y las métricas de infraestructura mejorada en la documentación de Compute Optimizer.

Compute Optimizer ofrece opciones y los ahorros asociados a cada opción, en función de la vCPU, la memoria, el almacenamiento, la red, el riesgo y el esfuerzo de migración. Puedes usar el panel de CloudWatch métricas para analizar los datos que se utilizan para hacer la recomendación. Con estos datos, puede ajustar el tamaño de EC2 las instancias que ejecutan cargas de trabajo de SQL Server. Para obtener más información sobre cómo cambiar el tipo de instancia, consulta Cambiar el tipo de instancia en la EC2 documentación de Amazon.

Recursos adicionales

- AWS Compute Optimizer identifica y filtra las cargas de trabajo de Microsoft SQL Server ()AWS
- Optimización del rendimiento y reducción de los costes de licencia: aprovechamiento AWS Compute Optimizer de las instancias de Amazon EC2 SQL Server (AWS blog)
- ¿Qué es? AWS Compute Optimizer (AWS documentación)
- Visualización de las recomendaciones de EC2 instancias (AWS documentación)

Revise Trusted Advisor las recomendaciones para las cargas de trabajo de SQL Server

Descripción general

AWS Trusted Advisorproporciona recomendaciones que le ayudan a seguir las AWS mejores prácticas. Al analizar el uso, la configuración y los gastos, Trusted Advisor ofrece recomendaciones prácticas para reducir los costes, mejorar la disponibilidad y el rendimiento del sistema o ayudar a

cerrar las brechas de seguridad. Esta sección se centra en las Trusted Advisor comprobaciones que pueden ayudarlo a reducir los costos de operación de las cargas de trabajo de SQL Server en. Nube de AWS

Recomendaciones de optimización de costos

Trusted Advisor proporciona recomendaciones que le ayudan a optimizar las cargas de trabajo de SQL Server en Amazon Elastic Compute Cloud (Amazon EC2). Las comprobaciones inspeccionan las cargas de trabajo de SQL Server y enumeran automáticamente las instancias que necesitan ser optimizadas. Poner en práctica Trusted Advisor las recomendaciones puede reducir los costos y mejorar la postura de seguridad de su organización.

Las siguientes son Trusted Advisor comprobaciones que se centran en Microsoft SQL Server:

- EC2 Instancias de Amazon sobreaprovisionadas para Microsoft SQL Server: esta comprobación analiza las EC2 instancias de Amazon que ejecutan SQL Server y te avisa si una instancia supera el límite de vCPU del software de SQL Server. Por ejemplo, una instancia con la edición estándar de SQL Server puede utilizar hasta 48 versiones. CPUs Una instancia con SQL Server Web puede utilizar hasta 32 vCPUs.

Edición	vCPU mínima	vCPU máx.
Web	4	32
Estándar	4	48
Enterprise	4	Límites del sistema operativo

- Consolidación de EC2 instancias de Amazon para Microsoft SQL Server: esta comprobación analiza sus EC2 instancias de Amazon y le avisa si su instancia tiene un número inferior al mínimo de licencias de SQL Server. Puede consolidar instancias de SQL Server más pequeñas para ayudar a reducir los costos. Si tiene muchas instancias pequeñas de SQL Server con licencia incluida, considere la posibilidad de consolidarlas. Según la guía de licencias de Microsoft SQL Server 2019, SQL Server requiere un mínimo de 4 licencias de vCPU por instancia. Si consolida estas bases de datos, puede ahorrar en costos de licencias. Puede tomar su decisión en función del número de bases de datos de la instancia, el tamaño máximo de la base de datos y el tamaño total de las bases de datos. La consolidación es compatible con las ediciones Web, Standard y Enterprise de SQL Server. Para obtener más información, consulte Consolidación de bases de datos de SQL Server (entrada del blog de Microsoft).

AWS no recomienda colocar bases de datos de producción de gran tamaño en un solo servidor. Sin embargo, puede consolidar las más pequeñas que se utilizan en entornos que no son de producción, como el desarrollo, las pruebas y la puesta en escena. Esto depende del uso actual de SQL Server; si tiene bases de datos de bajo uso, puede consolidarlas en un servidor.

Configurar Trusted Advisor

Haga lo siguiente para evaluar los registros centrados en Trusted Advisor SQL Server.

1. Inicie sesión en AWS Management Console.
2. Abra la consola de AWS Trusted Advisor.
3. En el panel de navegación, en Recomendaciones, elija Optimización de costes.
4. En la lista de comprobaciones de optimización de costes, revise el estado de la consolidación de las EC2 instancias de Amazon para Microsoft SQL Server y las EC2instancias de Amazon sobreaprovisionadas para las comprobaciones de Microsoft SQL Server.
   - Los símbolos de verificación verdes indican que tus EC2 instancias de Amazon están configuradas de forma óptima.
   - Los símbolos de alerta anaranjados indican que hay oportunidades de mejora.
5. Seleccione una marca para ver sus detalles y recomendaciones.
6. Siga las instrucciones que se proporcionan en la comprobación para optimizar las EC2 instancias de Amazon que ejecutan cargas de trabajo de SQL Server.
7. Supervise sus instancias con regularidad y actualice las comprobaciones periódicamente.

Recursos adicionales

- Trusted Advisor compruebe la referencia (AWS documentación)
- Microsoft SQL Server en AWS (AWS)
- Licencias de Microsoft en AWS (AWS)
- Precios de SQL Server 2019 (Microsoft)
- AWS Launch Wizard para SQL Server (AWS documentación)

Contenedores

La modernización es un proceso de transformación que ofrece muchas opciones, como descomponer monolitos en microservicios, rediseñar la arquitectura de las aplicaciones para que se basen en eventos mediante el uso de funciones sin servidor (AWS Lambda) y reutilizar bases de datos de SQL Server a Amazon Aurora o bases de datos gestionadas específicamente. Las vías de modernización para cambiar la plataforma de las aplicaciones de.NET Framework a contenedores de Linux y Windows requieren menos esfuerzo que otras opciones de modernización. Los contenedores ofrecen las siguientes ventajas:

- Acelere la innovación: la migración a los contenedores facilita la automatización de las etapas del ciclo de vida del desarrollo, que incluyen la creación, las pruebas y la implementación de aplicaciones. Al automatizar estos procesos, los equipos de desarrollo y operaciones disponen de más tiempo para centrarse en la innovación.
- Reduzca el costo total de propiedad (TCO): pasar a los contenedores también puede reducir su dependencia de las herramientas de administración de licencias y protección de terminales. Como los contenedores son unidades informáticas efímeras, puede automatizar y simplificar las tareas de administración, como la aplicación de parches, el escalado y las copias de seguridad y la restauración. Esto reduce el coste total de propiedad de la administración y el funcionamiento de las cargas de trabajo basadas en contenedores. Por último, los contenedores son más eficientes en comparación con las máquinas virtuales, ya que puede usarlos para maximizar la ubicación de las aplicaciones al proporcionar un mejor aislamiento. Esto aumenta la utilización de los recursos de infraestructura de la aplicación.
- Mejore la utilización de los recursos: los contenedores son más eficientes en comparación con las máquinas virtuales porque puede usarlos para maximizar la ubicación de las aplicaciones. Esto aumenta la utilización de los recursos de infraestructura de la aplicación al proporcionar un mejor aislamiento.
- Cierre la brecha de habilidades: AWS ofrece días de inmersión para capacitar a sus equipos de desarrollo en tecnología y DevOps prácticas de contenedores.

Esta sección abarca los siguientes temas:

- Mueva las aplicaciones de Windows a contenedores
- Optimice los costos de AWS Fargate las tareas en Amazon ECS
- Obtenga visibilidad de sus costos de Amazon EKS

- Cambie la plataforma de las aplicaciones de Windows con App2Container

Para obtener información sobre licencias, consulte la sección de licencias de Amazon Web Services y Microsoft: Preguntas frecuentes o envíe sus preguntas por correo electrónico a microsoft@amazon.com.

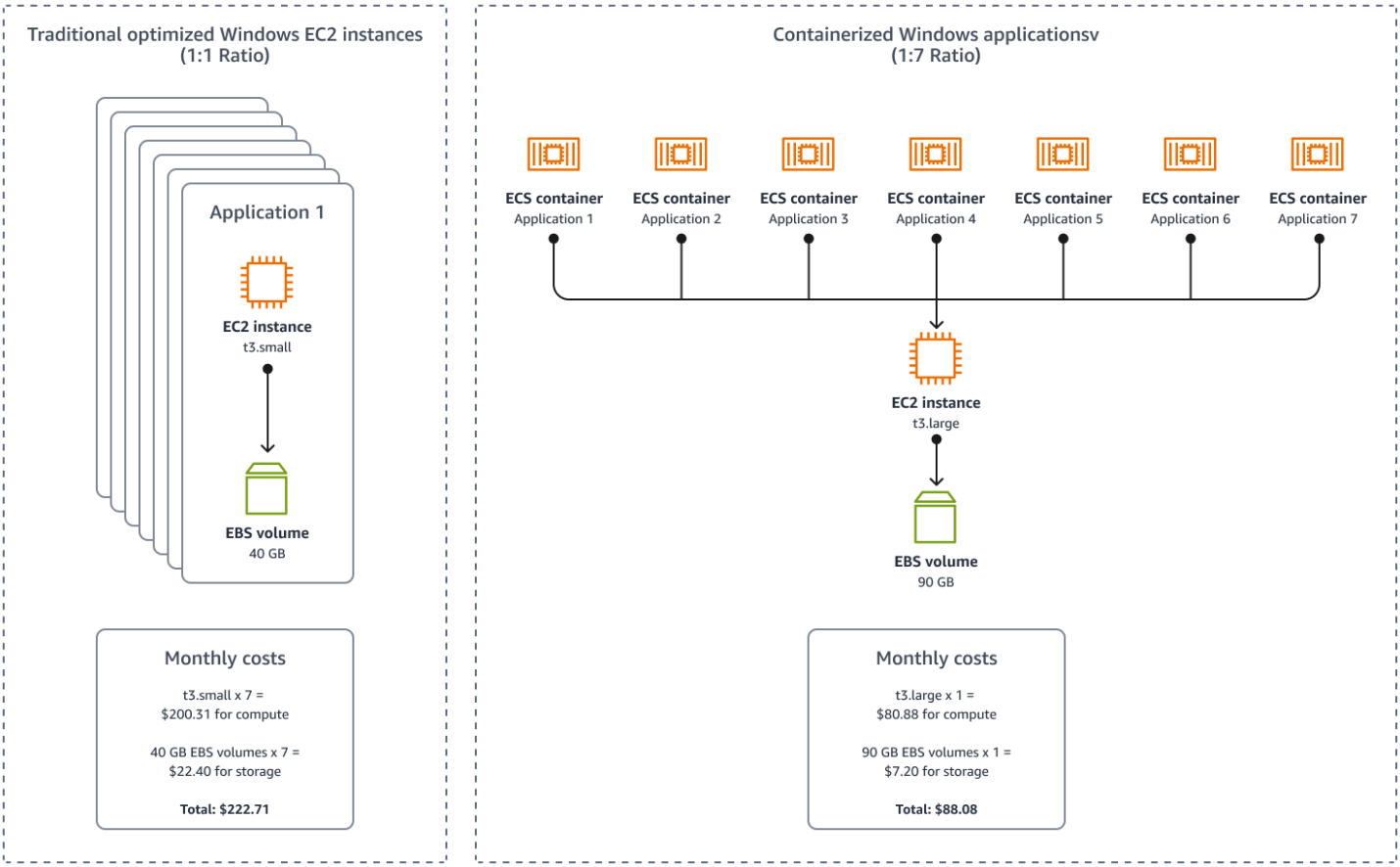
Mueva las aplicaciones de Windows a contenedores

Descripción general

Según la encuesta anual de 2021 de la CNCF, el 96 por ciento de las organizaciones utilizan o evalúan los contenedores para modernizar su infraestructura. Esto se debe a que los contenedores pueden ayudar a su organización a reducir el riesgo, aumentar la eficiencia y la velocidad operativas y permitir la agilidad. También puede usar contenedores para reducir el costo de ejecución de sus aplicaciones. En esta sección se ofrecen recomendaciones para ejecutar contenedores de forma rentable en todos los servicios de AWS contenedores, incluidos Amazon Elastic Container Service (Amazon ECS), Amazon Elastic Kubernetes Service (Amazon EKS) y. AWS Fargate

Beneficios de costos

La siguiente infografía muestra los ahorros de costos que una empresa puede lograr al consolidar sus aplicaciones de ASP.NET Framework en instancias de Amazon Elastic Compute Cloud (Amazon EC2) según una recomendación de evaluación de AWS optimización y licencias (AWS OLA). La siguiente infografía muestra los ahorros adicionales que se pueden lograr al mover las aplicaciones a un contenedor de Windows.



La AWS OLA recomendó que la empresa cambiara a instancias t3.small individuales. La empresa podría lograr estos ahorros ejecutando siete aplicaciones de ASP.NET en servidores locales, como se muestra en el siguiente análisis de utilización del rendimiento.

Not sharable | Nearing end of life | Approximately 15% overhead | The OS accounts for 1.2 GB per instance | Optimized for EC2

	Server name	Storage	Operating system	On-premises CPU AVG utilization	On-premises CPU peak utilization	On-premises RAM (GB)	On-premises RAM AVG utilization (GB)	On-premises RAM peak utilization (GB)	Instance size	vCPU	RAM (GB)
1	AppServer01	60	Windows Server 2012	7.00%	17.00%	8	13.50%	17.10%	t3.small	2	2
2	AppServer02	39	Windows Server 2012	20.07%	22.00%	16	7.50%	12.40%	t3.small	2	2
3	AppServer03	39	Windows Server 2012	24.00%	25.50%	16	8.80%	11.90%	t3.small	2	2
4	AppServer04	4	Windows Server 2012	21.40%	24.00%	16	7.80%	10.70%	t3.small	2	2
5	AppServer05	40	Windows Server 2012	21.30%	23.00%	16	8.20%	12.00%	t3.small	2	2
6	AppServer06	39	Windows Server 2012	21.50%	23.50%	16	7.90%	10.90%	t3.small	2	2
7	AppServer07	39	Windows Server 2012	21.60%	22.90%	16	8.40%	11.50%	t3.small	2	2

Un análisis más detallado reveló que la empresa podría ahorrar aún más en costos si ejecutara sus cargas de trabajo en contenedores. Los contenedores reducen la sobrecarga del sistema operativo en relación con los recursos del sistema, como el uso de la CPU, la RAM y el disco (se explica en la siguiente sección). En este escenario, la empresa podría consolidar las siete aplicaciones en una sola instancia t3.large y aún disponer de 3 GB de RAM de sobra. La migración a contenedores

puede ayudar a la empresa a lograr un ahorro promedio del 64 por ciento en costos de cómputo y almacenamiento al usar contenedores en lugar de Amazon EC2.

Recomendaciones de optimización de costos

En la siguiente sección, se ofrecen recomendaciones para optimizar los costes mediante la consolidación de las aplicaciones y el uso de contenedores.

Reduzca su EC2 huella de Windows en Amazon

Los contenedores de Windows pueden reducir el espacio que EC2 ocupa Windows en Amazon al permitirle consolidar más aplicaciones en menos EC2 instancias. Por ejemplo, supongamos que tiene 500 aplicaciones de ASP.NET. Si ejecutas un núcleo por aplicación para Windows en Amazon EC2, eso equivale a 500 instancias de Windows (t3.small). Si presupone una proporción de 1:7 (que puede aumentar considerablemente según el tipo y el tamaño de la EC2 instancia) para usar contenedores de Windows (con t3.large), solo necesitará aproximadamente 71 instancias de Windows. Esto representa una disminución del 85,8 por ciento en su EC2 presencia de Windows en Amazon.

Reduzca los costes de licencias de Windows

Si licencia una instancia de Windows, no necesita licenciar los contenedores que se ejecutan en esa instancia. Como resultado, la consolidación de las aplicaciones de ASP.NET mediante contenedores de Windows puede reducir considerablemente los costes de licencias de Windows.

Reduzca el espacio de almacenamiento

Cada vez que lanzas una EC2 instancia nueva, creas y pagas un nuevo volumen de Amazon Elastic Block Store (Amazon EBS) para alojar el sistema operativo. A medida que aumenta, el costo también aumenta. Si usa contenedores, puede reducir los costos de almacenamiento porque todos los contenedores comparten el mismo sistema operativo base. Además, los contenedores utilizan el concepto de capas para reutilizar partes inmutables de una imagen de contenedor para todos los contenedores en ejecución basados en esa imagen. En el ejemplo anterior, todos los contenedores ejecutan .NET Framework y, por lo tanto, todos comparten la capa intermedia e inmutable de ASP.NET Framework.

Migre end-of-support los servidores a contenedores

Support para Windows Server 2012 y Windows Server 2012 R2 finalizó el 10 de octubre de 2023. Puede migrar las aplicaciones que se ejecutan en Windows Server 2012 o versiones anteriores

agrupándolas en contenedores para que se ejecuten en nuevos sistemas operativos. De esta forma, evita ejecutar sus aplicaciones en sistemas operativos que no cumplen con las normas y, al mismo tiempo, aprovecha la rentabilidad, la reducción del riesgo, la eficiencia operativa, la velocidad y la agilidad que ofrecen los contenedores.

Una advertencia a tener en cuenta a la hora de adoptar este enfoque es si su aplicación requiere requisitos específicos APIs relacionados con la versión del sistema operativo que se utiliza actualmente (COM Interop, por ejemplo). En este caso, debe probar a mover la aplicación a una versión más reciente de Windows. Los contenedores de Windows alinean su imagen de contenedor base (por ejemplo, Windows Server 2019) con el sistema operativo del host del contenedor (por ejemplo, Windows Server 2019). Las pruebas y el traslado a contenedores pueden facilitar las actualizaciones del sistema operativo en el futuro al cambiar la imagen base de su Dockerfile e implementarla en un conjunto nuevo de hosts que ejecuten la última versión de Windows.

Elimine las licencias y las herramientas de administración de terceros

La administración de su flota de servidores requiere el uso de varias herramientas de operación del sistema de terceros para administrar los parches y la configuración. Esto puede hacer que la administración de la infraestructura sea compleja y, a menudo, se incurre en costos de licencias de terceros. Si utilizas contenedores AWS, no necesitas gestionar nada desde el punto de vista del sistema operativo. El tiempo de ejecución de los contenedores administra los contenedores. Esto significa que el host subyacente es efímero y se puede reemplazar fácilmente. Puede ejecutar sus contenedores sin necesidad de administrar directamente el host del contenedor. Además, puede utilizar herramientas gratuitas, como acceder fácilmente AWS Systems Manager Session Manager al host y solucionar problemas.

Mejore el control y la portabilidad

Los contenedores le proporcionan un control más detallado sobre los recursos del servidor, como la CPU y la RAM, que sobre EC2 las instancias. Por EC2 ejemplo, puedes controlar la CPU y la RAM seleccionando una familia de instancias, un tipo de instancia y opciones de CPU. Sin embargo, con los contenedores, puede definir exactamente la cantidad de CPU o RAM que desea asignar a un contenedor en las definiciones de tareas de ECS o a los pods de Amazon EKS. De hecho, recomendamos especificar la CPU y la memoria a nivel de contenedor para los contenedores de Windows. Este nivel de granularidad aporta beneficios en términos de costes. Considere el siguiente código de ejemplo:

```
json
```

```
{
    "taskDefinitionArn": "arn:aws:ecs:us-east-1:123456789012:task-definition/demo-
service:1",
    "containerDefinitions": [
        {
            "name": "demo-service",
            "image": "mcr.microsoft.com/dotnet/framework/samples:aspnetapp-
windowsservercore-ltsc2019",
            "cpu": 512,
            "memory": 512,
            "links": [],
            "portMappings": [
                {
                    "containerPort": 80,
                    "hostPort": 0,
                    "protocol": "tcp"
                }
            ],
```

Acelere la innovación

Pasar a los contenedores facilita la automatización de las etapas del ciclo de vida del desarrollo, que incluyen la creación, las pruebas y la implementación de aplicaciones. Si automatiza estos procesos, dará a sus equipos de desarrollo y operaciones más tiempo para centrarse en la innovación.

Reduzca el TCO

Pasar a los contenedores a menudo reduce la dependencia de las herramientas de administración de licencias y protección de terminales. Como los contenedores son unidades informáticas efímeras, puede automatizar y simplificar las tareas de administración, como la aplicación de parches, el escalado y las copias de seguridad y la restauración. Esto puede reducir el coste total de propiedad de la administración y el funcionamiento de las cargas de trabajo basadas en contenedores. Los contenedores son más eficientes en comparación con las máquinas virtuales porque permiten maximizar la ubicación de las aplicaciones para aumentar la utilización de los recursos de infraestructura de las aplicaciones.

Cierre la brecha de habilidades

AWS ofrece programas y jornadas de inmersión para capacitar a los equipos de desarrollo de clientes en materia de contenedores y DevOps tecnología. Esto incluye asesoramiento y capacitación prácticas.

Refactoriza a .NET 5+ y usa contenedores de Linux

Si bien puede reducir los costos trasladando sus aplicaciones de.NET Framework a contenedores, puede ahorrar aún más si refactoriza las aplicaciones de.NET heredadas para convertirlas en alternativas nativas de la nube. AWS

Elimine los costes de licencias

La refactorización de su aplicación de .NET Framework en Windows a .NET Core en Linux supone un ahorro de costes de aproximadamente un 45 por ciento.

Acceda a las mejoras más recientes

Al refactorizar sus aplicaciones de .NET Framework en Windows a .NET Core en Linux, tendrá acceso a las mejoras más recientes, como Graviton2. Graviton2 ofrece un precio de rendimiento un 40 por ciento superior al de instancias comparables.

Mejore la seguridad y el rendimiento

La refactorización de la aplicación de .NET Framework en Windows a .NET Core en contenedores de Linux mejora la seguridad y el rendimiento. Esto se debe a que obtiene los parches de seguridad más recientes, se beneficia del aislamiento de los contenedores y tiene acceso a nuevas funciones.

Utilice contenedores de Windows en lugar de ejecutar muchas aplicaciones en una sola instancia de IIS

Tenga en cuenta las siguientes ventajas de usar contenedores de Windows en lugar de ejecutar varias aplicaciones en una instancia de EC2 Windows con Internet Information Services (IIS):

- Seguridad: los contenedores proporcionan un nivel de seguridad listo para usar que no se logra mediante el aislamiento a nivel de IIS. Si un sitio web o una aplicación de IIS se ve comprometido, todos los demás sitios alojados quedan expuestos y son vulnerables. El escape de contenedores es poco frecuente y es una vulnerabilidad más difícil de explotar que hacerse con el control de un servidor a través de una vulnerabilidad web.
- Flexibilidad: la capacidad de ejecutar contenedores de forma aislada durante el proceso y disponer de su propia instancia permite disponer de opciones de red más detalladas. Los contenedores también ofrecen métodos de distribución complejos en muchos EC2 casos. Estas ventajas no se obtienen cuando se consolidan las aplicaciones en una sola instancia de IIS.
- Sobrecarga de administración: la indicación del nombre del servidor (SNI) genera una sobrecarga que requiere administración y automatización. Además, debe lidiar con las operaciones típicas

de administración del sistema operativo, como la aplicación de parches, la solución de problemas de BSOD (si no se implementa el escalado automático), la protección de puntos finales, etc. La configuración de los sitios de IIS de acuerdo con las mejores prácticas de seguridad es una actividad continua y que lleva mucho tiempo. Es posible que incluso necesite configurar niveles de confianza, lo que también aumenta la sobrecarga de administración. Los contenedores están diseñados para ser apátridas e inmutables. En última instancia, sus implementaciones son más rápidas, seguras y repetibles si en su lugar utiliza contenedores de Windows.

Pasos a seguir a continuación

Invertir en una infraestructura moderna para ejecutar sus cargas de trabajo heredadas aporta enormes beneficios a su organización. AWS Los servicios de contenedores facilitan la administración de la infraestructura subyacente, ya sea en las instalaciones o en la nube, para que pueda centrarse en la innovación y en las necesidades de su empresa. Casi el 80 por ciento de todos los contenedores de la nube funcionan en la AWS actualidad. AWS proporciona un amplio conjunto de servicios de contenedores para prácticamente todos los casos de uso. Para empezar, consulte Contenedores en AWS.

Recursos adicionales

- Optimice los costes de las cargas de trabajo en contenedores con los proveedores de capacidad de ECS y las instancias EC2 puntuales (AWS blog)
- Lista de verificación de optimización de costos para Amazon ECS y AWS Fargate (AWS blog)
- Amazon EKS en AWS Graviton2 está disponible de forma general: consideraciones sobre aplicaciones de arquitectura múltiple (Blog)AWS
- Optimización de costos para Kubernetes en (Blog) AWSAWS
- Optimización de los costes informáticos de Kubernetes con la consolidación de Karpenter (Blog)AWS

Optimice los costos de AWS Fargate las tareas en Amazon ECS

Descripción general

El dimensionamiento correcto de AWS Fargate las tareas es un paso importante para la optimización de costos. Con demasiada frecuencia, las aplicaciones se crean con tamaños arbitrarios para

las tareas de Fargate y nunca se revisan. Esto puede provocar un aprovisionamiento excesivo de las tareas de Fargate y gastos innecesarios. En esta sección, se muestra cómo ofrecer recomendaciones prácticas AWS Compute Optimizerpara optimizar la CPU y la memoria de las tareas para los servicios de Amazon Elastic Container Service (Amazon ECS) que se ejecutan en Fargate. Compute Optimizer también cuantifica el impacto en los costos de la adopción de estas recomendaciones. Esto le permite priorizar sus esfuerzos de optimización en función del tamaño de la oportunidad de ahorro. Las recomendaciones de Compute Optimizer proporcionan configuraciones de CPU y memoria a nivel de contenedor para tareas de reducción de tamaño.

Beneficios de costos

El tamaño correcto de las tareas de Amazon ECS en Fargate puede reducir los costos entre un 30 y un 70 por ciento en el caso de las tareas de ejecución prolongada. Sin revisar las métricas de rendimiento de las aplicaciones para dimensionar correctamente el tamaño de las tareas, puede aplicar la misma mentalidad que se utiliza en las instancias de EC2 cómputo al dimensionar los contenedores. Esto lleva a que las tareas de Fargate sobredimensionen y aumenten los costes de los recursos inactivos. Puedes usar Compute Optimizer para descubrir las oportunidades de dimensionamiento adecuadas de forma reactiva. Lo ideal es que el propietario de la aplicación revise las métricas de rendimiento específicas de la aplicación y elimine la sobrecarga del sistema operativo para garantizar que se especifique el tamaño de tarea adecuado. Para obtener más información, consulte la sección Mover aplicaciones de Windows a contenedores de esta guía.

Recomendaciones de optimización de costos

En esta sección, se ofrecen recomendaciones para usar Compute Optimizer para dimensionar correctamente las tareas de Amazon ECS en Fargate.

Como parte del proceso de optimización de costos, le recomendamos que haga lo siguiente:

- Activar Compute Optimizer
- Consume los resultados de Compute Optimizer
- Etiquete las tareas para que tengan el tamaño correcto
- Habilite la etiqueta de asignación de costos para que funcione con las herramientas AWS de facturación
- Implemente las recomendaciones de tamaño adecuadas
- Revise los costes de antes y después en Cost Explorer

Activar Compute Optimizer

Puede activarlo AWS Compute Optimizera nivel de organización o de cuenta única en AWS Organizations. La configuración de toda la organización proporciona informes continuos para las instancias nuevas y existentes en toda la flota y para todas las cuentas de los miembros. Esto permite que el dimensionamiento correcto sea una actividad recurrente en lugar de una actividad. point-in-time

Nivel de organización

Para la mayoría de las organizaciones, la forma más eficiente de usar Compute Optimizer es a nivel de la organización. Esto proporciona una visibilidad de la organización desde múltiples cuentas y regiones y centraliza los datos en una sola fuente para su revisión. Para habilitar esto a nivel de la organización, haga lo siguiente:

1. Inicie sesión en su cuenta de AWS Organizations administración con un rol que tenga los permisos necesarios y elija habilitar todas las cuentas de esta organización. Su organización debe tener habilitadas todas las características.
2. Después de habilitar la cuenta de administración, puedes iniciar sesión en ella, ver todas las cuentas de los demás miembros y consultar sus recomendaciones.

> **Note**
>
> Se recomienda configurar una cuenta de administrador delegado para Compute Optimizer. Esto te permite aplicar el principio de privilegios mínimos, lo que minimiza el acceso a la cuenta de AWS Organizations administración y, al mismo tiempo, proporciona acceso al servicio de toda la organización.

Nivel de cuenta único

Si te diriges a una cuenta con costes elevados pero a la que no tienes acceso AWS Organizations, puedes habilitar Compute Optimizer para esa cuenta y región. Para obtener más información sobre el proceso de suscripción, consulta Cómo empezar con. AWS Compute Optimizer

> **Note**
>
> Las recomendaciones se actualizan a diario y pueden tardar hasta 12 horas en generarse. Tenga en cuenta que Compute Optimizer requiere 24 horas de métricas en los últimos

14 días para generar recomendaciones para Amazon ECS en Fargate. Para obtener más información, consulte Requisitos para los servicios de Amazon ECS en Fargate en la documentación de Compute Optimizer.

Compute Optimizer analiza automáticamente las siguientes métricas de uso de CloudWatch Amazon y Amazon ECS para sus servicios de Amazon ECS en Fargate:

- `CPUUtilization`— El porcentaje de capacidad de la CPU que se utiliza en el servicio.
- `MemoryUtilization`— El porcentaje de memoria que se utiliza en el servicio.

Consume los resultados de Compute Optimizer

Considere un ejemplo que se centre en realizar los cambios de tamaño correctos en una sola cuenta y una sola región. En este ejemplo, Compute Optimizer está habilitado a nivel de organización en todas las cuentas. Ten en cuenta que el dimensionamiento correcto es un proceso disruptivo que, en la mayoría de los casos, los propietarios de las aplicaciones llevan a cabo con precisión durante un período de mantenimiento programado que dura varias semanas.

Si accedes a Compute Optimizer desde la cuenta de administración de una organización (como se muestra en los pasos siguientes), puedes elegir la cuenta que deseas investigar. En este ejemplo, una tarea se ejecuta en una sola cuenta que está sobreaprovisionada. `us-east-1` La atención se centra en cambiar el tamaño al tamaño recomendado para el servicio Amazon ECS.

1. Abre la consola Compute Optimizer.
2. En la página del panel de control, filtre por Findings=Over-provisioned para ver todos los servicios de Amazon ECS en Fargate.
3. Para revisar las recomendaciones detalladas sobre los servicios de ECS sobreaprovisionados en Fargate, desplázate hacia abajo y selecciona Ver recomendaciones.
4. Seleccione Exportar y guarde el archivo para usarlo en el futuro.

> **Note**
>
> Para guardar las recomendaciones para revisarlas en el futuro, debes tener un bucket de S3 disponible para que Compute Optimizer escriba en cada región. Para obtener más

información, consulte la política de bucket de Amazon S3 AWS Compute Optimizer en la documentación de Compute Optimizer.

Para ver las recomendaciones de Compute Optimizer, haz lo siguiente:

1. En la consola de Compute Optimizer, ve a la página de recomendaciones de exportación.
2. Para el destino del bucket de S3, elige tu bucket de S3.
3. En la sección Filtros de exportación, en Tipo de recurso, elija los servicios de ECS en Fargate.
4. En la página Recomendaciones para los servicios de ECS en Fargate, profundiza en uno de los servicios de ECS en Fargate y consulta las recomendaciones de CPU y memoria de Compute Optimizer. Por ejemplo, consulta las recomendaciones de las secciones Comparar la configuración actual con el tamaño recomendado de la tarea y Comparar la configuración actual con el tamaño de contenedor recomendado.

Para obtener la lista de servicios de ECS para Fargate que necesita ajustar al tamaño correcto, haga lo siguiente:

1. Abra la consola de Amazon S3.
2. En el panel de navegación, elija Buckets y, a continuación, elija el bucket al que exportó los resultados.
3. En la pestaña Objetos, selecciona tu objeto y elige Descargar.
4. En los resultados descargados, filtre la columna de búsqueda para mostrar solo los servicios OVER_PROVISIONED de Amazon ECS en Fargate. Aquí se muestran los servicios de Amazon ECS a los que planea dirigirse para obtener el tamaño correcto.
5. Guarde las definiciones de tareas en un editor de texto para utilizarlas más adelante.

Ajuste el tamaño correcto de las etiquetas para las tareas

Etiquetar tus cargas de trabajo es una herramienta poderosa para organizar tus recursos. AWS Puede usar etiquetas para obtener una visibilidad detallada de los costos y habilitar la devolución de cargos. Existen muchos métodos y estrategias para añadir etiquetas a los AWS recursos a fin de gestionar las devoluciones de gastos y la automatización. Para obtener más información, consulta el AWS documento técnico sobre las prácticas recomendadas para etiquetar los recursos. AWS

El siguiente ejemplo se utiliza AWS CloudShellpara etiquetar todas las tareas que forman parte de cualquier servicio de Amazon ECS dentro de la cuenta de destino y Región de AWS.

```
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="enabled"
# Get a list of ECS Clusters
ClustersArns=$( w secs list-clusters –query 'clusterArns' –output text)
for ClustersArn in $ClustersArns; do
 ServiceArns=$( w secs list-services –cluster $ClustersArn –query 'serviceArns' –output
 text)
 for ServiceArn in $ServiceArns; do
  TasksArns=$( w secs list-tasks –cluster $ClustersArn –service-name $ServiceArn –query
 'taskArns' –output text)
  for TasksArn in $TasksArns; do
    w secs tag-resource –resource-arn $TasksArn –tags key=$TAG_KEY,value=$TAG_VALUE
  done
 done
done
```

El siguiente ejemplo de código muestra cómo habilitar la propagación de etiquetas en todos los servicios de Amazon ECS.

```
#!/bin/bash
# Set variables
TAG_KEY="rightsizing"
TAG_VALUE="enabled"
# Get a list of ECS Clusters
ClustersArns=$(aws ecs list-clusters --query 'clusterArns' --output text)
for ClustersArn in $ClustersArns; do
 ServiceArns=$(aws ecs list-services --cluster $ClustersArn --query 'serviceArns' --
output text)
 for ServiceArn in $ServiceArns; do
  aws ecs update-service --cluster $ClustersArn --service $ServiceArn --propagate-tags
 SERVICE &>/dev/null
  aws ecs tag-resource --resource-arn $ServiceArn --tags key=$TAG_KEY,value=$TAG_VALUE
 done
done
```

Habilite la etiqueta de asignación de costos para que funcione con las herramientas AWS de facturación

Recomendamos activar la etiqueta de asignación de costes definida por el usuario. Esto permite reconocer y filtrar la etiqueta Rightsizing en las herramientas de AWS facturación (por ejemplo, y). AWS Cost Explorer AWS Cost and Usage Report Si no lo habilitas, la opción de filtrado de etiquetas y los datos no estarán disponibles. Para obtener información sobre el uso de etiquetas de asignación de costes, consulte Activar etiquetas de asignación de costes definidas por el usuario en la Administración de facturación y costos de AWS documentación.

Tras esperar 24 horas, podrá ver la etiqueta en Cost Explorer antes de implementar las recomendaciones de tamaño adecuadas en la siguiente sección. Para ello, busque la etiqueta Rightsizing en Cost Explorer.

Implemente las recomendaciones de tamaño adecuadas

Compute Optimizer proporcionará recomendaciones sobre el tamaño de las tareas o del contenedor. Para implementar las recomendaciones de tamaño correctas, haz lo siguiente.

1. Abra la consola de Amazon ECS.
2. En la barra de navegación, seleccione la Región que contiene la definición de tarea.
3. En el panel de navegación, elija Task Definitions (Definiciones de tareas).
4. En la página Task definitions (Definiciones de tareas), elija la tarea y, a continuación, elija Create new revision (Crear nueva revisión).
5. En la página Create new task definition revision (Crear nueva revisión de definición de tarea), realice cambios. Para actualizar la recomendación de tamaño del contenedor, actualice `cpu` y `memory` en el bloque ContainerDefinitions de la definición de la tarea de ECS. Por ejemplo:

```
"containerDefinitions": [
 {
  "name": "your-container-name",
  "image": "your-image",
  "cpu": 1024,
  "memory": 2048,
 }
],
```

6. Verifique la información y, luego, seleccione Create (Crear).

Para actualizar el servicio Amazon ECS, haga lo siguiente:

1. Abra la consola de Amazon ECS.
2. En la página Clusters (Clústeres), seleccione el clúster.
3. En la página Cluster overview (Información general del clúster), seleccione el servicio y, a continuación, elija Update (Actualizar).
4. En Task definition (Definición de tareas), elija la familia y la revisión de definiciones de tareas que se va a utilizar.

Para los operadores avanzados, puede utilizar CloudShell para actualizar el servicio Amazon ECS. Por ejemplo:

```
bash
#!/bin/bash
# Set variables
ClustersName="workshop-cluster"
ServiceName="lab7-fargate-service"
TaskDefinition="lab7-fargate-demo:3"
# update the service
aws ecs update-service --cluster $ClustersName --service $ServiceName --task-definition
 $TaskDefinition
```

Revise los costos anteriores y posteriores

Una vez que haya dimensionado correctamente sus recursos, puede usar el Explorador de costos para mostrar los costos anteriores y posteriores mediante la etiqueta Rightsizing. Recuerde que puede usar etiquetas de recursos para realizar un seguimiento de los costos. Al usar varias capas de etiquetas, puede lograr una visibilidad detallada de sus costos. En el ejemplo que se describe en esta guía, la etiqueta Rightsizing se utiliza para aplicar una etiqueta genérica a todas las instancias objetivo. A continuación, se utiliza una etiqueta de equipo para organizar aún más los recursos. El siguiente paso consiste en introducir etiquetas de aplicación para mostrar mejor el impacto en los costes de funcionamiento de una aplicación específica.

Considere un ejemplo de la reducción de costos que se puede lograr al usar la etiqueta Rightsizing para un solo nivel de cuenta. En este ejemplo, los costos operativos van de 30,26$ diarios a 7,56$ diarios. Suponiendo que haya 744 horas al mes, el coste anual, antes del dimensionamiento correcto, es de 11.044,9$. Con el tamaño correcto, el costo anual se reduce a 2.759,4 dólares. Esto

se traduce en una disminución del 75 por ciento en los costos de cómputo de esta cuenta. Imagine el impacto de esto en una organización grande.

Antes de embarcarse en la elección del tamaño correcto, tenga en cuenta lo siguiente:

- AWS ofrece muchas opciones para reducir los costos. Esto incluye AWS OLA, en el que se AWS revisan las instancias locales antes de mudarlas a AWS ellas. El AWS OLA también le proporciona recomendaciones sobre el tamaño correcto y orientación sobre licencias.
- Complete todas las tallas correctas antes de comprar Savings Plans. Esto puede ayudarlo a evitar compras excesivas en su compromiso con Savings Plans.

Pasos a seguir a continuación

Recomendamos los siguientes pasos:

1. Revise su panorama actual y considere la posibilidad de convertir los volúmenes gp2 de Amazon EBS en volúmenes gp3.
2. Revise Savings Plans.

Recursos adicionales

- Introducción a Compute Optimizer (documentación)AWS
- Prácticas recomendadas para etiquetar AWS recursos (documentos técnicos)AWS
- Windows Containers está activado AWS (AWS Workshop Studio)

Obtenga visibilidad de sus costos de Amazon EKS

Descripción general

Se necesita una visión holística para monitorear de manera efectiva el costo de una implementación de Kubernetes. El único coste fijo y conocido corresponde al plano de control de Amazon Elastic Kubernetes Service (Amazon EKS). Esto incluye todos los demás componentes que componen la implementación, desde el cómputo y el almacenamiento hasta las redes, y se trata de un importe variable en función de las necesidades de la aplicación.

Puedes usar Kubecost para analizar el costo de tu infraestructura de Kubernetes, desde los espacios de nombres y los servicios hasta los pods individuales, y luego mostrar los datos en un panel de

control. Kubecost muestra los costos dentro del clúster, como el cómputo y el almacenamiento, y los costos, out-of-cluster como los depósitos de Amazon Simple Storage Service (Amazon S3) y las instancias del Amazon Relational Database Service (Amazon RDS). Kubecost recomendará el tamaño correcto en función de estos datos y mostrará las alertas críticas que puedan afectar al sistema. Kubecost se puede integrar con AWS Cost and Usage ReportCompute Savings Plans, Reserved Instances y otros programas de descuentos para mostrar los ahorros.

Costos y beneficios

Kubecost proporciona informes y paneles que visualizan el costo de las implementaciones de Amazon EKS. Le permite desglosar desde el clúster cada uno de los distintos componentes, como los controladores, los servicios, los nodos, los pods y los volúmenes. Esto le proporciona una visión holística de las aplicaciones que se ejecutan en un entorno Amazon EKS. Al habilitar esta visibilidad, puede seguir las recomendaciones de Kubecost o ver los costos de cada aplicación de forma pormenorizada. El tamaño correcto de un grupo de nodos de Amazon EKS ofrece los mismos ahorros potenciales que EC2 las instancias estándar. Si puede dimensionar correctamente sus contenedores y nodos, puede eliminar la sobrecarga de cómputo del tamaño de la instancia necesaria para ejecutar el contenedor y la cantidad de EC2 instancias requeridas en el grupo de escalado automático.

Recomendaciones de optimización de costos

Para aprovechar Kubecost, te recomendamos que hagas lo siguiente:

1. Implemente Kubecost en su entorno
2. Obtenga un desglose detallado de los costos de las aplicaciones de Windows
3. Nodos de clúster del tamaño correcto
4. Solicitudes de contenedores del tamaño correcto
5. Gestione los nodos infrautilizados
6. Corrija las cargas de trabajo abandonadas
7. Actuar en función de las recomendaciones
8. Actualice los nodos autogestionados

Implemente Kubecost en su entorno

El taller Finhack de Amazon EKS le enseña cómo implementar un entorno de Amazon EKS configurado para usar Kubecost en una AWS cuenta propia. Esto le permite adquirir experiencia

práctica con la tecnología. Si está interesado en organizar este taller en su organización, póngase en contacto con su equipo de cuentas.

Para implementar Kubecost en su clúster de Amazon EKS con Helm, consulte la publicación en el AWS blog sobre la colaboración con Kubecost para ofrecer un monitoreo de costos a los clientes de EKS. AWS Como alternativa, puedes consultar la documentación oficial de Kubecost para obtener instrucciones sobre cómo instalar y configurar Kubecost. Para obtener información sobre el soporte de Kubecost para los nodos de Windows, consulte Soporte de nodos de Windows en la documentación de Kubecost.

Obtenga un desglose detallado de los costos de las aplicaciones de Windows

Si bien puede conseguir importantes ahorros de costes mediante el uso de Amazon EC2 Spot Instances, también puede beneficiarse del hecho de que las cargas de trabajo de Windows suelen estar llenas de estado. El uso de instancias puntuales depende de la aplicación y le recomendamos que compruebe si son aplicables a su caso de uso.

Para obtener un desglose detallado de los costos de sus aplicaciones de Windows, inicie sesión en Kubecost. En la página de navegación, selecciona Ahorros.

Nodos de clúster del tamaño correcto

En Kubecost, selecciona Ahorros en la barra de navegación y, a continuación, selecciona el tamaño correcto del nodo del clúster.

Considera un ejemplo en el que Kubecost informa que el clúster está sobreaprovisionado tanto en términos de vCPU como de RAM. En la siguiente tabla se muestran los detalles y las recomendaciones de Kubecost.

	Actuales	Recomendación: Sencillo	Recomendación: compleja
Recuento total	3462,57 dólares al mes	137,24 dólares estadounidenses al mes	303,68 dólares estadounidenses al mes
Recuento de nodos	4	5	4
CPU	74 VCPUs	10 VCPUs	8 VCPUs

	Actuales	Recomendación: Sencillo	Recomendación: compleja
RAM	152 GB	20 GB	18 GB
Desglose de instancias	2 c5.xlarge + 2 más	5 t3a.medianas	2 c5n.large + 1 más

Como se describe en la entrada del blog de Kubecost: Encuentre un conjunto óptimo de nodos para un clúster de Kubernetes, la opción simple utiliza un solo grupo de nodos, mientras que la compleja utiliza un enfoque de grupo de varios nodos. El botón Aprenda a adoptar permite cambiar el tamaño del clúster con un solo clic. Requiere la instalación del controlador de clúster Kubecost.

Si utilizas nodos de Windows autogestionados que no han sido creados por eksctl, consulta Actualizar un grupo de nodos autogestionado existente. Estas instrucciones le muestran cómo cambiar el tipo de instancia en la plantilla de EC2 lanzamiento de Amazon utilizada por el grupo Auto Scaling.

Solicitudes de contenedores del tamaño correcto

En Kubecost, selecciona Ahorros en la barra de navegación y, a continuación, ve a la página Solicitar recomendaciones sobre el tamaño correcto. En esta página se muestra la eficiencia de los módulos, las recomendaciones sobre el tamaño correcto y el ahorro de costes estimado. Puede usar el botón Personalizar para filtrar por clúster, nodo, espacio de nombres o controlador, etc.

Como ejemplo, considere que Kubecost calculó que algunos de sus pods están sobreaprovisionados en términos de CPU y RAM (memoria). A continuación, Kubecost recomienda que te adaptes a los nuevos valores de CPU y RAM para conseguir el ahorro mensual estimado. Para cambiar los valores de CPU y RAM, debes actualizar el archivo de manifiesto de despliegue.

Administre los nodos infrautilizados

En Kubecost, selecciona Ahorros en la barra de navegación y, a continuación, selecciona Administrar nodos infrautilizados.

Considere un ejemplo en el que la página muestre que un nodo del clúster está infrautilizado en términos de CPU y RAM (memoria) y, por lo tanto, se puede agotar y terminar o cambiar de tamaño. Si eliges los nodos que no pasan las comprobaciones de nodos y pods, obtendrás más información sobre los motivos por los que no se pueden drenar.

Corrija las cargas de trabajo abandonadas

En Kubecost, selecciona Ahorros en la barra de navegación y, a continuación, selecciona la página Cargas de trabajo abandonadas. En este ejemplo, filtra por un espacio de nombres denominado ventanas. En esta página se muestran los pods que no han alcanzado el umbral de tráfico y que se consideran abandonados. Los pods deben enviar o recibir cierta cantidad de tráfico de red durante el período definido.

Tras considerar detenidamente el abandono de uno o más pods, puede ahorrar costes reduciendo el número de réplicas, eliminando la implementación, redimensionándola para que consuma menos recursos o notificando al propietario de la aplicación que cree que la implementación está abandonada.

Siga las recomendaciones

En la sección sobre el tamaño correcto de los nodos del clúster, Kubecost analiza el uso de los nodos de trabajo del clúster y hace recomendaciones sobre el tamaño correcto de los nodos para reducir los costes. Hay dos tipos de grupos de nodos que se pueden usar con Amazon EKS: autogestionados y gestionados.

Actualice los nodos autogestionados

Para obtener información sobre la actualización de los nodos autogestionados, consulte Actualizaciones de nodos autogestionados en la documentación de Amazon EKS. Afirma que los grupos de nodos creados con no se `eksctl` pueden actualizar y deben migrarse a un nuevo grupo de nodos con la nueva configuración.

Por ejemplo, supongamos que tiene un grupo de nodos de Windows llamado `ng-windows-m5-2xlarge` (que usa una EC2 instancia m5.2xlarge) y desea migrar los pods a un nuevo grupo de nodos llamado `ng-windows-t3-large` (que está respaldado por una EC2 instancia t3.large para ahorrar costos).

Para migrar a un nuevo grupo de nodos cuando utilice grupos de nodos implementados por`eksctl`, haga lo siguiente:

1. Para buscar el nodo en el que se encuentra actualmente el pod, ejecute el `kubectl describe pod <pod_name> -n <namespace>` comando.
2. Ejecute el comando `kubectl describe node <node_name>`. El resultado muestra que el nodo se ejecuta en una instancia m5.2xlarge. También coincide con el nombre del grupo de nodos (). `ng-windows-m5-2xlarge`

3. Para cambiar la implementación para usar un grupo de nodosng-windows-t3-large, elimine el grupo de nodos ng-windows-m5-2xlarge y ejecutekubectl describe svc,deploy,pod -n windows. La implementación comienza a redesplegarse inmediatamente ahora que se ha eliminado su grupo de nodos.

> **Note**
>
> El servicio se interrumpirá cuando elimines el grupo de nodos.

4. Vuelva a ejecutar el kubectl describe svc,deploy,pod -n windows comando después de unos minutos. El resultado muestra que todos los pods están nuevamente en estado de ejecución.
5. Para mostrar que los pods se están ejecutando ahora en el grupo de nodosng-windows-t3-large, vuelve a ejecutar los kubectl describe node <node_name> comandos kubectl describe pod <pod_name> -n <namespace> y.

Métodos de redimensionamiento alternativos

Este método se aplica a cualquier combinación de grupos de nodos autogestionados o gestionados. La entrada del blog sobre cómo migrar sin problemas las cargas de trabajo de un grupo de nodos autogestionado por EKS a grupos de nodos gestionados por EKS proporciona orientación sobre cómo migrar las cargas de trabajo de un grupo de nodos con un tipo de instancia sobredimensionado a un grupo de nodos con el tamaño adecuado sin ningún tiempo de inactividad.

Pasos a seguir a continuación

Kubecost facilita la visualización del coste de sus entornos de Amazon EKS. La profunda integración de Kubecost con Kubernetes AWS APIs puede ayudarlo a encontrar posibles ahorros de costos. Puedes verlas como recomendaciones en el panel de ahorros de Kubecost. Kubecost también puede implementar algunas de estas recomendaciones a través de su función de controlador de clúster.

Te recomendamos que revises la step-by-step implementación y que Kubecost AWS colabore para ofrecer un seguimiento de los costes a los clientes de EKS, una entrada del blog de Containers. AWS

Recursos adicionales

- Taller de Amazon EKS (taller de Amazon EKS)

- AWS y Kubecost colaboran para ofrecer un seguimiento de los costes a los clientes de EKS (Blog)AWS
- Taller Amazon EKS Finhack (AWS Workshop Studio)
- Windows Containers está activado AWS (AWS Workshop Studio)

Cambie la plataforma de las aplicaciones de Windows con App2Container

Descripción general

AWS App2Containeres una herramienta de línea de comandos para migrar y modernizar aplicaciones web de Java y .NET a contenedores. App2Container analiza y crea un inventario de todas las aplicaciones que se ejecutan en máquinas virtuales, instancias de Amazon Elastic Compute Cloud EC2 (Amazon) o en otros proveedores de nube. Selecciona la aplicación que deseas almacenar en contenedores. App2Container empaqueta los artefactos y las dependencias de la aplicación en imágenes de contenedores, configura los puertos de red y genera los artefactos de despliegue necesarios de Amazon Elastic Container Service (Amazon ECS) y Amazon Elastic Kubernetes Service (Amazon EKS), que son plantillas de infraestructura como código (IaC). App2Container aprovisiona la infraestructura de nube y los canales de CI y CD necesarios para implementar la aplicación contenerizada en un entorno de producción. Para obtener más información, consulte Cómo funciona App2Container en la documentación de App2Container.

Con App2Container, puede migrar sus aplicaciones como contenedores AWS y modernizarlas y, al mismo tiempo, estandarizar la implementación y las operaciones de sus aplicaciones. Puede usar App2Container para crear rápidamente una prueba de concepto (PoC) o acelerar el despliegue de cargas de trabajo de producción en contenedores.

Hay varios aspectos que se deben tener en cuenta al trabajar con aplicaciones de Windows. App2Container admite la contenerización de aplicaciones ASP.NET implementadas en Microsoft Internet Information Services (IIS), incluidas las aplicaciones Windows Communication Foundation (WCF) alojadas en IIS que se ejecutan en Windows Server 2016, Windows Server 2019 o Windows Server Core 2004. Para obtener más información, consulte Aplicaciones compatibles para Windows en la documentación de App2Container. App2Container utiliza Windows Server Core como imagen base para sus artefactos de contenedor, haciendo coincidir la versión de contenedor de Windows Server Core con la versión del sistema operativo (SO) del servidor en el que se ejecutan los

comandos de contenedorización. Este enfoque desacopla la aplicación del sistema operativo subyacente para que pueda actualizar el sistema operativo sin realizar una migración tradicional.

Si utiliza una máquina de trabajo para alojar la aplicación en contenedores, la imagen base del contenedor, como el canal de servicio a largo plazo (LTSC) de Windows Server 2019, coincide con el sistema operativo de la máquina de trabajo, como Windows Server 2019. Si ejecuta la contenedorización directamente en los servidores de aplicaciones, la versión coincide con el sistema operativo del servidor de aplicaciones. Si sus aplicaciones se ejecutan en Windows Server 2008 o 2012 R2, puede seguir utilizando App2Container configurando una máquina de trabajo para los pasos de contenedorización e implementación. App2Container no admite aplicaciones que se ejecuten en sistemas operativos cliente Windows, como Windows 7 o Windows 10. App2Container es compatible con los marcos Tomcat, ToMEE y JBoss (en modo independiente) para procesos de Java. Para obtener más información, consulte Compatibilidad entre App2Container.

Costos y beneficios

La contenedorización y la consolidación de las aplicaciones pueden suponer hasta un 60% de ahorro informático en comparación con un patrón de diseño de despliegue one-application-to-one de servidores. App2Container ayuda a agilizar el proceso de contenerización de las aplicaciones. Los siguientes son algunos de los beneficios de usar App2Container para sus necesidades de modernización:

- App2Container se ofrece sin coste adicional.
- App2Container admite múltiples aplicaciones en una imagen de contenedor.
- Utilice App2Container para trasladar sus aplicaciones antiguas de.NET a contenedores para hacer frente a los sistemas operativos que se aproximan a la finalización del soporte. Puede pasarse a un sistema operativo más nuevo, evitar tener que pagar por un soporte adicional y reducir los riesgos de seguridad.
- Los contenedores son un método eficaz y rentable de empaquetar sus aplicaciones.NET. Consulte las ventajas de los contenedores en la recomendación de la MACO: Pasar a los contenedores.
- La consolidación de aplicaciones y la contenedorización ayudan a reducir el consumo de recursos informáticos, de almacenamiento y de licencias al utilizar los recursos informáticos de forma más eficiente.
- Pasar a los contenedores puede reducir los gastos operativos y los costes de infraestructura, así como aumentar la portabilidad del desarrollo y la agilidad de la implementación.

Recomendaciones de optimización de costos

Para obtener instrucciones sobre cómo usar App2Container, consulte Primeros pasos con. AWS App2ContainerPara obtener información sobre los comandos de App2Container, consulte la referencia de comandos de App2Container.

Pasos a seguir a continuación

App2Container puede acelerar el proceso de contenedorización de las aplicaciones y su implementación en Amazon EKS o Amazon ECS. La implementación de aplicaciones en contenedores reduce los costos de computación, redes y almacenamiento y reduce los gastos operativos para los operadores de aplicaciones.

Para obtener una experiencia práctica con App2Container, consulte el taller sobre cómo modernizar con App2Container. AWS App2Container Si quieres tener una experiencia de aprendizaje profunda, pide a tu equipo de AWS cuentas que organice un día de inmersión en App2Container.

Recursos adicionales

- Contenerización de aplicaciones Windows complejas de varios niveles mediante (entrada del blog) AWS App2ContainerAWS
- Contenerización de aplicaciones ASP.NET antiguas mediante (entrada de blog) AWS App2ContainerAWS
- Aplicaciones compatibles con App2Container (documentación)AWS
- Modernice con AWS App2Container Workshop (Workshop Studio)AWS
- AWS App2Container FAQs(AWS sitio web)

Almacenamiento

Elegir el almacenamiento adecuado para sus cargas de trabajo de Microsoft es una decisión arquitectónica fundamental. Como parte del proceso de toma de decisiones, le recomendamos que desarrolle un plan de almacenamiento y determine los requisitos funcionales de sus aplicaciones y servicios. En este capítulo se proporciona una descripción general de las siguientes opciones de almacenamiento que podrían tenerse en cuenta en su planificación.

Secciones:

- Amazon EBS
- Amazon FSx
- AWS Storage Gateway

Amazon EBS

Amazon Elastic Block Store (Amazon EBS) es un servicio de almacenamiento en bloques totalmente gestionado que le permite almacenar volúmenes de almacenamiento persistentes a nivel de bloque que puede utilizar con instancias de Amazon Elastic Compute Cloud (Amazon). EC2 Puede aprovechar varias funciones de Amazon EBS para gestionar y optimizar eficazmente los recursos de almacenamiento para las cargas de trabajo de Windows en la nube. Por ejemplo, puede usar Amazon EBS para aprovisionar la cantidad exacta de IOPS y el rendimiento que necesita para su carga de trabajo, seleccionar entre una variedad de tipos de volumen que se adapten a sus requisitos de carga de trabajo y utilizar herramientas para identificar y eliminar los recursos de almacenamiento desperdiciados. Este control pormenorizado del rendimiento y el uso del almacenamiento le ayuda a optimizar sus recursos de almacenamiento y, al mismo tiempo, a evitar costes innecesarios.

Esta sección abarca los siguientes temas:

- Migre los volúmenes de Amazon EBS de gp2 a gp3
- Modifique las instantáneas de Amazon EBS
- Eliminar volúmenes de Amazon EBS no adjuntos

Migre los volúmenes de Amazon EBS de gp2 a gp3

Descripción general

Una unidad de estado sólido (SSD) es la opción de almacenamiento estándar para cargas de trabajo de producción y de alto rendimiento. Amazon EBS ofrece un volumen SSD de uso general para cargas de trabajo de rendimiento medio a alto. El estándar en muchos Servicios de AWS (incluido Amazon EC2) es gp2, la segunda generación de estos volúmenes SSD de uso general. La tercera generación de dispositivos de uso general SSDs, denominada gp3, se lanzó en diciembre de 2020.

La gama gp3 introdujo mejoras significativas en los aspectos de personalización del rendimiento con respecto a la generación anterior. En el caso de los volúmenes gp2 de Amazon EBS, el rendimiento está estrechamente relacionado con el tamaño del volumen. Por cada 1 GB de capacidad, los volúmenes gp2 obtienen un rendimiento de 3 IOPS. Es decir, un volumen gp2 de 2000 GB es capaz de generar 6000 IOPS. En el caso de los volúmenes gp3, el rendimiento se puede personalizar independientemente de la capacidad de almacenamiento. Esto permite que incluso los volúmenes de pequeña capacidad alcancen capacidades de rendimiento de hasta 16 000 IOPS y un rendimiento de 1000 MB/s.

Otro cambio importante de los volúmenes gp3 es el rendimiento básico de las IOPS. Los volúmenes de gp3 comienzan con 3000 IOPS. En comparación, los volúmenes de gp2 deben alcanzar un tamaño de 1 TiB antes de alcanzar la misma capacidad de rendimiento. Para Windows Server, que normalmente tiene una unidad C: mucho más pequeña que 1 TiB, la actualización de gp2 a gp3 supone una mejora significativa del rendimiento.

Por último, el precio de los volúmenes gp3 es una de las mayores mejoras en comparación con los volúmenes gp2. Los volúmenes gp3 ofrecen capacidades de rendimiento mejoradas a un coste un 20 por ciento inferior al de los volúmenes gp2.

Impacto del costo

Con la capacidad de escalar el rendimiento de forma independiente de la capacidad, es importante comprender los aspectos relacionados con los precios que conlleva añadir IOPS y rendimiento adicionales. Para los volúmenes de gp2, el precio se basa en la capacidad aprovisionada a 0,10 USD por GiB al mes. En el caso de los volúmenes gp3, el precio es similar al de los volúmenes SSD de IOPS aprovisionados de alto rendimiento, que tienen un coste por capacidad y otro por IOPS y rendimiento adicionales.

Como se indica en la siguiente tabla, los volúmenes de gp3 tienen un precio de capacidad de 0,08 USD por GiB al mes (un 20 por ciento menos que el de gp2) y unos costes independientes de IOPS

de 0,005 USD por mes aprovisionadas (más de 3000 000) y 0,04 USD por mes aprovisionado de más de 125 USD por rendimiento. MiBs MiBs

	gp3	gp2
Tamaño del volumen	1 GiB – 16 TiB	1 GiB – 16 TiB
Las IOPS de referencia	3000	3 IOPs/GiB (mínimo 100 IOPS) hasta un máximo de 16 000 IOPS Los volúmenes de menos de 1 TiB pueden alcanzar velocidad es de hasta 3000 IOPS
Máximo de IOPS por volumen	16,000	16,000
Rendimiento de referencia	125 MiBs	El límite de rendimiento está entre 128 y 250 MiBs MiBs, según el tamaño del volumen
Rendimiento máximo por volumen	1000 MiBs	250 MiBs
Precio	0,08 dólares/Gib-mes 3000 IOPS gratuitas y 0,005 USD por IOPS aprovisionadas al mes o más de 3000 125$ MiBs gratuitas y MiBs 0,04$ aprovisionadas al mes (más de 125$ MiBs	0,10 USD/GiB al mes

> **Important**
>
> Aunque los volúmenes gp3 tienen costes de capacidad y rendimiento separados, los volúmenes gp3 siempre son más baratos que los volúmenes gp2 si están configurados con los mismos niveles de rendimiento.

En las siguientes tablas se muestran ejemplos de los ahorros de costos que se pueden lograr al convertir los volúmenes gp2 en gp3 con distintas configuraciones de capacidad y rendimiento.

Ejemplo de configuración gp2

Tamaño del volumen (GiB)	IOPS máximos	Rendimiento () MiBs	Coste (USD/mes)
30	3 000	128	3,00$
100	3 000	128	10,00 DÓLARES
500	3 000	250	50,00 DÓLARES
1 000	3 000	250	100,00 DÓLARES
2000	6000	250	200,00 DÓLARES
6000	16 000	250	600,00 DÓLARES

Ejemplo de configuración gp3 (línea base)

IOPS máximas	Rendimiento () MiBs	Coste (USD/mes)	Reducción de costes (en comparación con gp2)
3 000	125	2,40$	20%
3 000	125	8,00 DÓLARES	20%
3 000	125	40,00 DÓLARES	20%

IOPS máximas	Rendimiento () MiBs	Coste (USD/mes)	Reducción de costes (en comparación con gp2)
3 000	125	80,00 DÓLARES	20%
3 000	125	160,00 DÓLARES	20%
3 000	125	480,00 DÓLARES	20%

Ejemplo de configuración gp3 (coincidencia con gp2)

IOPS máximos	Rendimiento () MiBs	Coste (USD/mes)	Reducción de costes (en comparación con gp2)
3 000	128	2,52$	16%
3 000	128	8,12 DÓLARES	19%
3 000	250	45,00 DÓLARES	10%
3 000	250	85,00 DÓLARES	15%
6000	250	180,00 DÓLARES	10%
16 000	250	550,00 DÓLARES	8%

Para obtener un análisis de costos, consulte la sección sobre la calculadora de ahorro de costos de la migración de la gp2 a la gp3 de EBS en el recurso de Amazon EBS. Puede descargar la calculadora y utilizarla para averiguar cuánto puede ahorrar al migrar sus volúmenes de gp2 a gp3.

Recomendaciones de optimización de costos

Para obtener instrucciones sobre cómo completar el proceso de migración, consulte la publicación Migrar sus volúmenes de Amazon EBS de gp2 a gp3 y ahorrar hasta un 20% en costes en el AWS blog sobre almacenamiento.

Recursos adicionales

- Migre sus volúmenes de Amazon EBS de gp2 a gp3 y ahorre hasta un 20% en costes (AWS blog de almacenamiento)
- Cree una regla AWS Config personalizada para optimizar los tipos de volúmenes de Amazon EBS (blog de operaciones y migraciones AWS en la nube)
- Controle sus AWS costes mediante la eliminación de los volúmenes de Amazon EBS no utilizados (blog de operaciones y migraciones en la AWS nube)
- Utilidad de migración de Amazon EBS () GitHub
- Cómo ahorrar con los anuncios de re:Invent de 2020 (AWS Cloud Financial Management)
- Taller de optimización de costos (AWS Well-Architected Labs)
- Calculadora de ahorro de costes por la migración de gp2 a gp3 (descargar)

Modifique las instantáneas de Amazon EBS

Descripción general

Eliminar los volúmenes de EBS y gestionar la retención y el archivado de las instantáneas es un aspecto importante para controlar los costes desde el principio. Puede realizar copias de seguridad de los datos de sus volúmenes de EBS en Amazon Simple Storage Service (Amazon S3) realizando instantáneas. point-in-time Las instantáneas son copias de seguridad incrementales, por lo que solo guardan los bloques en los dispositivos que se modificaron después de la instantánea más reciente. Esto disminuye el tiempo necesario para crearlo y ahorra costos de almacenamiento, ya que no se duplican los datos. Cada instantánea contiene toda la información necesaria para restaurar los datos (desde el momento en que se creó la instantánea) en un nuevo volumen de EBS.

Los cargos por las instantáneas de EBS se calculan por gigabytes mensuales. Se le facturará según el tamaño de la instantánea y el tiempo que la conserve. El precio varía según el nivel de almacenamiento. En el caso del nivel Estándar, solo se facturarán los bloques modificados que se almacenen. En el nivel de archivo, se te facturarán todos los bloques de instantáneas que estén almacenados. También se le facturará por la recuperación de instantáneas del nivel de archivo. Los siguientes son ejemplos de escenarios para cada nivel de almacenamiento:

- Nivel estándar: tiene un volumen que almacena 100 GB de datos. Se le facturarán los 100 GB completos de datos de la primera instantánea (instantánea A). En el momento de la siguiente

instantánea (instantánea B), dispondrá de 105 GB de datos. En ese caso, solo se le facturarán los 5 GB de almacenamiento adicionales para la instantánea B incremental.

- Nivel de archivado: se archiva la instantánea B. A continuación, la instantánea se mueve al nivel de archivo y se le factura el bloque completo de instantáneas de 105 GB.

Puede usar Amazon Data Lifecycle Manager como ayuda para configurar un ciclo de vida que le permita conservar y administrar las instantáneas según lo programado.

Impacto del costo

Los cargos por los volúmenes e instantáneas de EBS se gestionan por separado. Las instantáneas de EBS se facturan a una tarifa inferior a la de los volúmenes de EBS activos. Cuando una instancia finaliza, el valor del DeleteOnTermination atributo de cada volumen de EBS adjunto determina si se debe conservar o eliminar el volumen. De forma predeterminada, el `DeleteOnTermination` atributo se establece en `True` para el volumen raíz. Está configurado `False` para todos los demás tipos de volumen. Esto crea situaciones en las que el operador tiene la intención de eliminar una EC2 instancia, pero deja atrás los volúmenes que se agregaron a la instancia además del volumen raíz. Para obtener instrucciones sobre cómo comprobar los volúmenes (y sus instantáneas asociadas) que ya no necesita, consulte Ver información sobre un volumen de Amazon EBS en la documentación de Amazon EBS.

De forma predeterminada, cuando crea una instantánea, se almacena en el nivel Amazon EBS Snapshot Standard (nivel estándar). Las instantáneas almacenadas en el nivel estándar son progresivas. Eso significa que solo los bloques del volumen que han cambiado después de la última instantánea se guardan. El Amazon EBS Snapshots Archive es un nuevo nivel de almacenamiento que puede utilizar para almacenar a largo plazo y bajo coste las instantáneas a las que se accede con poca frecuencia y que no requieren una recuperación rápida o frecuente. La diferencia de precio entre el estándar y el de archivo es significativa y debería ser un factor clave a tener en cuenta a la hora de configurar su estrategia de instantáneas. El archivo de instantáneas de Amazon EBS ofrece una reducción de hasta el 75 % menos de los costos de almacenamiento de instantáneas para las instantáneas que planea almacenar durante 90 días o más y a las que rara vez necesita acceder.

Almacenamiento de instantáneas en Amazon EBS	Costo
Estándar	0,05 USD por GB al mes

Almacenamiento de instantáneas en Amazon EBS	Costo
Archivado	0,0125 dólares/GB/mes

En entornos más pequeños, es posible que el ahorro de costes no sea significativo. Los ahorros son más significativos a gran escala, cuando hay varias cuentas y miles de EC2 instancias con instantáneas TBs de EBS almacenadas incluso cuando se han eliminado los volúmenes de EBS.

En la siguiente tabla se comparan los niveles estándar y de archivado por mes con un uso de tan solo 50 TB. Incluso a esta escala más baja, se siguen ahorrando miles de dólares al año.

Almacenamiento de instantáneas en Amazon EBS	Coste por mes	Coste por año
50 TB estándar	312,50 DÓLARES	3.750 DÓLARES
Archivo: 50 TB	78,13$	937,60 DÓLARES
	Ahorros anuales	2.812,40$

Recomendaciones de optimización de costos

Eliminar una instantánea no reduce necesariamente los costos de almacenamiento de datos de la organización. Es posible que otras instantáneas hagan referencia a los datos de esta, y los datos a los que se hace referencia se conservan siempre. Por ejemplo, al tomar la primera snapshot de un volumen con 10 GiB de datos, el tamaño de la snapshot también es de 10 GiB. Como las instantáneas son incrementales, la segunda instantánea que tome del mismo volumen contiene solamente los bloques de datos que hayan cambiado desde la primera instantánea. Además, la segunda snapshot hace referencia a los datos de la primera snapshot. Si cambia 4 GiB de datos y toma una segunda instantánea, el tamaño de la segunda instantánea es de 4 GiB. Asimismo, la segunda snapshot hace referencia a los 6 GiB que no han cambiado de la primera snapshot. Para obtener más información, consulte ¿Por qué mis costes de almacenamiento no se redujeron después de eliminar una instantánea de mi volumen de EBS y, a continuación, eliminar el volumen propiamente dicho? en el AWS Knowledge Center.

Considere lo siguiente:

- No se te facturan las instantáneas que otra persona Cuenta de AWS posea y comparta con tu cuenta. Solo se te facturará cuando copies la instantánea compartida en tu cuenta. También se le facturarán los volúmenes de EBS que cree a partir de la instantánea compartida.
- Si otra instantánea (instantánea B) hace referencia a una instantánea (instantánea A), es posible que eliminar la instantánea B no reduzca los costes de almacenamiento. Al eliminar una instantánea, solo se eliminan los datos exclusivos de esa instantánea. Los datos a los que hacen referencia otras instantáneas permanecen y se le facturan estos datos referenciados. Para eliminar una instantánea incremental, consulte Eliminación incremental de instantáneas en la documentación de Amazon EBS.

La limpieza de las instantáneas es una práctica operativa estándar cuando se ejecutan cargas de trabajo. AWS Con el tiempo, las instantáneas pueden suponer costosos gastos por datos que no necesita.

Recursos adicionales

- Controle sus AWS costes mediante la eliminación de los volúmenes de Amazon EBS no utilizados (blog de operaciones y migraciones en la AWS nube)
- Eliminar una instantánea de Amazon EBS (documentación de Amazon EBS)
- Taller de optimización de costos (AWS Well-Architected Labs)
- Archive automáticamente las instantáneas de Amazon EBS con Amazon Data Lifecycle Manager (blog sobre AWS almacenamiento)

Eliminar volúmenes de Amazon EBS no adjuntos

Descripción general

Los volúmenes de EBS independientes (huérfanos) pueden generar costes de almacenamiento innecesarios en su entorno. AWS Es esencial incorporar la revisión y la eliminación periódicas de los volúmenes de EBS no utilizados y no utilizados como parte de la higiene del entorno. AWS Se recomienda contar con un proceso para revisar continuamente el uso de los volúmenes de EBS. Puede usarlo para revisar las instancias AWS Compute Optimizerinfrautilizadas. Esta sección le ayuda a identificar, administrar y eliminar los volúmenes de EBS que están desconectados o infrautilizados.

Amazon EBS

Amazon Elastic Block Store (Amazon EBS) es un dispositivo a nivel de bloque que ofrece volúmenes de almacenamiento para instancias de Amazon Elastic Compute Cloud (Amazon). EC2 EBS proporciona almacenamiento persistente, con la flexibilidad de conectarse y desconectarse de las instancias. EC2 Esto significa que el ciclo de vida de los volúmenes de EBS persiste incluso si se termina una EC2 instancia. El DeleteOnTerminationatributo es una función que controla si se deben conservar o eliminar los volúmenes de EBS adjuntos al finalizar la instancia. De forma predeterminada, el atributo se establece en el volumen raíz, lo que provoca su eliminación. `True` Está configurado `False` para otros volúmenes, lo que permite su conservación.

Impacto del costo

Los volúmenes de EBS independientes, también denominados volúmenes huérfanos o no utilizados, incurren en los mismos cargos que los volúmenes adjuntos, según el tamaño y el tipo de almacenamiento aprovisionados. Si bien el coste medio de los cargos de Amazon EBS puede parecer mínimo, de 0,10 USD por GB al mes, es fundamental reconocer que la acumulación de volúmenes de EBS no utilizados puede generar costes significativos a lo largo del tiempo.

Por ejemplo, considere las consecuencias de conservar 50 volúmenes de EBS no utilizados, cada uno aprovisionado con un tamaño de almacenamiento de 100 GB, como se muestra en la siguiente tabla.

Número de volúmenes de almacenamiento	Tipo de volumen	Tamaño	Coste mensual total
50 volúmenes	gp2 (0,10 USD)	100 GB	100 GB (50,00$, volúmenes de EBS al mes, 0,10 USD = 500,00 USD)

El escenario de la tabla anterior arroja una reducción de costos de aproximadamente 500 dólares al mes o 6000 dólares al año. Se trata de un paso eficaz hacia la reducción de costes. Asegúrese de incorporar la eliminación de los volúmenes de EBS no conectados como práctica habitual en la higiene de su AWS entorno.

Recomendaciones de optimización de costos

Puede utilizarlo AWS para automatizar fácilmente la eliminación de volúmenes de EBS no adjuntos. Por ejemplo, puede usar AWS Lambda Amazon CloudWatch y AWS Systems Manager para definir los criterios para eliminar los volúmenes no adjuntos en función de la antigüedad, las etiquetas y otras especificaciones. AWS Config También puede utilizarlos Servicios de AWS para automatizar el proceso de limpieza a escala.

Para evitar consecuencias imprevistas, le recomendamos que actúe con la diligencia debida antes de eliminar los volúmenes de EBS no adjuntos.

Administre los volúmenes de EBS independientes

Le recomendamos que tenga en cuenta las siguientes prácticas recomendadas:

- Cumpla con los requisitos de conformidad: compruebe que la eliminación de los volúmenes de EBS independientes cumpla con los requisitos de gobernanza y cumplimiento de su organización.
- Establezca políticas de respaldo y retención de datos: antes de eliminar un volumen de EBS no adjunto, haga una copia de seguridad de los datos importantes en otro repositorio de almacenamiento (por ejemplo, Amazon S3). Para la retención de datos, las instantáneas de Amazon EBS son una forma más rentable de retener los datos que los volúmenes de EBS, y pueden restaurar el volumen si es necesario en el futuro. Para obtener más información sobre la administración eficaz de las instantáneas, consulte la sección Modificar las instantáneas de Amazon EBS de esta guía.
- Compruebe si hay dependencias: compruebe si hay dependencias entre los volúmenes de EBS no conectados y otros recursos. AWS Puede utilizar la API AWS Management Console o una para recopilar información descriptiva sobre sus volúmenes de EBS, como el tamaño, el estado y los recursos asociados. Este es un paso importante para evitar que se eliminen recursos temporalmente no asociados.
- Cree una política de retención: establezca un período de retención para los volúmenes de EBS no adjuntos. Esto puede ayudarlo a identificar el momento adecuado para eliminar los volúmenes no adjuntos y garantizar que su AWS entorno permanezca optimizado. Por ejemplo, puede crear una EventBridge regla de Amazon para iniciar una función Lambda de forma programada. La función Lambda puede usar el AWS SDK para identificar activamente cualquier volumen de EBS no adjunto, aplicar un mecanismo de etiquetado para facilitar el seguimiento y enviar notificaciones cuando un volumen de EBS no adjunto alcance o supere un umbral definido.
- Etiquetar volúmenes de EBS independientes: etiquetar los volúmenes de EBS es una práctica útil que puede ayudar a organizar e identificar los volúmenes en función de atributos como el entorno,

la aplicación o el propietario. Esto puede resultar especialmente útil a la hora de decidir qué volúmenes independientes se van a eliminar, ya que permite identificar rápidamente los volúmenes que ya no se necesitan en función de sus etiquetas.

- Garantice una eliminación segura: comprobar cuándo se adjuntó por última vez un volumen de EBS puede ayudarle a determinar si es seguro eliminarlo. Para obtener más información, consulte ¿Cómo uso AWS CLI los comandos para enumerar el historial de adjuntos o separaciones de un volumen específico de Amazon EBS? en el Centro de conocimiento AWS .
- Identifique los volúmenes de EBS infrautilizados: identificar y eliminar los volúmenes de EBS infrautilizados es una práctica muy recomendable para reducir los costos de almacenamiento y mantener un entorno optimizado. AWS AWS Trusted Advisor y AWS Compute Optimizerpuede ayudarlo a identificar los volúmenes de EBS infrautilizados y a ofrecer recomendaciones para reducir los costos y mejorar la eficiencia. Por ejemplo, consulte Configurar la automatización para optimizar los volúmenes de EBS con AWS Trusted Advisor (GitHub), Establecer un panel de Trusted Advisor organización (TAO) (AWS Workshop Studio) y Optimizar los costes de los volúmenes de Amazon EBS mediante AWS Compute Optimizer (AWS Storage Blog).

Automatice la limpieza de los volúmenes de EBS independientes

Le recomendamos que tenga en cuenta las siguientes herramientas para ayudarle a automatizar la limpieza de los volúmenes de EBS no conectados:

- AWS APIs (DescribeVolumes) — Puede filtrar y buscar volúmenes de EBS independientes mediante AWS SDKs o el AWS Command Line Interface ()AWS CLI. Puede ahorrar tiempo y esfuerzo automatizando este proceso con un script o una función Lambda que se ejecute según una programación. Un ejemplo de script GitHub muestra cómo funciona esto. El script usa Lambda para analizar los AWS CloudTrail registros e identificar los volúmenes de EBS no conectados.
- AWS Systems Manager Automatización: esto le permite automatizar las tareas rutinarias de mantenimiento y corrección en su infraestructura. Para empezar, cree un manual de automatización, que defina una serie de pasos que deben ejecutarse en un orden específico. Por ejemplo, puede crear un manual que cree primero una instantánea del volumen de EBS no adjunto y, a continuación, elimine el volumen propiamente dicho. Esto puede ayudarlo a automatizar tareas que, de otro modo, llevarían mucho tiempo y serían propensas a errores si se realizaran manualmente.
- AWS Config— Esto le permite evaluar, auditar y realizar un seguimiento de los cambios en sus AWS recursos a lo largo del tiempo. Al capturar los cambios de configuración, puede utilizarlos AWS Config para evaluar el cumplimiento, la gobernanza y la utilización de los recursos en

su entorno. Por ejemplo, AWS Config puede identificar los volúmenes de EBS no utilizados. Además, puede asociar la AWS Systems Manager automatización AWS Config para corregir automáticamente la eliminación de los volúmenes de EBS no utilizados.

Recursos adicionales

- Elimine los volúmenes no utilizados de Amazon Elastic Block Store (Amazon EBS) AWS Config utilizando AWS Systems Manager y AWS (Guía prescriptiva)
- Controle sus AWS costes mediante la eliminación de los volúmenes de Amazon EBS no utilizados (blog de operaciones y migraciones en la AWS nube)
- AWSConfigRemediation-DeleteUnusedEBSVolume(Referencia del AWS Systems Manager manual de automatización)

Amazon FSx

Amazon FSx for Windows File Server es un servicio de almacenamiento de archivos totalmente gestionado que está optimizado para las cargas de trabajo de Windows. Le proporciona una solución sencilla y escalable para ejecutar sus aplicaciones y cargas de trabajo basadas en Windows, sin necesidad de gestionar una infraestructura de almacenamiento compleja. Puede utilizar Windows File Server FSx para aprovisionar y acceder fácilmente al almacenamiento de archivos compartidos que admite sus aplicaciones de Windows de forma nativa, incluidas Microsoft SQL Server SharePoint, Microsoft y aplicaciones.NET personalizadas. Además, FSx para Windows, File Server le ayuda a administrar los costos al ofrecer opciones de precios flexibles, como cuotas de almacenamiento, pay-as-you-go y deduplicación automática de datos para reducir el espacio de almacenamiento y optimizar el rendimiento y los costos.

Esta sección abarca los siguientes temas:

- Elija el almacenamiento de archivos SMB adecuado
- Habilite la deduplicación de datos en Amazon FSx
- Conozca la fragmentación de datos en Windows FSx File Server
- Comprenda el uso del volumen del disco duro en Amazon FSx
- Utilice una única zona de disponibilidad

Elija el almacenamiento de archivos SMB adecuado

Descripción general

AWS ofrece una variedad de servicios de almacenamiento totalmente gestionados que le brindan las amplias capacidades de los servicios de archivos líderes del sector, al tiempo que combinan las últimas innovaciones de AWS infraestructura y seguridad. Puede incorporar AWS servicios en los flujos de trabajo de infraestructura como código (IaC) e integrarlos con los servicios de AWS computación, supervisión y protección de datos. Para las cargas de trabajo de Windows, puede elegir entre dos servicios de archivos totalmente gestionados que se pueden utilizar para satisfacer las necesidades de sus aplicaciones: FSx para Windows File Server y Amazon FSx para NetApp ONTAP.

FSx para Windows File Server

Amazon FSx for Windows File Server proporciona almacenamiento compartido totalmente gestionado integrado en Windows Server y ofrece una amplia gama de funciones de acceso, gestión y administración de datos. FSx para Windows File Server se integra fácilmente en los entornos de Windows porque es un servicio nativo de Windows. Se recomienda utilizar el servidor de archivos de Windows FSx para compartir usuarios y grupos, las instancias de clúster de conmutación por error Always On para SQL Server, las aplicaciones de Windows y la infraestructura de escritorios virtuales (VDI). FSx para Windows File Server también se integra bien con Amazon FSx File Gateway, Amazon Kendra, los registros de auditoría de Amazon S3 y Amazon Data Firehose.

FSx para ONTAP

FSx for ONTAP se basa NetApp en el sistema de archivos ONTAP patentado por ONTAP. Requiere cierto nivel de capacitación y se recomienda principalmente a los usuarios locales existentes. NetApp Entre los casos de uso habituales se incluyen los recursos compartidos de usuarios y grupos, las instancias de clúster de conmutación por error Always On para SQL Server y las aplicaciones de Windows. FSx para ONTAP, admite varios protocolos, sistemas de archivos de más de 64 TB (a escala PB sin un servidor de espacio de nombres DFS), la clonación, la replicación, las instantáneas, la compresión (eficiencia del almacenamiento) y la organización inteligente de los datos en niveles.

Impacto del costo

FSx para Windows File Server

FSx para Windows File Server fue la primera solución de almacenamiento compartido AWS para implementar instancias de clúster de conmutación por error para SQL Server. Con FSx el servidor

de archivos de Windows, puede lanzar instancias de clúster de conmutación por error utilizando las licencias de la edición SQL Standard. Sin embargo, esto le impide confiar en los grupos de disponibilidad de Always On, que requieren licencias de la edición Enterprise de SQL Server. Al cambiar de la edición SQL Server Enterprise Standard a la edición SQL Server Standard, podría ahorrar entre un 65 y un 75 por ciento en sus licencias de SQL Server.

Puede utilizarlas FSx para Windows File Server for Failover Cluster Instancias para reducir la carga de E/S de almacenamiento del almacenamiento típico de EBS. Al reducir la carga de E/S a FSx Windows File Server, podría reducir las EC2 instancias, que dependen de un alto rendimiento de Amazon EBS y de IOPS, sin que ello afecte al rendimiento del almacenamiento.

FSx para ONTAP

Puede utilizar ONTAP FSx para ejecutar su clúster de conmutación por error de Microsoft en el protocolo de bloques iSCSI y beneficiarse de la inicialización instantánea de archivos de SQL Server, el uso SnapMirror de la replicación entre regiones, el soporte de antivirus y la clonación. Si crea varias copias de bases de datos para realizar pruebas, la clonación puede suponer una diferencia significativa tanto en el consumo de espacio como en la rapidez con la que se pueden crear esas copias de bases de datos. Además, puede utilizar NetApp SnapCenter ONTAP para gestionar las funciones de copia de seguridad, restauración y clonación de EC2 las instancias de SQL Server. FSx FSx for ONTAP también ofrece una organización automática en niveles, desde unidades SSD hasta un almacenamiento agrupado de capacidad de bajo coste, con una combinación de rendimiento y rentabilidad.

FSx para ONTAP NetApp es compatible con el sistema de archivos (ONTAP), a diferencia del servidor de archivos FSx de Windows, que admite un sistema de archivos NTFS nativo de Windows. El tamaño mínimo FSx para ONTAP es de 1024 GB, mientras que FSx para Windows File Server puede empezar con tan solo 32 GB.

Integración con el sistema de archivos distribuido de Microsoft

FSx para Windows File Server y FSx ONTAP, se integran con el sistema de archivos distribuido (DFS) de Microsoft para lograr una integración perfecta en las implementaciones existentes. Tenga en cuenta lo siguiente al planificar su arquitectura:

- FSx para Windows File Server y FSx ONTAP, admiten espacios de nombres DFS (DFSN) en ambos tipos de implementación (varias zonas de disponibilidad y zonas de disponibilidad únicas).
- Solo FSx para Windows, el servidor de archivos admite la replicación DFS (DFSR) y solo cuando se utilizan zonas de disponibilidad únicas.

Recomendaciones de optimización de costos

Tanto para Windows File Server como FSx FSx para ONTAP, el rendimiento depende en gran medida de la configuración, al igual que su precio. FSx para Windows File Server, el precio depende principalmente de la capacidad y el tipo de almacenamiento, la capacidad de rendimiento, las copias de seguridad y los datos transferidos. Con ONTAP, usted paga FSx por el almacenamiento en SSD, las IOPS de SSD, el uso del conjunto de capacidades, la capacidad de rendimiento y las copias de seguridad.

Servicio de archivos	Coste de 5 TB de almacenamiento	Configuración	Región
FSx para Windows File Server	982,78$	Una sola zona de disponibilidad SSD (15 000 IOPS) 32 MBps 5 TB de respaldo (sin ahorros en la deduplicación)	Este de EE. UU. (Norte de Virginia)
FSx para ONTAP	979,28 DÓLARES	Una sola zona de disponibilidad 100% SSD Nivel de capacidad de lectura y escritura de 15 000 15 000 IOPS EN SSD 128 MBps 5 TB de respaldo (sin ahorros en la deduplicación)	Este de EE. UU. (Norte de Virginia)

Tenga en cuenta lo siguiente:

- La deduplicación y la compresión le permiten almacenar más datos en dispositivos físicos al reducir el tamaño de los datos, pero usted paga por el almacenamiento aprovisionado en unidades de estado sólido (SSD) o disco duro (HDD).
- Puede utilizar ONTAP FSx para organizar sus datos en niveles. Es extremadamente raro que se pueda acceder al 100 por ciento de los datos con regularidad y que se requiera almacenamiento en SSD. Puede mover los datos inactivos y a los que se accede con poca frecuencia a un nivel de capacidad para ahorrar costes.
- Los precios que se mencionan aquí se calculan con un 100 por ciento de los datos en el nivel SSD y 15 000 IOPS en el nivel SSD.

Copia de seguridad

De forma predeterminada, tanto FSx para ONTAP como FSx para Windows File Server almacenan sus copias de seguridad totalmente gestionadas en Amazon S3. Sin embargo, FSx para ONTAP hay una opción adicional para el uso de copias de seguridad SnapVault, que permite configurar las copias de seguridad para que residan en el nivel de capacidad. La copia de seguridad SnapVault es un mecanismo autogestionado que resulta más rentable que la opción de copia de seguridad totalmente gestionada por defecto. La opción de backup totalmente gestionado cuesta 0,05 USD por GB al mes. El coste de la SnapVault copia de seguridad de ONTAP (unidad SSD FSx por unidad de almacenamiento con capacidad de 10:1) cuesta 0,03221 USD (0,9 x 0,0219+0,1 x 0,125).

Tenga en cuenta lo siguiente:

- AWS Las copias de seguridad gestionadas ofrecen una granularidad de una hora. SnapVaultle permite trabajar en tan solo cinco minutos.
- Puede usar NetApp las herramientas (como la CLI y la API) para configurar SnapVault las relaciones y la replicación de instantáneas.
- Habilite la política de `all` estratificación en un SnapVault volumen para usar el nivel de capacidad como almacenamiento para los datos de respaldo.
- SnapVault los destinos pueden estar en el mismo lugar Región de AWS, entre regiones o locales. Por lo general, se dirige a un destino de copia de seguridad del sistema de archivos de una sola zona de disponibilidad o de varias zonas de disponibilidad. En comparación, AWS Backup está respaldado por la resiliencia regional de Amazon S3.

Tamaño correcto

También puede ahorrar costes y sacar el máximo partido a su sistema de archivos ajustando el tamaño y evitando el sobreaprovisionamiento.

Para ajustar el tamaño, haga lo siguiente:

1. Identifique sus necesidades actuales en función de los datos. Para las cargas de trabajo típicas de Windows, puede utilizar herramientas integradas del sistema operativo, como el Monitor de rendimiento.
2. En el Monitor de rendimiento, utilice los siguientes contadores para evaluar sus necesidades de rendimiento actuales. El intervalo de captura se establece en un segundo, con un tamaño de registro máximo de 1000 MB y está habilitada la sobrescritura.

```
Logman.exe create counter PerfLog-Short -o "c:\perflogs\PerfLog-Long.blg" -f bincirc
 -v mmddhhmm -max 1024 -c "\LogicalDisk(*)\*" "\Memory\*" "\.NET CLR Memory(*)\*"
 "\Cache\*" "\Network Interface(*)\*" "\Paging File(*)\*" "\PhysicalDisk(*)\*"
 "\Processor(*)\*" "\Processor Information(*)\*" "\Process(*)\*" "\Thread(*)\*"
 "\Redirector\*" "\Server\*" "\System\*" "\Server Work Queues(*)\*" "\Terminal
 Services\*" -si 00:00:01
```

3. Para iniciar la captura de registros, ejecute el `logman start PerfLog-Short` comando. Para detener la captura del registro, ejecute el `logman stop PerfLog-Short` comando.

> **Note**
>
> Puede encontrar los archivos de registro de rendimiento en c:\perflogs, en el servidor que ejecuta la captura. Para obtener más información, consulte Descripción general del Monitor de rendimiento de Windows en la documentación de Microsoft.

4. Tras identificar la configuración correcta, comprueba si tu estimación es correcta en el sistema de FSx archivos de Amazon mediante herramientas de stress de disco como Microsoft DISKSPD.
5. Si está satisfecho con el rendimiento, pase al recurso compartido de archivos.

Recomendamos un enfoque conservador de la capacidad de almacenamiento, ya que solo se puede ampliar. La capacidad de rendimiento se puede aumentar o reducir según sea necesario.

Recursos adicionales

- Amazon FSx for NetApp ONTAP FAQs (AWS sitio web)
- Optimización del rendimiento del servidor de archivos de Amazon FSx para Windows con nuevas métricas (blog sobre AWS almacenamiento)

Habilite la deduplicación de datos en Amazon FSx

Descripción general

La deduplicación de datos es una función que le permite almacenar sus datos de manera más eficiente y con menos requisitos de capacidad. Implica encontrar y eliminar la duplicación en los datos sin comprometer su fidelidad o integridad. La deduplicación de datos utiliza la fragmentación y la compresión de subarchivos de tamaño variable, lo que ofrece índices de optimización de 2:1 para los servidores de archivos generales y de hasta 20:1 para los datos de virtualización. La deduplicación de datos es mucho más eficaz que la compresión NTFS. La resiliencia ante los fallos de hardware es inherente a la arquitectura de deduplicación, con una validación completa de los datos y los metadatos mediante una suma de comprobación, incluida la redundancia de los metadatos y de los fragmentos de datos a los que se accede con mayor frecuencia.

FSx para Windows, File Server es totalmente compatible con la deduplicación de datos. Su uso puede suponer un ahorro medio del 50 al 60% en el caso de los archivos compartidos de uso general. Con las acciones, los ahorros oscilan entre el 30 y el 50% en los documentos de usuario y entre el 70 y el 80% en los conjuntos de datos de desarrollo de software. Es importante comprender que los ahorros de almacenamiento que se pueden lograr con la deduplicación de datos dependen de la naturaleza del conjunto de datos, incluida la cantidad de duplicados que existan entre los archivos. La deduplicación no es una buena opción si los datos almacenados son de naturaleza dinámica.

Impacto del costo

Para hacer frente al crecimiento del almacenamiento de datos en la empresa, los administradores consolidan los servidores y hacen que el escalamiento de la capacidad y la optimización de los datos sean objetivos clave. La configuración predeterminada de la deduplicación de datos puede suponer un ahorro inmediato, o los administradores pueden ajustar la configuración para obtener beneficios adicionales. Por ejemplo, puede configurar la deduplicación para que se ejecute solo en determinados tipos de archivos o puede crear un cronograma de trabajo personalizado.

En términos generales, la deduplicación tiene tres tipos de tareas: optimización, recolección de elementos no utilizados y limpieza. Tenga en cuenta que no se liberará espacio hasta que ejecute un trabajo de recolección de basura después de la optimización. Puede programar el trabajo o ejecutarlo manualmente. Todos los ajustes disponibles al programar un trabajo de deduplicación de datos también están disponibles al iniciar un trabajo manualmente (excepto los que son específicos de la programación).

Si bien la deduplicación supone solo un 25 por ciento de ahorro efectivo, Windows File Server supone un importante ahorro de costes. FSx Estos ahorros proyectados se basan en una estimación del. Calculadora de precios de AWS

Recomendaciones de optimización de costos

La deduplicación activada FSx para los sistemas de archivos del servidor de archivos de Windows no está habilitada de forma predeterminada. Para habilitar la deduplicación mediante la administración remota activada PowerShell, debe ejecutar el `Enable-FSxDedup` comando y, a continuación, usarlo para establecer la configuración. `Set-FSxDedupConfiguration` Para obtener más información, consulte Administrar sistemas de archivos en la documentación del servidor FSx de archivos de Windows.

Para habilitar la deduplicación, ejecute el siguiente comando:

```
PS C:\Users\Admin> Invoke-Command -ComputerName amznfsxzzzzzzzz.corp.example.com -
ConfigurationName FSxRemoteAdmin -ScriptBlock {Enable-FsxDedup }
```

Para verificar la configuración de deduplicación, ejecute el siguiente comando:

```
Invoke-Command -ComputerName amznfsxzzzzzzzz.corp.example.com -ConfigurationName
 FSxRemoteAdmin -ScriptBlock {
Set-FSxDedupSchedule -Name "CustomOptimization" -Type Optimization -Days
 Mon,Tues,Wed,Sat -Start 09:00 -DurationHours 7
}
```

Al ejecutar el PowerShell `Measure-DedupFileMetadata` cmdlet, puede determinar cuánto espacio potencial de disco se puede recuperar en un volumen si elimina un grupo de carpetas, una sola carpeta o un solo archivo y, a continuación, ejecuta un trabajo de recolección de elementos no utilizados. En concreto, el `DedupDistinctSize` valor indica cuánto espacio se recupera si se eliminan esos archivos. Los archivos suelen tener fragmentos que se comparten en otras carpetas,

por lo que el motor de deduplicación calcula qué fragmentos son únicos y se eliminarían tras la tarea de recolección de elementos no utilizados.

Los cronogramas predeterminados de los trabajos de deduplicación de datos están diseñados para funcionar bien con las cargas de trabajo recomendadas y ser lo menos intrusivos posible (excepto el trabajo de optimización de prioridades que está habilitado para el tipo de uso de la copia de seguridad). Si las cargas de trabajo requieren grandes recursos, le recomendamos programar los trabajos para que se ejecuten únicamente durante las horas de inactividad o para reducir o aumentar la cantidad de recursos del sistema que puede consumir un trabajo de deduplicación de datos.

De forma predeterminada, la deduplicación de datos utiliza el 25 por ciento de la memoria disponible. Sin embargo, esto se puede aumentar mediante el uso `-memory switch` de. Para los trabajos de optimización, le recomendamos que establezca un rango entre 15 y 50. Para los trabajos programados, puede utilizar un mayor consumo de memoria. Por ejemplo, con los trabajos de recolección de basura y limpieza (que normalmente se programan para que se ejecuten fuera del horario laboral), puede establecer un consumo de memoria más alto (por ejemplo, 50).

Para obtener información adicional sobre la configuración de deduplicación de datos, consulte Reducir los costos de almacenamiento mediante la deduplicación de datos en la documentación del servidor FSx de archivos de Windows.

Recursos adicionales

- Descripción de la deduplicación de datos (documentación de Microsoft)
- Reducir los costos de almacenamiento con la deduplicación de datos (FSx para la documentación del servidor de archivos de Windows)

Conozca la fragmentación de datos en Windows FSx File Server

Descripción general

FSx para Windows File Server, el rendimiento depende de la configuración. Se basa principalmente en el tipo de almacenamiento, la capacidad de almacenamiento y la configuración del rendimiento. La capacidad de rendimiento que seleccione determina los recursos de rendimiento disponibles para el servidor de archivos, incluidos los límites de E/S de la red, la CPU y la memoria y los límites de E/S del disco impuestos por el servidor de archivos. La capacidad de almacenamiento y el tipo de almacenamiento que seleccione determinan los recursos de rendimiento disponibles para los volúmenes de almacenamiento: los límites de E/S de disco impuestos por los discos de

almacenamiento. Además del rendimiento, las opciones de configuración también influyen en el coste. FSx para Windows File Server, el precio depende principalmente de la capacidad y el tipo de almacenamiento, la capacidad de rendimiento, las copias de seguridad y los datos transferidos.

Si tiene requisitos de rendimiento y almacenamiento de archivos relativamente grandes, puede beneficiarse de la fragmentación de datos. La fragmentación de datos implica dividir los datos de los archivos en conjuntos de datos más pequeños (fragmentos) y almacenarlos en diferentes sistemas de archivos. Las aplicaciones que acceden a los datos desde varias instancias pueden alcanzar niveles de rendimiento altos al leer y escribir en estos fragmentos en paralelo. Al mismo tiempo, aún puede presentar una vista unificada en un espacio de nombres común para las aplicaciones. Además, también puede ayudar a escalar el almacenamiento de datos de archivos más allá de lo que admite cada sistema de archivos (64 TB) para conjuntos de datos de archivos grandes (hasta cientos de petabytes).

Impacto del costo

En el caso de conjuntos de datos de gran tamaño, suele ser más eficaz implementar varios sistemas de archivos pequeños FSx para Windows File Server, en lugar de una unidad SSD grande para lograr el mismo nivel de rendimiento. El uso de una combinación FSx de los tipos de almacenamiento en disco duro y SSD del servidor de archivos de Windows permite ahorrar costes y adaptar la carga de trabajo al mejor subsistema de disco subyacente. En las siguientes tablas, puede ver la diferencia entre un único sistema de archivos de 17 TB y compararla con varios sistemas de archivos más pequeños que aumentan la misma capacidad.

Sistema de archivos SSD de gran tamaño con múltiples cargas de trabajo

Nombre del servidor	Costo	Configuración	Región
Servidor FSx de archivos Amazon para Windows	5.716 DÓLARES	17 TB DE SSD 30 por ciento de deduplicación 256 Mbps 17 TB de respaldo	Este de EE. UU. (Norte de Virginia)

Carga de trabajo particionada mediante DFSN

Nombre del servidor	Costo	Configuración	Región	Share
Servidor FSx de archivos Amazon para Windows	1.024 DÓLARES	SSD DE 2 TB 20% de deduplicación 128 Mbps 2 TB de respaldo Multi-AZ	Este de EE. UU. (Norte de Virginia)	Comparte 1
Servidor FSx de archivos Amazon para Windows	2.132 DÓLARES	SSD DE 5 TB 30% de deduplicación 256 Mbps 5 TB de respaldo Multi-AZ	Este de EE. UU. (Norte de Virginia)	Comparte 2
Servidor FSx de archivos Amazon para Windows	1.036 DÓLARES	DISCO DURO DE 10 TB 40% de deduplicación 128 Mbps Copia de seguridad de 10 TB Multi-AZ	Este de EE. UU. (Norte de Virginia)	Comparte 3

Nombre del servidor	Costo	Configuración	Región	Share
Instancias de Windows EC2 de DFSN	27 DÓLARES	t3a.medium 2 v CPUs Memoria de 4 GiB	Este de EE. UU. (Norte de Virginia)	Instancias DFSN

El coste anual de un sistema de archivos SSD de gran tamaño es de 68.592$. El coste anual de una carga de trabajo particionada es de 50.640$. En este ejemplo, se puede lograr un ahorro del 26 por ciento y, al mismo tiempo, adaptar la carga de trabajo al almacenamiento de back-end adecuado. Para obtener más información sobre la estimación de precios, consulte la Calculadora de precios de AWSestimación.

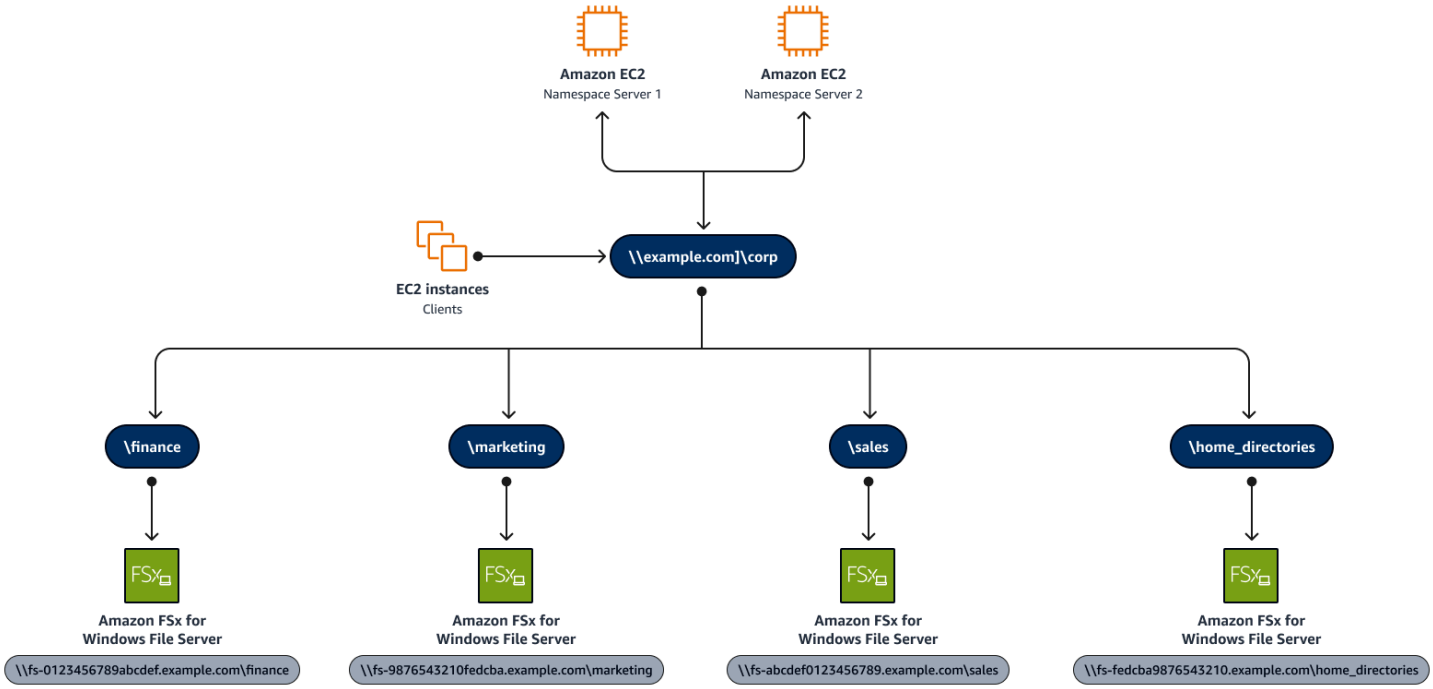
Recomendaciones de optimización de costos

Para implementar una solución de deduplicación de datos, debe configurar un espacio de nombres DFS de Microsoft en función del tipo de datos, el tamaño de E/S y el patrón de acceso de E/S. Cada espacio de nombres admite hasta 50 000 archivos compartidos y cientos de petabytes de capacidad de almacenamiento en total.

Lo más eficaz es elegir una convención de fragmentación que distribuya la E/S de manera uniforme entre todos los sistemas de archivos que vaya a utilizar. Supervisar la carga de trabajo le ayudará a lograr una optimización adicional o a reducir los costos. Si necesita ayuda para medir la información de rendimiento del sistema de FSx archivos de Amazon, consulte el rendimiento del servidor FSx de archivos de Windows en la documentación del servidor FSx de archivos de Windows.

Tras elegir una estrategia de fragmentación, puede agrupar los sistemas de archivos para acceder fácilmente a sus recursos compartidos mediante espacios de nombres DFS. Esto permite a los usuarios ver un sistema de archivos homogéneo, cuando en realidad están accediendo a una variedad de sistemas de archivos diferentes con casos de uso específicos. Es importante crear los recursos compartidos con una convención de nomenclatura adecuada para que los usuarios finales puedan descifrar fácilmente para qué carga de trabajo están diseñados los recursos compartidos. También es importante etiquetar los recursos compartidos de producción y los que no son de producción, de modo que los usuarios finales no coloquen los archivos por error en el sistema de archivos incorrecto.

El siguiente diagrama muestra cómo se puede utilizar un único espacio de nombres DFS como punto de acceso para varios sistemas de archivos de Amazon FSx .



Tenga en cuenta lo siguiente:

- Puede agregar los recursos compartidos existentes FSx para el servidor de archivos de Windows a un árbol DFS.
- No se FSx puede añadir Amazon a la raíz de la ruta de acceso compartido de DFS. Solo tiene una subcarpeta.
- Debe implementar una EC2 instancia que sirva para la configuración del espacio de nombres DFS.

Para obtener más información acerca de la configuración del DFS-N, consulte Introducción a los espacios de nombres DFS en la documentación de Microsoft. Para obtener más información sobre el uso de los espacios de nombres DFS, consulte el vídeo Uso de los espacios de nombres DFS con Amazon FSx para Windows File Server sobre. YouTube

Recursos adicionales

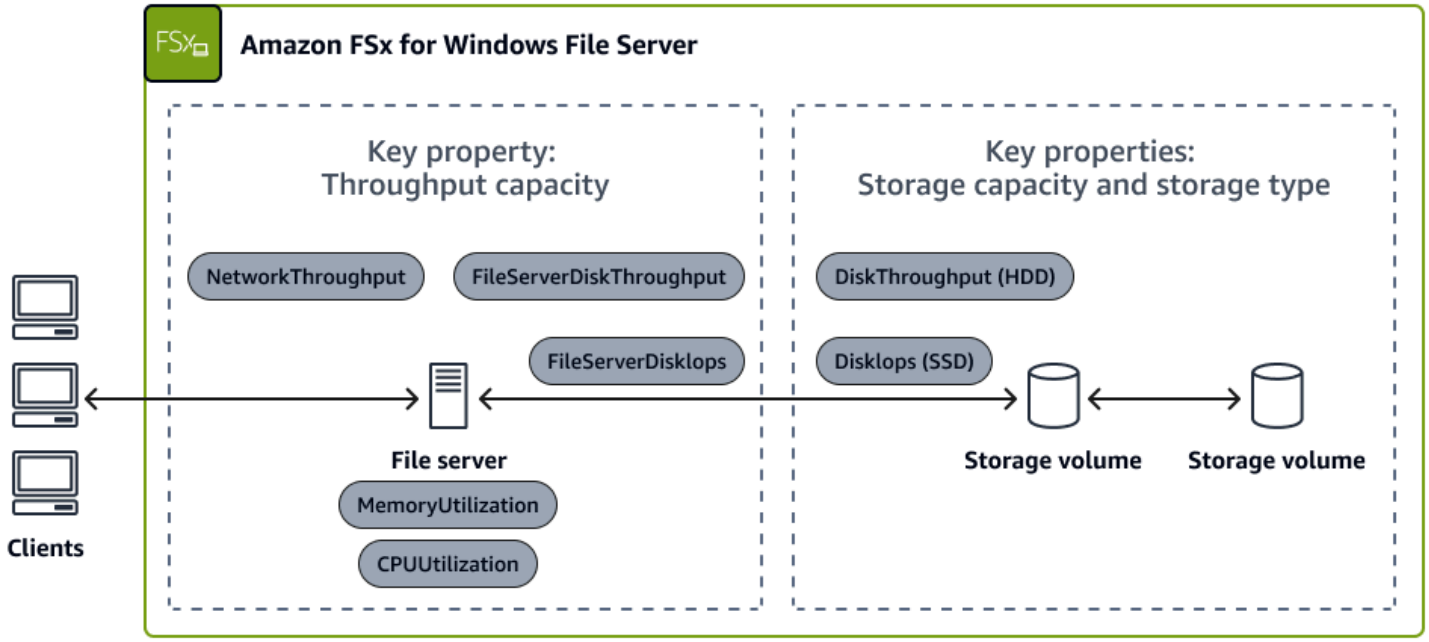
- Agrupación de varios sistemas de archivos con espacios de nombres DFS (documentación de Amazon) FSx
- Tutorial 6: Ampliar el rendimiento con fragmentos (documentación de Amazon FSx)
- Uso de espacios de nombres DFS con Amazon FSx for Windows File Server (Labs)AWS

Comprenda el uso del volumen del disco duro en Amazon FSx

Descripción general

Amazon FSx para Windows File Server ofrece la flexibilidad de elegir el rendimiento independientemente de la capacidad del sistema de archivos. Hay dos ajustes de capacidad disponibles: HDD (unidad de disco duro) y SSD (unidad de estado sólido). Las unidades EBS st1 se utilizan para el almacenamiento del sistema de archivos en el disco duro. Las unidades EBS io1 se utilizan para las SSD.

El siguiente diagrama muestra la relación entre la configuración de rendimiento y almacenamiento.



Con el almacenamiento basado en disco duro, recibe una base de 12 IOPS con 80 IOPS de disco en ráfaga (por IOPs TiB de almacenamiento) y un rendimiento de 12 Megabytes/second baseline with 80 burst Megabytes/second (por TiB de almacenamiento). Por ejemplo, si su recurso compartido tiene un tamaño de 50 TB, obtendrá 50 * 12 = 600 como referencia tanto para el rendimiento como para las IOPS.

El servidor de archivos de Amazon FSx para Windows proporciona 80 ráfagas de IOPS. Los créditos de ráfaga se recargan automáticamente cuando la utilización es inferior a la tasa de referencia y se consumen automáticamente cuando la utilización supera la tasa de referencia. Por ejemplo, si su carga de trabajo solo utiliza un 10% (IOPS/TB for an hour (2 IOPS/TB below your baseline rate), you can then utilize 14 IOPS/TB (2 IOPS/TBpor encima de su valor de referencia) durante la hora siguiente antes de volver a quedarse sin créditos acumulados.

Para las operaciones de archivos, Amazon FSx for Windows File Server proporciona latencias consistentes de menos de milisegundos con almacenamiento SSD y latencias de milisegundos de un solo dígito con almacenamiento en disco duro. Para todos los sistemas de archivos, incluidos aquellos con almacenamiento en disco duro, Amazon FSx para Windows File Server proporciona una caché rápida (en memoria) en el servidor de archivos, de modo que puedes obtener un alto rendimiento y latencias de menos de milisegundos para los datos a los que se accede activamente, independientemente del tipo de almacenamiento.

Cuando es apropiado, el uso del almacenamiento en disco duro puede ayudar a reducir el coste de la capacidad total de almacenamiento y proporcionar una plataforma de almacenamiento fiable para sus necesidades.

Impacto del costo

El rendimiento del servidor de archivos de Amazon FSx para Windows depende de tres factores: la capacidad de almacenamiento, el tipo de almacenamiento y el rendimiento. El rendimiento de E/S de la red y el tamaño de la caché en memoria vienen determinados únicamente por la capacidad de rendimiento, mientras que el rendimiento de E/S del disco viene determinado por una combinación de la capacidad de rendimiento, el tipo de almacenamiento y la capacidad de almacenamiento.

Si bien el SSD se recomienda para cargas de trabajo con un uso intensivo de E/S, existen diversas cargas de trabajo cuyas necesidades pueden satisfacerse con las especificaciones de rendimiento del disco duro. El almacenamiento en el disco duro está diseñado para una amplia gama de cargas de trabajo, incluidos los directorios principales, los recursos compartidos de usuarios y departamentos y los sistemas de administración de contenido. Por ejemplo, si sus usuarios solo necesitan un acceso de baja latencia a los datos que respaldan los proyectos actuales, el acceso a la mayoría de los datos que está almacenando será poco frecuente.

Puedes usar el Calculadora de precios de AWSpara comparar una SSD de 20 TB con un sistema de archivos HDD. `us-east-1` Como se muestra en la siguiente tabla, aunque no se ahorre en la deduplicación, la diferencia de costes es significativa cuando se comparan los sistemas de archivos HDD con los sistemas de archivos SSD.

Configuración FSx del sistema de archivos de Amazon	Costes mensuales
SSD Multi-AZ de 20 TB () `us-east-1`	4.699,30 DÓLARES
Disco duro Multi-AZ de 20 TB () `us-east-1`	542,88$

Configuración FSx del sistema de archivos de Amazon	Costes mensuales
Ahorros mensuales estimados	4.156,42$

> **Note**
>
> FSx Para obtener más ahorros en Windows File Server, consulte la FSx sección Habilitar la deduplicación de datos en Amazon de esta guía.

Al identificar correctamente sus necesidades de rendimiento, puede seleccionar el almacenamiento adecuado para su carga de trabajo y reducir sus costes.

Recomendaciones de optimización de costos

Si decide utilizar el almacenamiento en disco duro, pruebe su sistema de archivos para asegurarse de que cumple sus requisitos de rendimiento. El almacenamiento en disco duro tiene un costo menor en comparación con el almacenamiento en SSD, pero con niveles más bajos de rendimiento del disco e IOPS de disco por unidad de almacenamiento. Puede ser adecuado para recursos compartidos de usuarios de uso general y directorios principales con bajos requisitos de E/S, sistemas de administración de contenido de gran tamaño en los que los datos se recuperan con poca frecuencia o conjuntos de datos con un número reducido de archivos grandes.

El tipo de almacenamiento de un sistema de archivos existente no se puede cambiar. Para convertir el tipo de almacenamiento de un sistema de archivos de Amazon FSx para Windows File Server, debe realizar una copia de seguridad del sistema de archivos existente y restaurarlo en un nuevo sistema de archivos con el tipo de almacenamiento deseado. Si deseas convertir un sistema de archivos SSD existente en un sistema de archivos HDD, ten en cuenta que el HDD tiene una capacidad mínima mucho mayor, de 2 TB.

Para restaurar una copia de seguridad con un tipo de almacenamiento diferente, haga lo siguiente:

1. Realice una copia de seguridad del sistema de archivos existente.
2. Cree un nuevo sistema de FSx archivos de Amazon con el tipo de almacenamiento HDD.
3. Restaure la copia de seguridad en el nuevo sistema de archivos con el tipo de almacenamiento deseado.

4. Compruebe que el nuevo sistema de archivos tenga el tipo de almacenamiento correcto y que sus datos estén intactos.

Antes de pasar los cambios a producción, te recomendamos que analices el rendimiento de tu sistema de FSx archivos de Amazon y compruebes que el cambio es aceptable. Para obtener más información, consulta la publicación Optimización del rendimiento del servidor de archivos de Amazon FSx para Windows con nuevas métricas en el blog sobre AWS almacenamiento.

Recursos adicionales

- Optimización de costes con Amazon FSx (FSx documentación de Amazon)

Utilice una única zona de disponibilidad

Descripción general

En esta sección se explica cuándo es más beneficioso utilizar una implementación de zona de disponibilidad única de Amazon FSx para Windows File Server. Abarca situaciones en las que el traslado a una única zona de disponibilidad reduce los costes y, al mismo tiempo, le permite utilizar Amazon FSx for Windows File Server como su servicio de almacenamiento de archivos gestionado. Le recomendamos que implemente una única zona de disponibilidad para Amazon FSx para las cargas de trabajo de producción. Esto puede ayudar a garantizar la redundancia de varias zonas de disponibilidad.

Impacto del costo

Un único sistema de archivos de zonas de disponibilidad ofrece una reducción de costes de aproximadamente un 40 por ciento en comparación con una implementación de varias zonas de disponibilidad. Con un sistema de archivos de múltiples zonas de disponibilidad, paga 0,230$ por GB al mes en SSD y 0,025$ por GB al mes en disco duro, frente a los 0,130$ por GB al mes en SSD y 0,013$ por GB al mes en disco duro con un único sistema de archivos de Availability Zone. Puede ver una comparación de los costes y crear sus propias estimaciones utilizando el. Calculadora de precios de AWS

En el caso de un sistema de archivos de 10 TB, esto puede suponer la diferencia entre pagar aproximadamente 1200$ al mes para varias zonas de disponibilidad o 680$ al mes para una sola zona de disponibilidad. En este ejemplo, se utiliza un sistema de archivos de 10 TB FSx para Windows File Server con SSD. El ahorro estimado en la deduplicación es del 50 por ciento. En

general, una sola zona de disponibilidad tiene un costo de entrada más bajo, pero incluye algunas advertencias que se describen en la siguiente sección.

Recomendaciones de optimización de costos

Implementaciones de una sola zona de disponibilidad

Para asegurarse de que una única zona de disponibilidad es la adecuada, tenga en cuenta su propia zona interna SLAs para almacenar los datos del servidor de FSx archivos de Windows. Esto implica comprender si tiene que proporcionárselo SLAs a sus clientes (internos y externos) y si los tres nueves de disponibilidad de una FSx sola zona de disponibilidad de Amazon le permitirán cumplirlos SLAs. FSx en el caso de Windows, el servidor de archivos con una única zona de disponibilidad sigue teniendo un tiempo de actividad del 99,9 por ciento. El SLA de Amazon FSx para múltiples zonas de disponibilidad es superior al 99,99 por ciento. Para las cargas de trabajo de misión crítica, le recomendamos que utilice varias zonas de disponibilidad en una sola zona de disponibilidad, incluso con un coste adicional.

Las implementaciones en una sola zona de disponibilidad son ideales para cargas de trabajo como las copias de seguridad de bases de datos de SQL Server. Pueden proporcionar almacenamiento de bajo costo con un nivel de disco duro y, al mismo tiempo, proporcionarle un tiempo de actividad constante. Si necesita un mayor nivel de disponibilidad para una carga de trabajo de producción, como servidores SQL de alta disponibilidad o acceso a aplicaciones de producción, una única zona de disponibilidad no es la opción adecuada para sus cargas de trabajo. En el caso de las copias de seguridad, las pruebas no relacionadas con la producción y los entornos de desarrollo, la implementación de una FSx única zona de disponibilidad de Amazon puede reducir los costes operativos.

Un caso de uso en el que un sistema de archivos de zona de disponibilidad FSx única de Amazon funciona bien es en una situación de producción en la que se utilizan varios sistemas de archivos de zona de disponibilidad FSx única de Amazon, como almacenamiento por servidor en un clúster de SQL Server de alta disponibilidad mediante grupos de disponibilidad Always On. Para obtener más información, consulte la sección Optimización de los costes de las implementaciones de SQL Server de alta disponibilidad publicada en el AWS blog sobre AWS almacenamiento.

Replicación multirregional

Una posible opción para reducir los costos con un único sistema de archivos de zona de disponibilidad (uno en el que solo funcione un sistema de archivos de zona de disponibilidad) es si desea aprovechar la replicación multirregional con Amazon FSx. Puede implementar sistemas de archivos Single-AZ que admitan el uso con el DFS-R nativo de Microsoft. El DFS-R tiene la

capacidad de replicar automáticamente los datos entre regiones y varios sitios. Para obtener más información sobre la configuración del DFS-R mediante Amazon FSx, consulte Uso de la replicación del sistema de archivos distribuido de Microsoft en la documentación de Amazon FSx.

Otra alternativa para ahorrar costes en varias regiones es utilizar. AWS Storage Gateway Esto le permite implementar un Amazon FSx File Gateway en otra región para el acceso multirregional de Amazon FSx. Para obtener más información, consulte la sección AWS Storage Gateway de esta guía.

Si trabaja en varias regiones, debe tener en cuenta el costo de transferencia de datos del tráfico de datos entre regiones. El tráfico que se mueve entre regiones conlleva un cargo de 0,02 USD por GB. Por lo tanto, si sus datos cambian de forma constante en grandes volúmenes, esto aumentará su costo total. Por ejemplo, 1 TB de transferencia de datos equivale aproximadamente a 20,48$.

Periodo de mantenimiento

El período de mantenimiento es una consideración clave si utilizas una zona de disponibilidad única con Amazon FSx. Durante el período de mantenimiento, el sistema de FSx archivos de Amazon no estará disponible durante unos 20 minutos, debido a la aplicación rutinaria de parches de software para el servidor Windows subyacente. Si utilizas el sistema de archivos para realizar copias de seguridad nocturnas, ajusta el período de FSx mantenimiento de Amazon en consecuencia para evitar interrupciones durante la copia de seguridad. Puedes ajustar la ventana de mantenimiento después de crear tu sistema de FSx archivos de Amazon.

Recursos adicionales

- Disponibilidad y durabilidad: sistemas de archivos Single-AZ y Multi-AZ (documentación de Amazon FSx)
- Precios de servidores de archivos de Amazon FSx para Windows (AWS sitio web)

AWS Storage Gateway

AWS Storage Gateway es un servicio de almacenamiento en la nube híbrida que conecta los entornos locales con el almacenamiento en AWS la nube. Le permite integrar sin problemas su infraestructura local existente AWS, lo que le permite almacenar y recuperar datos de la nube y ejecutar aplicaciones en un entorno híbrido. Para las cargas de trabajo de Windows, puede usar Storage Gateway para almacenar y acceder a los datos mediante protocolos nativos de Windows, como SMB y NFS. Puede usar Storage Gateway para reducir los costos asociados con la ejecución

de cargas de trabajo de Windows AWS mediante el uso de hardware y software locales como puente hacia la nube. Esto le permite aprovechar la escalabilidad y la rentabilidad que supone AWS no tener que realizar cambios significativos en su infraestructura existente.

Bajo el paraguas de Storage Gateway, encontrará Amazon S3 File Gateway, Amazon FSx File Gateway, Tape Gateway y Volume Gateway. S3 File Gateway y FSx File Gateway se utilizan con mayor frecuencia con las cargas de trabajo de Microsoft.

Puerta de enlace de archivos Amazon S3

Amazon S3 File Gateway le permite almacenar sus archivos en Amazon S3 y, al mismo tiempo, proporciona acceso a sus usuarios mediante el uso de recursos compartidos SMB tradicionales. Esto proporciona una interfaz de usuario familiar y ayuda a reducir los costes al almacenar los datos en Amazon S3 y aprovechar los distintos niveles de almacenamiento de Amazon S3. Puede implementar Storage Gateway con S3 Intelligent Tiering para ayudarlo a mover automáticamente los archivos del ciclo de vida a los niveles de almacenamiento de menor costo para reducir aún más sus costos. Recomendamos S3 File Gateway para ofrecer escalabilidad horizontal y acceso de solo lectura, lecturas rápidas y repetidas (desde la memoria caché) y volcados de bases de datos. Por lo general, no se recomienda para escrituras, edición de archivos o comparticiones departamentales de alto rendimiento o alta disponibilidad.

Amazon FSx File Gateway

Amazon FSx File Gateway también puede ofrecer un ahorro de costes al trabajar con los sistemas de archivos de Amazon FSx Windows. Puede instalar un FSx File Gateway para proporcionar acceso localizado a un sistema de FSx archivos de Amazon en otra región para evitar los costes de tener dos sistemas de archivos independientes. Esto también puede resultar útil si tiene varios servidores de archivos locales y desea consolidarlos para evitar tener que pagar por varios dispositivos de hardware.

Impacto del costo

Puerta de enlace de archivos Amazon S3

Configurar S3 File Gateway es fácil porque puede usar el asistente de inicio de Storage Gateway. Puede implementar la puerta de enlace en cuestión de minutos mediante una EC2 instancia de su AWS entorno. Una vez configurada la puerta de enlace, puede configurar los recursos compartidos de Storage Gateway para que sean accesibles a través de los protocolos SMB y NFS. Para las cargas de trabajo típicas de Windows, también puede usar esta configuración para aprovechar un

entorno de Active Directory y establecer permisos en los recursos compartidos de archivos. Puede integrar eficazmente un Storage Gateway en su uso normal, ya que funcionará como un recurso compartido de archivos típico de Windows. Los archivos y las carpetas se almacenan como objetos y las listas de control de acceso NTFS (ACLs) como metadatos.

En la siguiente tabla se comparan los costes de 10 TB de almacenamiento con las tres opciones de almacenamiento disponibles:

- FSx para Windows File Server
- Puerta de enlace de archivos Amazon S3
- Amazon Elastic Block Store (Amazon EBS)

El precio de almacenar 10 TB de almacenamiento es considerablemente más económico si utilizas Amazon S3, ya que puedes particionar tus datos en varios niveles de uso. En la estimación de precios, se utiliza S3 Intelligent Tiering por su flexibilidad de precios. Esto incluye el 80 por ciento en S3 Standard, el 10 por ciento en Infrequent Access y el 10 por ciento en Amazon S3 Glacier. Si bien puede utilizar S3 Glacier, es importante establecer las reglas de ciclo de vida adecuadas para garantizar que no sea necesario acceder inmediatamente a los archivos que se trasladen a S3 Glacier. S3 Glacier es exclusivamente para uso de archivos, no para un uso de acceso regular.

Sistemas de almacenamiento	Coste de 10 TB de almacenamiento	Región
FSx para Windows File Server (suponiendo un ahorro del 50% en la deduplicación)	USD 683,20 USD en SSD	Este de EE. UU. (Norte de Virginia)
Puerta de enlace de archivos Amazon S3	Organización inteligente en niveles por 449,51$	Este de EE. UU. (Norte de Virginia)
Amazon EBS	1.335,69 USD GP3	Este de EE. UU. (Norte de Virginia)

Considere lo siguiente:

- En S3 Glacier, recibe errores de E/S genéricos a menos que utilice la RestoreObjectAPI para restaurar el objeto en Amazon S3. Le recomendamos que utilice una notificación para este error

de E/S mediante Amazon CloudWatch Events. De esta forma, su equipo de operaciones podrá reaccionar si un usuario recibe este error en un archivo al que podría necesitar acceder. Para obtener más información sobre estos errores, consulte Error: InaccessibleStorageClass en la documentación de Amazon S3 File Gateway.

- Además de la limitación de acceso de S3 Glacier, solo se ACLs permiten 10 por objeto o carpeta en Storage Gateway. Antes de decidir usar Storage Gateway, asegúrese de no necesitar más de 10 entradas de ACL.

Amazon FSx File Gateway

Al igual que una puerta de enlace de archivos de Amazon S3, una puerta de enlace de FSx archivos proporciona acceso a un sistema de archivos que conserva los datos a largo plazo. En Amazon S3 File Gateway, los datos residen en Amazon S3. En el caso de FSx File Gateway, sus datos residen en FSx el servidor de archivos de Windows. Si bien las opciones Multi-AZ están disponibles FSx para Windows File Server, no hay ninguna opción multirregional. Si tiene una empresa global o una oficina remota, es posible que necesite proporcionar una plataforma de almacenamiento compartido que esté geográficamente más cerca del usuario final para evitar la latencia. Si tuviera que implementar otro sistema de FSx archivos de Amazon, esto añadiría el coste de un sistema de archivos Amazon FSx for Windows File Server completamente nuevo y el almacenamiento necesario. Para evitar la creación de un sistema de archivos completamente nuevo y la duplicación de costes, puede implementar FSx File Gateway en la región secundaria. Esto proporciona a los usuarios acceso localizado a los archivos y, al mismo tiempo, ayuda a reducir los costos generales.

Sistemas de almacenamiento	Coste de 10 TB de almacenamiento	Región
Servidor FSx de archivos Amazon para Windows	USD 683,20 USD EN SSD	Este de EE. UU. (Norte de Virginia)
Amazon FSx File Gateway	503,70 dólares/Puerta de enlace única	Este de EE. UU. (Norte de Virginia)

> **Note**
>
> Los precios de la tabla anterior se basan en los precios de Storage Gateway.

Tenga en cuenta lo siguiente:

- FSx File Gateway puede ayudarle a ahorrar aproximadamente 180$ al mes (o 2100$ al año) en cargas de trabajo multirregionales.
- Los gastos de transferencia de datos son mucho más bajos con FSx File Gateway, ya que solo necesita almacenar en caché los archivos a los que se accede con regularidad y no una copia secundaria completa.
- Aunque puede tener dos implementaciones de FSx Windows File Server en distintas regiones y mantenerlas actualizadas con AWS Backup o AWS DataSync, ninguna de las dos opciones es prácticamente en tiempo real.

Recomendaciones de optimización de costos

Puerta de enlace de archivos Amazon S3

S3 File Gateway ofrece una opción económica para almacenar archivos, pero hay que tener en cuenta algunos aspectos relacionados con la forma en que implementa y utiliza el sistema de archivos. Por ejemplo, S3 File Gateway requiere el uso de una máquina virtual para ejecutar el software Storage Gateway. En AWS, Storage Gateway se implementa en Amazon EC2 mediante una instancia m5.xlarge de forma predeterminada. Si desea reducir los costos de almacenamiento local, puede implementar Storage Gateway como un dispositivo virtual en plataformas de virtualización como VMware Hyper-V.

Consideraciones sobre la alta disponibilidad

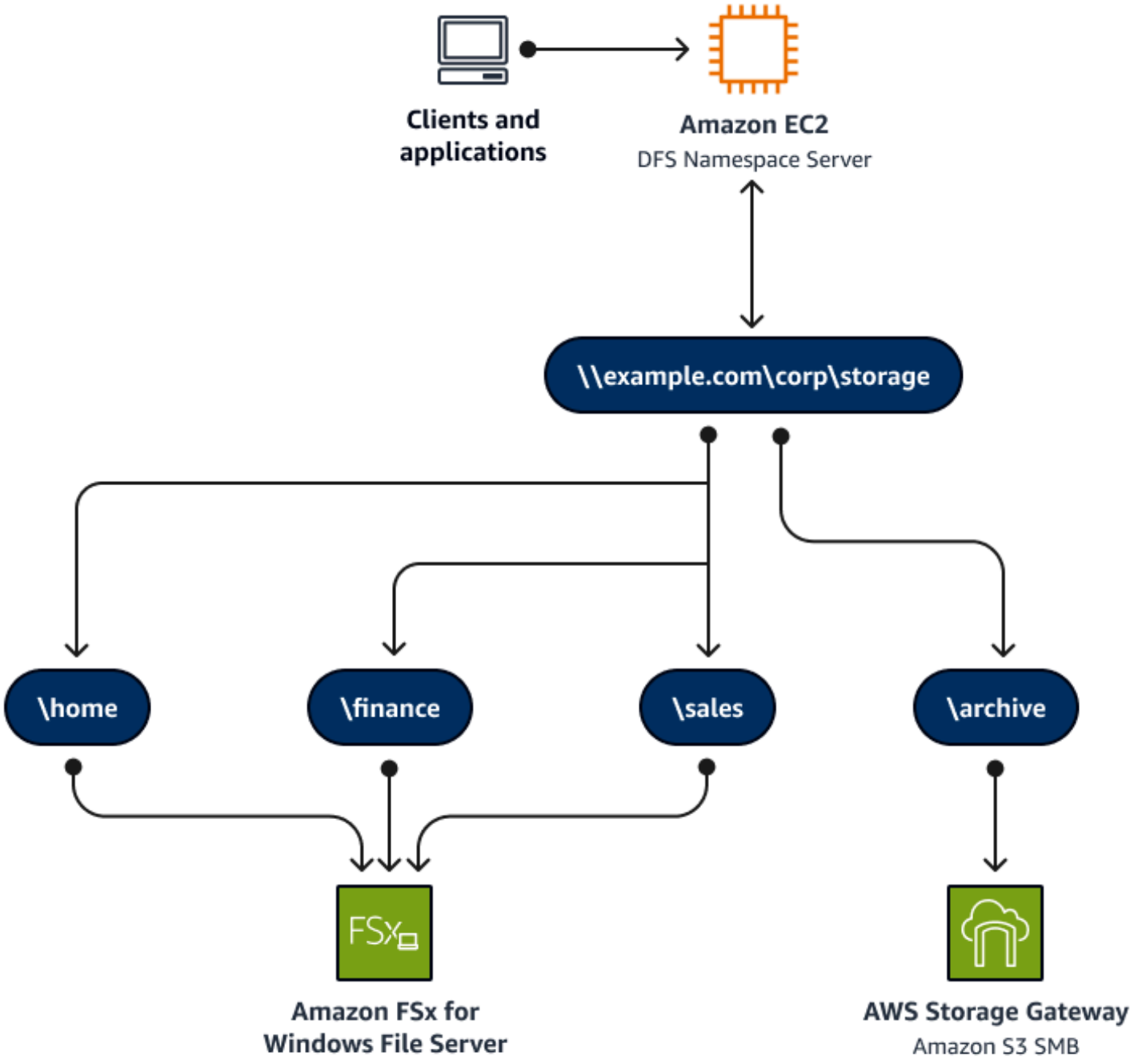
La ejecución de Storage Gateway es un único punto de error para acceder a los archivos. Para evitar tiempos de inactividad innecesarios, le recomendamos que implemente un control de acceso estricto que permita a los usuarios realizar cambios o detener e iniciar la instancia de Storage Gateway. Además, para las implementaciones en marcha AWS, resulta beneficioso utilizar Amazon Data Lifecycle Manager para crear instantáneas de enrutamiento que permitan recuperar rápidamente la implementación de Storage Gateway. Si ejecuta Storage Gateway de forma local VMware, puede configurarlo para una alta disponibilidad.

Ejecuta varios sistemas de archivos

Separar las cargas de trabajo de archivos de uso diario de las cargas de trabajo de archivado puede ayudarle a evitar costes de almacenamiento innecesarios. Storage Gateway tiene la capacidad de implementarse junto con un sistema de archivos FSx para Windows File Server. Al usar los espacios

de nombres DFS, puede presentar su almacenamiento principal de uso diario que se ejecuta en Windows File Server y su almacenamiento que se ejecuta en FSx Amazon S3 (al que se accede a través de Storage Gateway).

En el siguiente diagrama, se muestra cómo se puede utilizar un único espacio de nombres DFS como punto de acceso frontend para diferentes opciones de almacenamiento de backend.



Los clientes se dirigen a una estructura de carpetas, como\\ example.com\ storage. Este directorio principal contiene los subdirectorios. Un sistema de archivos de servidor de archivos FSx para Windows contiene los recursos compartidos de archivos a los que se accede de forma normal. Puede usar un recurso compartido de archivos creado en Storage Gateway para archivar datos. Los usuarios pueden archivar los elementos manualmente en la carpeta de archivado o usted puede crear un proceso para automatizar el traslado de algunos archivos de los recursos compartidos habituales a la carpeta de archivado.

Considere lo siguiente:

- Revise sus requisitos de almacenamiento y proporcione el almacenamiento adecuado para la memoria caché.
- Añada la puerta de enlace a la configuración de Active Directory y utilice Windows estándar ACLs para acceder a los archivos.

FSx File Gateway

La implementación de FSx File Gateway es similar a la implementación de S3 File Gateway, pero es aún más fácil si usa el asistente de inicio. Para obtener instrucciones detalladas, consulte el paso 3: Crear y activar un Amazon FSx File Gateway en la documentación de Amazon FSx File Gateway. Tras implementar FSx File Gateway en su entorno, podrá asociarlo a sus sistemas de FSx archivos de Amazon existentes y obtener acceso a sus archivos.

El almacenamiento es la consideración principal a la hora de implementar FSx File Gateway. El almacenamiento predeterminado proporciona 150 GB, que es una cantidad de espacio decente para almacenar archivos en caché. La creación de alertas de supervisión cuando haya poco espacio libre puede ayudar a ajustar el tamaño del almacenamiento sin una sobreasignación.

Recursos adicionales

- AWS Storage Gateway recursos (documentación)AWS

Active Directory

Amazon Elastic Compute Cloud (Amazon EC2) que ejecuta Windows Server es un entorno seguro, fiable y de alto rendimiento para implementar aplicaciones y cargas de trabajo basadas en Windows. Puede aprovisionar instancias rápidamente y ampliarlas o reducirlas según sea necesario, pagando solo por lo que utilice. Los servicios de Active Directory se utilizan como fuente principal de administración de identidades en los entornos de Windows Server.

Esta sección abarca los siguientes temas:

- Active Directory autogestionado en Amazon EC2
- AWS Managed Microsoft AD
- Conector de AD

Active Directory autogestionado en Amazon EC2

Descripción general

En esta sección se proporcionan recomendaciones para reducir el costo de ejecutar Active Directory en Amazon Elastic Compute Cloud (Amazon EC2). El objetivo principal es asegurarse de que puede dimensionar los controladores de dominio de Active Directory de forma adecuada y utilizar la flexibilidad que ofrecen Nube de AWS para ajustarlos según sea necesario a su entorno. AWS puede ayudarle a detener fácilmente una instancia y redimensionarla para adaptarla a sus necesidades cambiantes, o a reducir el tamaño de la instancia si la amplía demasiado rápido. Elegir el tamaño y el tipo de instancia correctos puede generar ahorros significativos.

Impacto del costo

En la siguiente tabla, se muestra la diferencia entre elegir una instancia de la familia de instancias con fragmentación y una instancia de uso general. Esta elección puede ahorrarle una cantidad considerable de dinero cada mes. La planificación y el tamaño adecuados de la instancia pueden ayudarle a gestionar los costes.

Tipo de instancia	Número de instancias	vCPU	Memoria	Costo
t3a.medium	2	2.	8	81,76$ al mes

Tipo de instancia	Número de instancias	vCPU	Memoria	Costo
m5a.large	2	2.	8	259,88$ al mes

Para obtener más información sobre los costos, consulta la estimación. Calculadora de precios de AWS

Un ahorro de 178,12$ al mes acaba siendo más de 2.000$ de ahorro al año para sus controladores de dominio. Tenga en cuenta que se trata de un espacio reducido de tan solo dos controladores de dominio en una cuenta. Si se escalan con varias cuentas y controladores de dominio adicionales, estos ahorros pueden suponer una importante reducción de costes.

Recomendaciones de optimización de costos

Microsoft proporciona recomendaciones de planificación de la capacidad para implementar su entorno de Active Directory. Le recomendamos que tenga en cuenta los siguientes componentes principales al planificar o escalar su entorno de Active Directory:

- Memoria
- Network
- Almacenamiento
- Procesador

Teniendo en cuenta estos componentes principales, puede seleccionar un tipo de instancia que se adapte a su entorno de Active Directory AWS. En esta sección se describen algunos ejemplos de escenarios de AWS implementación de Active Directory. Estos escenarios dejan claro que no es necesario replicar el entorno local si no se planea gestionar el mismo número de usuarios y ordenadores que en el entorno local. AWS

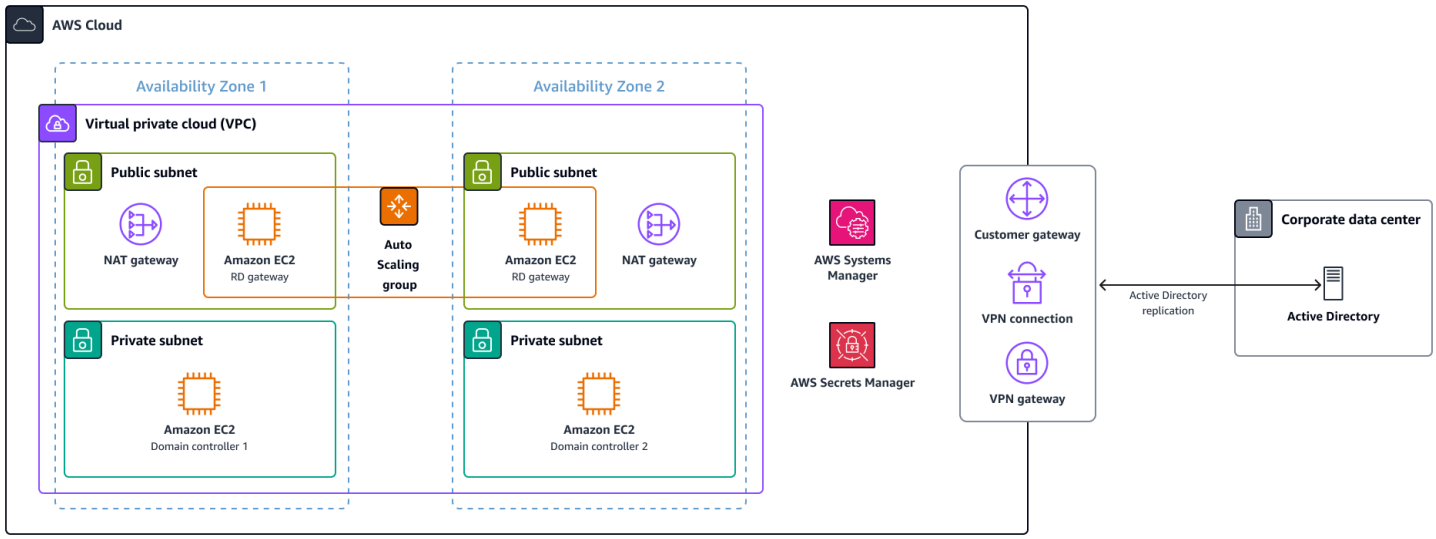
En la siguiente tabla, se destacan los componentes importantes relacionados con la vCPU, la memoria y el disco para su AWS espacio físico.

Componente	Estimaciones
Tamaño de almacenamiento/base de datos	40 a 60 KB para cada usuario

Componente	Estimaciones
RAM	Tamaño de base de datos Recomendaciones básicas sobre el sistema operativo Aplicaciones de terceros
Network	1 GB
CPU	1000 usuarios simultáneos para cada núcleo

Escenario de despliegue híbrido

El siguiente diagrama muestra un ejemplo de arquitectura para una implementación híbrida de Active Directory.



Como se muestra en el diagrama, normalmente se tiene un espacio local y, a continuación, se amplía al Nube de AWS. En las fases iniciales de una migración, normalmente no tendrá todos los usuarios y servidores desplegados AWS. Por eso es importante implementar inicialmente un espacio más pequeño para ahorrar dinero en los esfuerzos de migración.

Si va a mantener un entorno local en el que los servidores y los usuarios se autentiquen de forma local, no necesitará el mismo espacio para los controladores de dominio. AWS Si sigue las prácticas recomendadas de Active Directory, puede implementar los sitios y servicios de Active Directory

adecuados para autenticar a los usuarios y los equipos en su entorno local y, al mismo tiempo, autenticar únicamente su AWS espacio en los controladores de dominio. AWS Esto le permite evitar sobredimensionar el espacio de Active Directory AWS al limitar el uso únicamente a los AWS recursos y no a toda la infraestructura local. Para obtener orientación sobre el diseño de una configuración híbrida, consulte la ubicación adecuada de los controladores de dominio y las consideraciones del sitio en la documentación de Microsoft.

Optimice para una AWS migración ajustando el tamaño

Si va a implementar una nueva instancia de Active Directory para sus usuarios o planea migrar completamente a ella AWS para su infraestructura de Active Directory, le recomendamos que planifique el tamaño según las recomendaciones de Microsoft en cuanto a vCPU, memoria y espacio en disco para las instancias seleccionadas en la tabla anterior.

Si se trata de un espacio nuevo, puede empezar poco a poco y aprovechar la posibilidad de cambiar fácilmente los tipos de instancias para cambiar el tamaño del entorno a medida que crece. AWS La EC2 sección Windows en Amazon de esta guía le muestra cómo monitorear y revisar el uso de la CPU y la memoria en AWS. De esta forma, sabrá cuándo aumentar el tamaño de la EC2 instancia.

Si vas a migrar por completo tu entorno local de Active Directory a un entorno de Active Directory AWS, puedes implementar los mismos planes de tamaño para garantizar un rendimiento adecuado. Antes de duplicar lo que tiene en las instalaciones AWS, le recomendamos que realice una revisión exhaustiva de su entorno de Active Directory. Esto puede ayudarle a evitar el sobreaprovisionamiento. Asegúrese de usar el Monitor de rendimiento para recopilar información sobre la cantidad de tráfico y la utilización de sus controladores de dominio actuales. Esto le permitirá comprender el uso general para poder ajustar el tamaño y, en última instancia, reducir sus costos.

Optimice Active Directory en AWS

Si está ejecutando Active Directory AWS, también es importante que supervise continuamente la utilización y que cambie el tamaño de las instancias según sea necesario para reducir sus gastos. Puede AWS Compute Optimizer utilizarlos para obtener información sobre los recursos que utiliza AWS. Para obtener información sobre el uso de Compute Optimizer para ajustar el tamaño de tus cargas de trabajo de Windows, consulta la EC2 sección Windows en Amazon de esta guía. Para obtener un análisis más exhaustivo, puedes usar Performance Monitor para monitorear el uso de tus controladores de dominio de Active Directory, evaluar el rendimiento y, luego, cambiar el tamaño en consecuencia.

También se puede utilizar CloudWatch para supervisar el rendimiento de los controladores de dominio. Para optimizar tus controladores de dominio (ampliándolos o reduciéndolos), puedes usar las métricas disponibles CloudWatch para ayudarte a tomar las decisiones correctas. Puede usar el CloudWatch agente para configurar métricas personalizadas del Monitor de rendimiento que se enviarán para la recopilación de datos. Para obtener instrucciones, consulte ¿Cómo puedo usar el CloudWatch agente para ver las métricas del Monitor de rendimiento en un servidor Windows? en el Centro de AWS conocimiento.

Tras implementar el CloudWatch agente, puede configurar las siguientes métricas en el archivo de configuración del agente que se encuentra en`metrics_collected`:

Categoría métrica	Nombre de métrica
De base de datos a instancias (NTDSA)	% de aciertos en la caché de base
La base de datos de E/S lee la latencia media	
Lecturas de bases de datos de E/S por segundo	
Latencia media de escritura en el registro de E/S	
DirectoryServices (NTDS)	Tiempo de enlace de LDAP
Operaciones de replicación pendientes del DRA	
Sincronizaciones de replicación pendientes de DRA	
DNS	Consultas recursivas por segundo
Error de consulta recursiva/segundo	
Consulta TCP recibida/segundo	
Número total de consultas recibidas/segundo	
Total de respuestas enviadas/segundo	

Categoría métrica	Nombre de métrica
Consulta UDP recibida/segundo	
LogicalDisk	Longitud media de la cola de discos
% de espacio libre	
Memoria	% de bytes comprometidos en uso
Duración media de la memoria caché en espera a largo plazo	
Interfaz de red	Bytes enviados por segundo
Bytes recibidos por segundo	
Ancho de banda actual	
NTDS	Retraso de cola estimado por ATQ
Latencia de solicitud de ATQ	
Lecturas del directorio DS por segundo	
Búsquedas en el directorio DS por segundo	
El directorio DS escritura/segundo	
sesiones de cliente LDAP	
Búsquedas LDAP por segundo	
El LDAP se enlaza satisfactoriamente por segundo	
Procesador	% de tiempo de procesador
Estadísticas de todo el sistema de seguridad	Autenticaciones Kerberos
Autenticaciones NTLM	

Recursos adicionales

- Active Directory Domain Services en AWS: Guía de implementación de soluciones para socios (documentación)AWS
- Planificación de la capacidad de los servicios de dominio de Active Directory (documentación de Microsoft)
- Consideraciones de diseño para ejecutar Active Directory en EC2 instancias (AWS documentos técnicos)

AWS Managed Microsoft AD

Descripción general

AWS Directory Service for Microsoft Active Directory, también conocido como AWS Managed Microsoft AD, funciona con un Active Directory de Windows Server y es administrado por AWS. Se puede utilizar AWS Managed Microsoft AD para migrar una amplia gama de aplicaciones compatibles con Active Directory a. Nube de AWS AWS Managed Microsoft AD funciona con una variedad de aplicaciones y servicios nativos de Active Directory. También es compatible con aplicaciones y servicios AWS administrados. Si bien el servicio y sus mecanismos de facturación no AWS Managed Microsoft AD cuentan con muchas herramientas de optimización de costes, existen algunos principios de diseño que pueden ayudarle a mantener los costes al mínimo.

Impacto del costo

Dado que AWS Managed Microsoft AD se trata de un servicio gestionado basado en el presente SKUs, el dimensionamiento es un proceso relativamente sencillo. Actualmente hay dos tamaños SKUs disponibles: las ediciones Standard y Enterprise. Otras SKUs incluyen el uso compartido de directorios, la adición de controladores de dominio adicionales (incluidas regiones adicionales) y la transferencia de datos entre regiones.

Recomendaciones de optimización de costos

Existen diferencias entre la Edición AWS Managed Microsoft AD Estándar y la Edición AWS Managed Microsoft AD Empresarial. La Edición Enterprise admite hasta 500 000 objetos de Active Directory, 125 cuentas compartidas (límite limitado) y es compatible con varias regiones. La edición Standard admite hasta 30 000 objetos de Active Directory y cinco cuentas compartidas (el límite es de aproximadamente 30 como máximo) y no es compatible con varias regiones.

Las preguntas que hay que tener en cuenta antes de seleccionar el tipo de directorio son las siguientes:

- ¿Se requiere soporte multirregional?
- ¿Se va a compartir el directorio con más de 30 cuentas?
- ¿El número de objetos de Active Directory será superior a 30 000?

Si la respuesta a alguna de las preguntas anteriores es afirmativa, se requiere la versión Enterprise Edition. Si la respuesta a todas las preguntas es negativa, le recomendamos que comience con la Edición Estándar.

> **Note**
>
> Puede actualizar un directorio de la Edición Estándar a la Edición Enterprise, pero un directorio no se puede degradar. La implementación de la Standard Edition no es un proceso de un solo sentido. Si desea actualizar su directorio a Enterprise Edition, póngase en contacto con AWS.

Si compartes directorios en AWS Managed Microsoft AD Enterprise Edition, hay un coste por cada recurso compartido. Esto es inferior al coste de implementar un directorio en cada cuenta, pero ten en cuenta que los costes de uso compartido pueden aumentar si no se controlan. Le recomendamos que solo comparta directorios con cuentas que contengan Amazon Relational Database Service (Amazon RDS) y FSx Amazon for Windows File Server, ya que solo esos servicios admiten esta función. Tenga en cuenta que tiene la opción de integrar FSx Windows File Server con su Active Directory autogestionado, incluido un. AWS Managed Microsoft AD Si solo FSx se requiere Amazon en otra cuenta, puedes realizar un FSx despliegue autogestionado de Amazon AWS Managed Microsoft AD sin necesidad de compartir el directorio.

Al decidir cuándo implementar controladores de dominio adicionales, tenga en cuenta que solo AWS Managed Microsoft AD admite dos subredes en zonas de disponibilidad independientes en la misma VPC. Agregar controladores de dominio adicionales no permite agregar subredes adicionales. Para determinar si debe agregar controladores de dominio adicionales debido a problemas de rendimiento, consulte las métricas de rendimiento de los controladores de dominio en CloudWatch. Esto le indica si uno o todos los controladores de dominio están sobrecargados. Si determina que solo un controlador de dominio está sobrecargado, agregar controladores de dominio adicionales no aliviará la carga y tendrá que profundizar en las aplicaciones y no en el equilibrio de carga entre

los controladores de dominio disponibles actualmente. Si todos los controladores de dominio se utilizan de forma intensiva, añadir un controlador de dominio adicional podría reducir la carga de los controladores de dominio existentes. Para obtener instrucciones sobre cómo automatizar el escalado, consulte Cómo automatizar el AWS Managed Microsoft AD escalado en función de las métricas de uso en el blog AWS de seguridad.

Si ha ampliado su directorio a varias regiones, le recomendamos que no utilice el directorio compartido NETLOGON o SYSVOL para almacenar archivos. Todos los controladores de dominio replican el contenido de esos recursos compartidos. Al no utilizar los recursos compartidos para el almacenamiento de archivos, los costes de transferencia de datos se reducen al mínimo.

También tiene la opción de suscribirse a un acuerdo empresarial con AWS. Los acuerdos empresariales le ofrecen la opción de personalizar los acuerdos que mejor se adapten a sus necesidades. Para obtener más información, consulte Clientes empresariales.

Recursos adicionales

- AWS Managed Microsoft AD cuotas (AWS Directory Service documentación)
- AWS Directory Service Precios (AWS sitio web)
- Los servicios de dominio de Active Directory están AWS disponibles (AWS documentos técnicos)

Conector de AD

Descripción general

AD Connector es un servicio de proxy que proporciona una forma sencilla de conectar tu Microsoft Active Directory local existente a AWS aplicaciones compatibles, como Amazon WorkSpaces QuickSight, Amazon y una unión de dominios integrada para instancias de Amazon Elastic Compute Cloud (Amazon EC2), sin almacenar en caché ninguna información en la nube. Puede usar AD Connector para agregar una cuenta de servicio a Active Directory. AD Connector elimina la necesidad de sincronización de directorios o el coste y la complejidad de alojar una infraestructura de federación. Si bien AD Connector no cuenta con muchas herramientas de optimización de costes debido a la naturaleza del servicio y sus mecanismos de facturación, puede seguir las recomendaciones de diseño de esta sección para mantener los costes al mínimo.

Impacto del costo

AD Connector es un servicio gestionado basado en un ajuste preestablecido SKUs. Esto hace que el dimensionamiento sea un proceso sencillo. Hay dos tamaños SKUs disponibles: tamaño pequeño y grande. Puede usarlo Calculadora de precios de AWSpara estimaciones de costos relacionadas con AD Connector.

Recomendaciones de optimización de costos

Aparte de los recursos informáticos de fondo, no hay diferencia entre los tamaños de conector pequeños y grandes.

Las cuestiones que hay que tener en cuenta antes de seleccionar el tipo de directorio son las siguientes:

- ¿Hay un gran número (más de 10 000) de usuarios activos que utilizan AWS aplicaciones integradas con el AD Connector?
- ¿El usuario es miembro de muchos grupos anidados, profundos o circulares?

Si la respuesta a ambas preguntas es negativa, le recomendamos que comience con el tamaño pequeño. Si respondes afirmativamente a alguna de las preguntas anteriores, quizás valga la pena considerar una talla grande. Puede empezar con un AD Connector de tamaño pequeño y, si el directorio se deteriora debido al rendimiento, puede solicitar que se actualice el directorio a un tamaño más grande.

> **Note**
>
> Puede actualizar un conector AD de pequeño a grande, pero un conector AD no se puede degradar.

La mayoría de los problemas de rendimiento no están relacionados con el AD Connector, sino con la sobrecarga de los controladores de dominio de Active Directory locales debido a que muchos usuarios son miembros de muchos grupos anidados, profundos o circulares.

También tienes la opción de suscribirte a un acuerdo empresarial con. AWS Los acuerdos empresariales le ofrecen la opción de personalizar los acuerdos que mejor se adapten a sus necesidades. Para obtener más información, consulte Clientes empresariales.

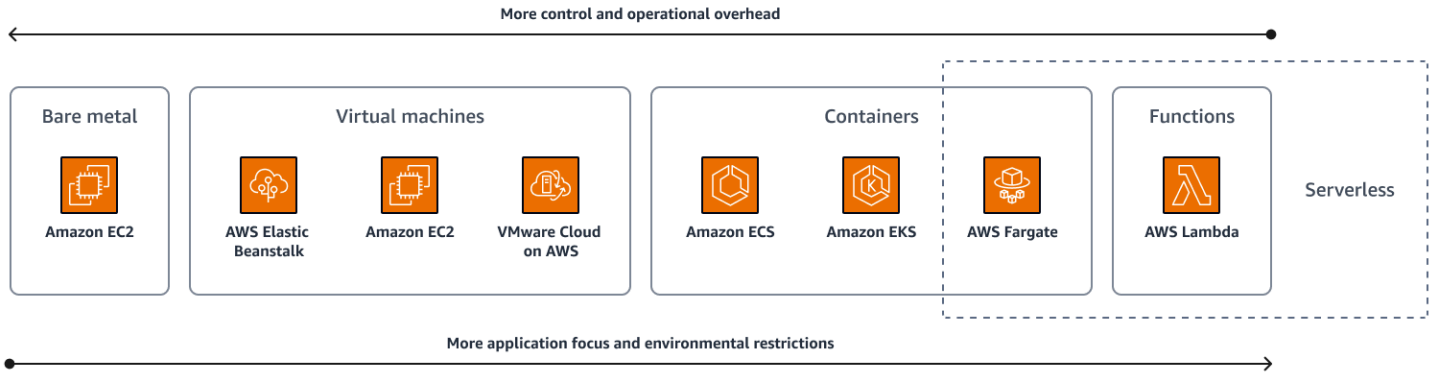
Recursos adicionales

- Cuotas de AD Connector (AWS Directory Service documentación)
- Precios de otros tipos de directorios (AWS sitio web)
- Los servicios de dominio de Active Directory están AWS disponibles (AWS documentos técnicos)

.NET

Desarrollar e implementar aplicaciones.NET es una clave importante para ayudarle a alcanzar la escala y la agilidad que ofrece la computación en nube. Para muchas aplicaciones de .NET heredadas, la opción de procesamiento más adecuada para ejecutar aplicaciones AWS es usar máquinas virtuales, ya sea a través de Amazon Elastic Compute Cloud (Amazon EC2) AWS Elastic Beanstalk o Amazon. También es posible ejecutar aplicaciones.NET en contenedores de Windows y Linux.

La introducción de .NET core le permite diseñar aplicaciones.NET modernas que aprovechan todas las ventajas de la nube. Las aplicaciones modernas pueden utilizar el conjunto tradicional de opciones informáticas y, además, orientarse a varios tipos de entornos sin servidor, incluidos AWS Fargate o AWS Lambda. .NET 6+ ahora ofrece un alojamiento eficiente de cargas de trabajo en ARM64 EC2 instancias como las familias Graviton2. EC2 Esto permite el acceso a la última generación de procesadores disponibles en Amazon EC2. Esto significa que sus aplicaciones se pueden alojar en sistemas informáticos especializados para su tipo de carga de trabajo, como la codificación de vídeo, los servidores web y la informática de alto rendimiento (HPC).



En esta sección se proporcionan recomendaciones que le ayudarán a adaptar sus aplicaciones.NET a fin de aprovechar las ventajas de la nube, centrándose en la rentabilidad.

Esta sección abarca los siguientes temas:

- Refactoriza la versión moderna de.NET y cámbiate a Linux
- Contenedore las aplicaciones.NET
- Utilice instancias y contenedores de Graviton
- Support dinámico para aplicaciones estáticas de.NET Framework
- Utilice el almacenamiento en caché para reducir la demanda de bases de datos

- Considere la posibilidad de utilizar .NET sin servidor
- Considere la posibilidad de crear bases de datos especialmente diseñadas

Refactoriza la versión moderna de.NET y cámbiate a Linux

Descripción general

La modernización de las aplicaciones antiguas de .NET Framework puede ayudarle a mejorar la seguridad, el rendimiento y la escalabilidad. Una forma eficaz de modernizar las aplicaciones de.NET Framework consiste en migrarlas a una versión moderna de.NET (superior a 6). Estas son algunas de las principales ventajas de migrar estas aplicaciones a .NET de código abierto:

- Para reducir los costos de licencias de Windows ejecutándolas en un sistema operativo Linux
- Aproveche la disponibilidad de los lenguajes modernos
- Obtenga un rendimiento optimizado para ejecutarse en Linux

Muchas organizaciones siguen utilizando versiones anteriores de.NET Framework. Esto puede plantear riesgos de seguridad, ya que Microsoft ya no aborda las vulnerabilidades de las versiones anteriores. Microsoft ha dejado de dar soporte a las versiones recientes de.NET Framework 4.5.2, 4.6 y 4.6.1. Es muy importante evaluar los riesgos y beneficios de seguir utilizando las versiones anteriores del marco. Para reducir el riesgo y los costos, puede valer la pena invertir tiempo y esfuerzo en la refactorización a una versión moderna de .NET.

Impacto del costo

Considere un tipo de EC2 instancia de uso general (m5), que ofrezca un equilibrio de recursos de cómputo, memoria y red. Estas instancias son adecuadas para una variedad de aplicaciones, como servidores web, bases de datos medianas y repositorios de código fuente.

Por ejemplo, una instancia m5.xlarge bajo demanda con 4 v CPUs y 16 GB de memoria en Windows Server (licencia incluida) en el este de EE. UU. (Virginia del Norte) cuesta 274,48$ al mes. Los mismos recursos en un servidor Linux cuestan 140,16$ al mes. En este ejemplo, hay una reducción del 49 por ciento en el costo al migrar la aplicación de.NET Framework a una versión moderna de .NET y ejecutar la aplicación en un servidor Linux. El costo puede variar en función de las opciones (por ejemplo, el tipo de instancia, el sistema operativo o el almacenamiento) que elija al seleccionar una EC2 instancia. Puede optimizar aún más los costos mediante Savings Plans o

Reserved Instances. Para obtener más información, utilice el Calculadora de precios de AWSpara realizar estimaciones de costes. En el caso de las instancias incluidas en Windows, el coste de la licencia es de 0,046 USD por vCPU por hora, independientemente del modelo de precios.

La migración de estas aplicaciones.NET Framework a la versión moderna de .NET requiere el esfuerzo del desarrollador. Debe evaluar sus aplicaciones y sus dependencias para comprobar si son compatibles con la versión de la plataforma de destino.AWS El asistente de portabilidad para .NET es una herramienta de asistencia que analiza las aplicaciones de.NET Framework y genera una evaluación de la compatibilidad de.NET, lo que le ayuda a portar sus aplicaciones para que sean compatibles con Linux con mayor rapidez. El asistente de portabilidad para .NET identifica las incompatibilidades con .NET, busca sustitutos conocidos y genera una evaluación de compatibilidad detallada. Tras migrar la solución, debe realizar cambios manuales en el código para que el proyecto se compile correctamente con las dependencias. Esto reduce el esfuerzo manual que implica la modernización de las aplicaciones a Linux. Si su aplicación es compatible con los procesadores ARM, al migrar a Linux podrá utilizar instancias de Graviton. Esto puede ayudarlo a lograr una reducción adicional de costos del 20 por ciento. Para obtener más información, consulte Cómo potenciar .NET 5 con AWS Graviton2: puntos de referencia en el AWS blog sobre informática.

Existen otras herramientas, como el AWS kit de herramientas para la refactorización de.NET y el asistente de actualización de.NET, que pueden ayudarte a migrar aplicaciones antiguas de .NET Framework a la versión moderna de .NET.

Recomendaciones de optimización de costos

Para migrar aplicaciones de.NET Framework, haga lo siguiente:

1. Requisitos previos: para utilizar Porting Assistant para .NET, debe instalar .NET 5+ en la máquina en la que va a analizar el código fuente de la aplicación. Los recursos de la máquina deben tener una velocidad de GHz procesamiento mínima de 1,8, 4 GB de memoria y 5 GB de espacio de almacenamiento. Para obtener más información, consulte los requisitos previos en la documentación del Asistente de portabilidad para.NET.
2. Evaluación: descargue Porting Assistant para .NET como un archivo ejecutable (de descarga). Puede descargar e instalar la herramienta en su máquina para iniciar la evaluación de sus aplicaciones. La página de evaluación contiene proyectos y paquetes portados APIs que son incompatibles con la versión moderna de.NET. Por este motivo, se producen errores de compilación en la solución después de la evaluación. Puede ver o descargar los resultados de la evaluación en un archivo CSV. Para obtener más información, consulte Cómo portar una solución en la documentación del Asistente de portabilidad para .NET.

3. Refactorización: después de evaluar la aplicación, puede migrar sus proyectos a la versión del marco de trabajo de destino. Al migrar una solución, el asistente de portabilidad modificará los archivos del proyecto y parte del código. Puede consultar los registros para revisar los cambios en el código fuente. En la mayoría de los casos, el código requerirá un esfuerzo adicional para completar la migración y las pruebas a fin de que esté listo para la producción. Según la aplicación, algunos de los cambios pueden incluir la estructura de la entidad, la identidad y la autenticación. Para obtener más información, consulte Cómo portar una solución en la documentación del Asistente de portabilidad para .NET.

Este es el primer paso para modernizar sus aplicaciones a contenedores. La modernización de las aplicaciones de.NET Framework y convertirlas en contenedores de Linux podría ser una serie de factores empresariales y técnicos. Uno de los principales impulsores es la reducción del coste total de propiedad al pasar de un sistema operativo Windows a uno Linux. Esto reduce los costos de licencia al migrar la aplicación a una versión multiplataforma de .NET y a contenedores para optimizar la utilización de los recursos.

Una vez que la aplicación se haya migrado a Linux, podrá utilizarla AWS App2Containerpara contenedorizarla. App2Container utiliza Amazon ECS o Amazon EKS como servicios de punto final en los que puede implementarlos directamente. App2Container proporciona toda la infraestructura necesaria como artefactos de despliegue de código (IaC) para contenedorizar sus aplicaciones de forma repetida.

Consideraciones y recursos adicionales

- Si tiene aplicaciones creadas en VB.NET (un marco heredado de 2002) y desea migrarlas a .NET 6, consulte la publicación Cómo portar aplicaciones antiguas de VB.NET a .NET 6.0 con Porting Assistant para.NET en el blog Microsoft Workloads on. AWS
- Si tiene aplicaciones antiguas en Windows Communication Foundation (WCF) y desea ejecutarlas en la versión moderna de.NET, puede adoptar CoreWCF. Para obtener más información, consulte la publicación Modernización de aplicaciones WCF heredadas a CoreWCF mediante Porting Assistant para .NET en el blog Microsoft Workloads on. AWS
- Puede añadir un asistente de portabilidad como una extensión a su IDE de Visual Studio. Esto le permite realizar todas las tareas necesarias para convertir el código sin necesidad de cambiar entre el IDE y la herramienta Porting Assistant para .NET. Para obtener más información, consulte la publicación sobre cómo acelerar la modernización de las aplicaciones.NET con la extensión IDE de Porting Assistant para .NET Visual Studio en el AWS blog Microsoft Workloads on.

- AWS Porting Assistant para .NET es ahora una herramienta de código abierto con los componentes de análisis de código fuente y compatibilidad incluidos en la evaluación. Esto puede animar a sus desarrolladores a utilizar y compartir los conocimientos y las mejores prácticas de portabilidad de .NET.
- Puede migrar aplicaciones de.NET Framework a la versión moderna de.NET en Linux mediante el AWS kit de herramientas para la refactorización de.NET. Para obtener más información, consulte la publicación Acelere la modernización de.NET con el AWS kit de herramientas para la refactorización de.NET en el blog Microsoft Workloads on. AWS
- Puede acelerar la contenedorización y la migración de las aplicaciones de ASP.NET Core para utilizarlas. AWS AWS App2Container

Contenedore las aplicaciones.NET

Descripción general

Los contenedores son una forma ligera y eficiente de empaquetar e implementar aplicaciones de forma coherente y reproducible. En esta sección se explica cómo puede utilizar AWS Fargate un servicio de contenedores sin servidor para reducir los costos de sus aplicaciones.NET y, al mismo tiempo, proporcionar una infraestructura escalable y confiable.

Impacto del costo

Algunos factores que influyen en la eficacia del uso de contenedores para ahorrar costos incluyen el tamaño y la complejidad de la aplicación, la cantidad de aplicaciones que deben implementarse y el nivel de tráfico y demanda de las aplicaciones. En el caso de aplicaciones pequeñas o sencillas, es posible que los contenedores no supongan un ahorro de costes significativo en comparación con los enfoques de infraestructura tradicionales, ya que la sobrecarga que supone gestionar los contenedores y los servicios asociados puede, de hecho, aumentar los costes. Sin embargo, en el caso de aplicaciones más grandes o complejas, el uso de contenedores puede suponer un ahorro de costes al mejorar la utilización de los recursos y reducir la cantidad de instancias necesarias.

Le recomendamos que tenga en cuenta lo siguiente cuando utilice contenedores para ahorrar costes:

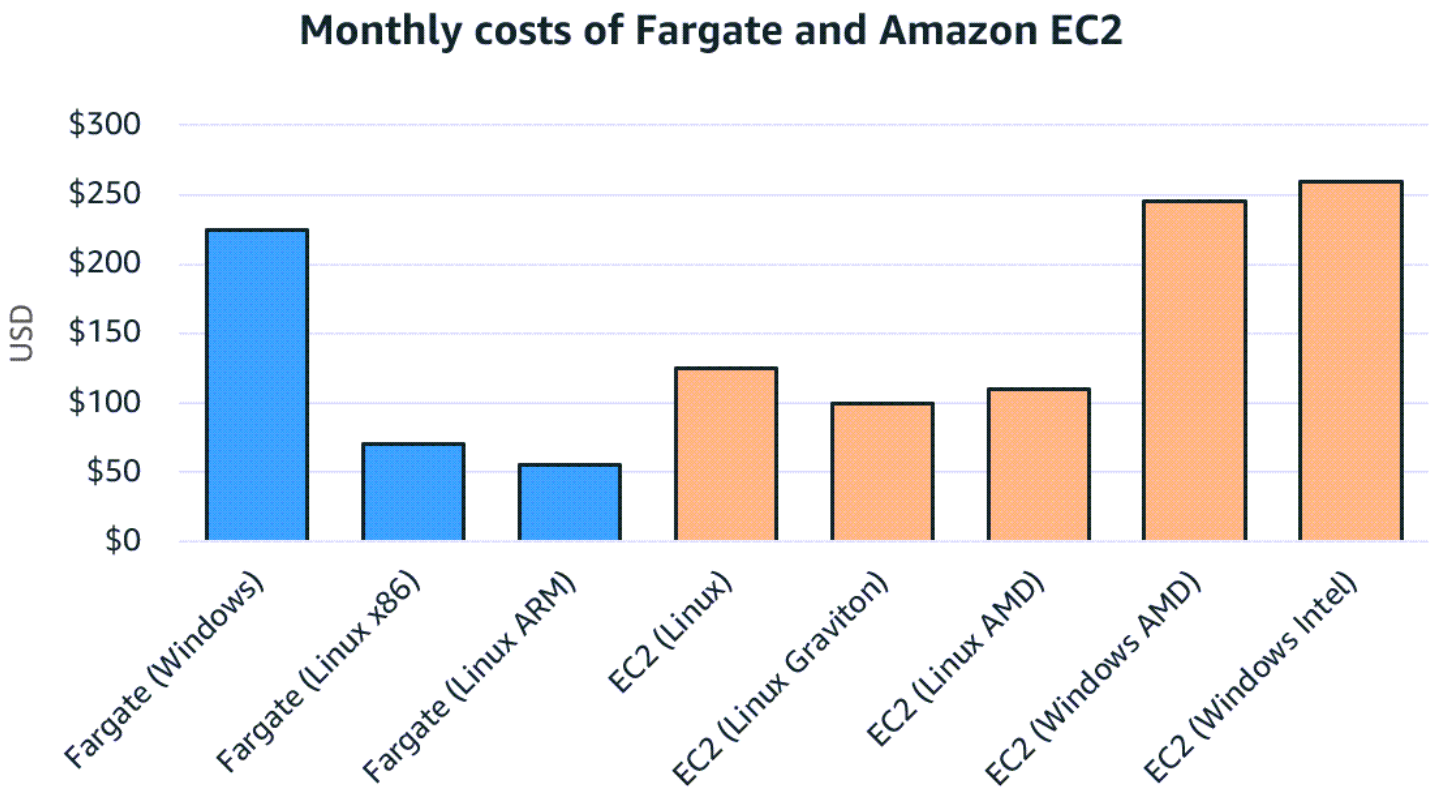
- Tamaño y complejidad de las aplicaciones: las aplicaciones más grandes y complejas son más adecuadas para la contenedorización porque suelen requerir más recursos y pueden beneficiarse más de una mejor utilización de los recursos.

- Cantidad de aplicaciones: cuantas más aplicaciones deba implementar su organización, mayor será el ahorro de costos que se logrará mediante la contenedorización.
- Tráfico y demanda: las aplicaciones que experimentan un tráfico y una demanda elevados pueden beneficiarse de la escalabilidad y la elasticidad que ofrecen los contenedores. Esto puede suponer un ahorro de costes.

Las diferentes arquitecturas y sistemas operativos afectan a los costos de los contenedores. Si utiliza contenedores de Windows, es posible que los costos no disminuyan debido a consideraciones de licencia. Los costos de licencia son más bajos o inexistentes con los contenedores de Linux. AWS Fargate En el siguiente gráfico se utiliza una configuración básica de la región EE.UU. Este (Ohio) con los siguientes ajustes: 30 tareas al mes, ejecutándose durante 12 horas cada una con 4 V CPUs y 8 GB de memoria asignados.

Puede elegir entre dos plataformas informáticas principales para ejecutar sus contenedores AWS: hosts de contenedores EC2 basados y sin servidor o AWS Fargate. Si utilizas Amazon Elastic Container Service (Amazon ECS) en lugar de Fargate, debes mantener el cómputo (instancias) en ejecución para permitir que el motor de colocación cree instancias de los contenedores cuando sea necesario. Si usa Fargate en su lugar, solo se aprovisiona la capacidad de cómputo necesaria.

En el siguiente gráfico se muestra la diferencia entre contenedores equivalentes que utilizan Fargate y Amazon. EC2 Gracias a la flexibilidad de Fargate, las tareas de una aplicación pueden ejecutarse 12 horas al día, sin utilizarlas fuera del horario laboral. Sin embargo, en el caso de Amazon ECS, debe controlar la capacidad informática mediante un grupo de EC2 instancias de Auto Scaling. Esto puede hacer que la capacidad funcione las 24 horas del día, lo que, en última instancia, puede aumentar los costos.



Recomendaciones de optimización de costos

Utilice contenedores de Linux en lugar de Windows

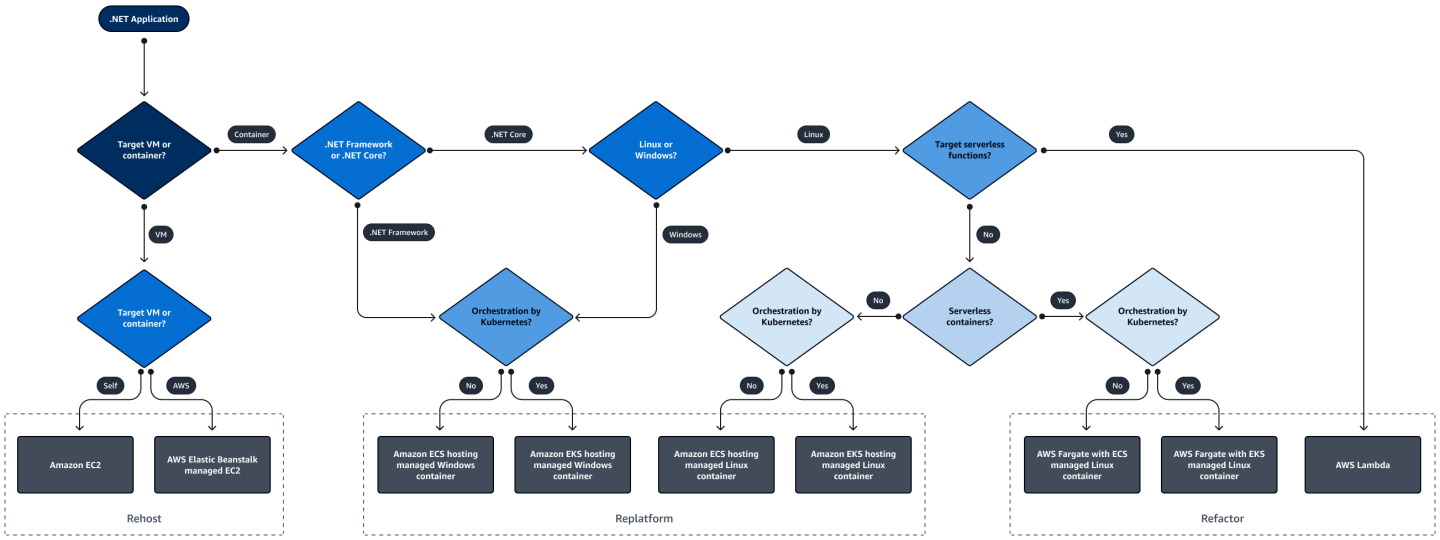
Puede lograr ahorros significativos si utiliza contenedores de Linux en lugar de contenedores de Windows. Por ejemplo, puede ahorrar aproximadamente un 45 por ciento en los costos de procesamiento si ejecuta .NET Core en EC2 Linux en lugar de ejecutar .NET Framework en EC2 Windows. Puede obtener un ahorro adicional del 40 por ciento si utiliza la arquitectura ARM (AWS Graviton) en lugar de la x86.

Si planea ejecutar contenedores basados en Linux para las aplicaciones de .NET Framework existentes, debe migrar estas aplicaciones a versiones modernas y multiplataforma de .NET (por ejemplo, .NET 6.0) para poder utilizar contenedores de Linux. Una consideración importante es sopesar el costo de la refactorización en comparación con los ahorros de costos que se obtienen al reducir el costo de los contenedores de Linux. Para obtener más información sobre la migración de sus aplicaciones a la versión moderna de .NET, consulte el Asistente de portabilidad para .NET en la AWS documentación.

Otra ventaja de migrar a la versión moderna de .NET Framework (es decir, alejarla de la versión de.NET Framework) es que se ofrecen más oportunidades de modernización. Por ejemplo, puede

considerar la posibilidad de rediseñar la arquitectura de su aplicación para convertirla en una arquitectura basada en microservicios que sea más escalable, ágil y rentable.

El siguiente diagrama ilustra el proceso de toma de decisiones para explorar las oportunidades de modernización.



Aproveche los Savings Plans

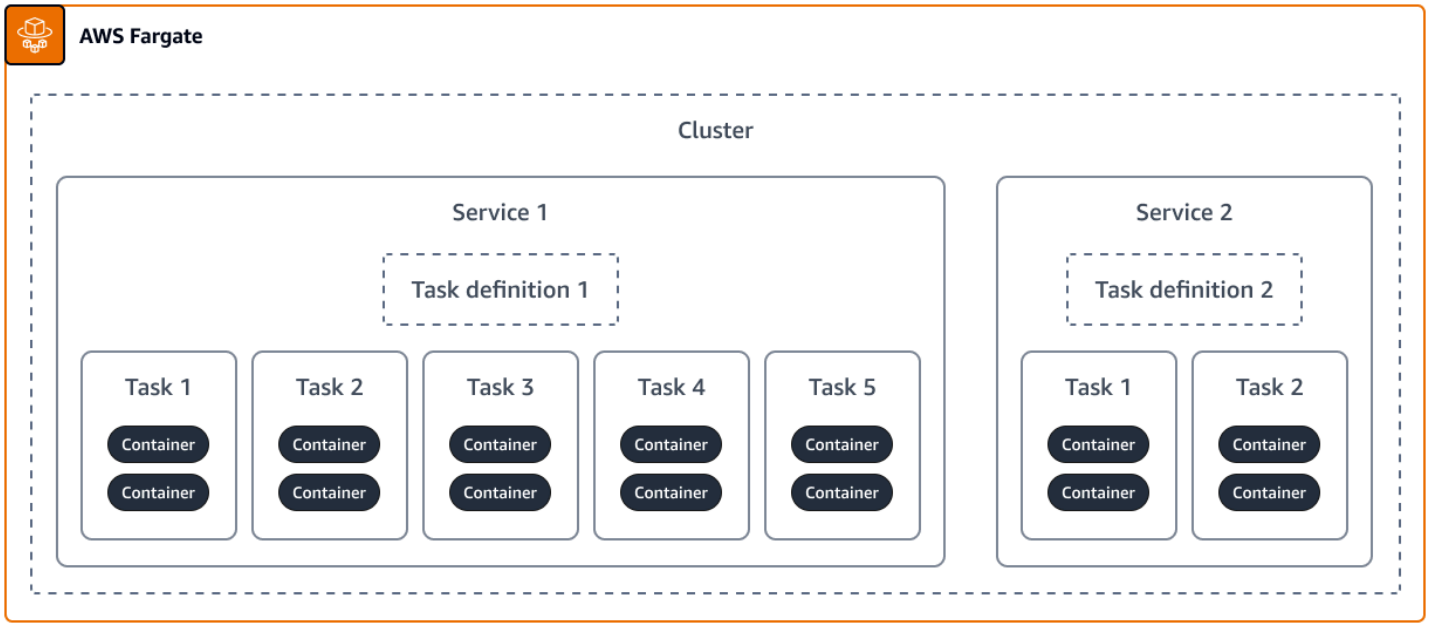
Los contenedores pueden ayudarte a aprovechar Compute Savings Plans para reducir los costos de Fargate. El modelo de descuento flexible ofrece los mismos descuentos que las instancias reservadas convertibles. Los precios de Fargate se basan en los recursos de vCPU y memoria utilizados desde el momento en que se empieza a descargar la imagen del contenedor hasta que finaliza la tarea de Amazon ECS (redondeados al segundo más cercano). Los Savings Plans for Fargate ofrecen ahorros de hasta un 50 por ciento en el uso de Fargate a cambio del compromiso de usar una cantidad específica de uso informático (medida en dólares por hora) durante un período de uno o tres años. Puede usarlo como ayuda AWS Cost Explorerpara elegir un Savings Plan.

Es importante entender que los Compute Savings Plans se aplican primero al uso que te permita ahorrar más. Por ejemplo, si ejecuta una instancia Linux t3.medium `us-east-2` y una instancia t3.medium de Windows idéntica, la instancia Linux recibe primero el beneficio del Savings Plan. Esto se debe a que la instancia de Linux tiene un potencial de ahorro del 50 por ciento, mientras que la misma instancia de Windows tiene un potencial de ahorro del 35 por ciento. Si tiene otros recursos aptos para el Savings Plan Cuenta de AWS, como Amazon EC2 o Lambda, no es necesario que su Savings Plan se aplique primero a Fargate. Para obtener más información, consulta Cómo se aplican los Savings Plans a tu AWS uso en la documentación de Savings Plans y en la EC2 sección Optimizar el gasto para Windows en Amazon de esta guía.

Tareas Fargate del tamaño correcto

Es importante asegurarse de que las tareas de Fargate tengan el tamaño correcto para lograr el máximo grado de optimización de costos. Con frecuencia, los desarrolladores no disponen de toda la información de uso necesaria a la hora de determinar inicialmente las configuraciones de las tareas de Fargate utilizadas en sus aplicaciones. Esto puede provocar un aprovisionamiento excesivo de tareas y, posteriormente, generar gastos innecesarios. Para evitarlo, le recomendamos que cargue las aplicaciones de prueba que se ejecutan en Fargate para comprender cómo funciona la configuración de una tarea específica en diferentes escenarios de uso. Puede utilizar los resultados de las pruebas de carga, la vCPU, la asignación de memoria de las tareas y las políticas de autoescalado para encontrar el equilibrio adecuado entre rendimiento y coste.

En el siguiente diagrama, se muestra cómo Compute Optimizer genera recomendaciones para el tamaño óptimo de la tarea y el contenedor.



Un enfoque consiste en utilizar una herramienta de pruebas de carga, como la que se describe en Distributed Load Testing on AWS, para establecer una línea base para la utilización de la vCPU y la memoria. Después de ejecutar la prueba de carga para simular una carga de aplicación típica, puede ajustar la configuración de la vCPU y la memoria para la tarea hasta alcanzar la utilización de referencia.

Recursos adicionales

- Lista de verificación de optimización de costos para Amazon ECS y AWS Fargate (entrada del blog AWS Containers)
- Optimización teórica de costes mediante Amazon ECS tipo de lanzamiento: Fargate vs. EC2 (entrada del blog de AWS Containers)
- Asistente de portabilidad para .NET (AWS documentación)
- Las pruebas de carga distribuidas están AWS activas (biblioteca de AWS soluciones)
- AWS Compute Optimizer lanza el soporte para los servicios de Amazon ECS en AWS Fargate (entrada del blog AWS Cloud Financial Management)

Utilice instancias y contenedores de Graviton

Descripción general

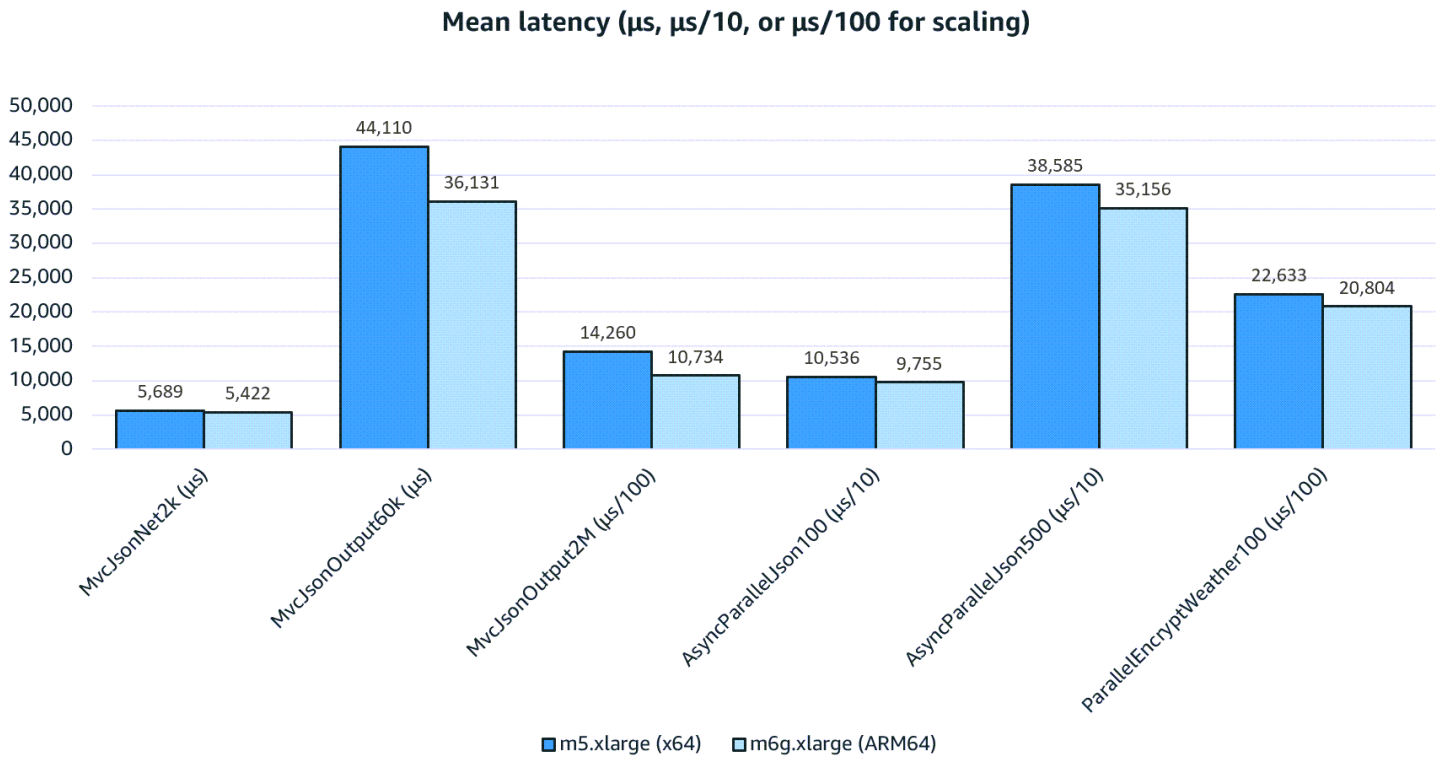
AWS Las instancias Graviton funcionan con procesadores ARM diseñados para ofrecer la mejor AWS relación precio-rendimiento para las cargas de trabajo en la nube que se ejecutan en Amazon Elastic Compute Cloud EC2 (Amazon), incluidos los contenedores que se ejecutan en ellas. AWS Actualmente hay tres generaciones de Graviton disponibles para su uso en Amazon EC2. Esta guía se centra en el uso de Graviton 2 y 3 con aplicaciones.NET, ya que supone un importante ahorro de costes al utilizar las versiones más recientes de Graviton. Tenga en cuenta que las instancias de Graviton solo ejecutan el sistema operativo Linux. Como resultado, las instancias de Graviton son una buena opción para .NET que se ejecuta en Linux, pero no son una opción para el sistema operativo Windows ni para las aplicaciones heredadas de .NET Framework.

Graviton 3 es un 60 por ciento más eficiente que EC2 las instancias comparables, con un rendimiento hasta un 40 por ciento mejor. Esta guía se centra en los beneficios económicos del uso de Graviton, pero es importante tener en cuenta que Graviton ofrece los beneficios adicionales de mejorar el rendimiento y mejorar la sostenibilidad ambiental.

Impacto del costo

Puedes ahorrar hasta un 45 por ciento si te cambias a Graviton. Tras refactorizar cualquier aplicación antigua de .NET Framework para convertirla en una versión moderna de .NET, podrá utilizar instancias de Graviton. La migración a Graviton es una técnica eficaz de optimización de costes para los desarrolladores de.NET.

El ejemplo de la siguiente tabla muestra las posibles mejoras de rendimiento que se pueden lograr al migrar a las instancias de Graviton.



Para obtener un desglose completo y una explicación del enfoque de evaluación comparativa utilizado para crear los resultados del diagrama anterior, consulte Potenciar .NET 5 con AWS Graviton2: puntos de referencia en el blog de informática. AWS

Una de las razones de la mejora de la eficiencia es la diferencia en el significado de vCPU entre x86 y Graviton. En la arquitectura x86, una vCPU es un núcleo lógico que se logra mediante hipersubprocesos. En Graviton, vCPU equivale a un núcleo físico que permite que la vCPU se comprometa plenamente con la carga de trabajo.

El resultado con Graviton2 es una relación precio-rendimiento un 40 por ciento superior a la de las instancias x86/x64 comparables. Graviton3 ofrece lo siguiente en comparación con Graviton2:

- Un perfil de rendimiento mejorado con un rendimiento hasta un 25 por ciento mejor
- Rendimiento de punto flotante hasta dos veces superior
- Rendimiento de las cargas de trabajo criptográficas hasta dos veces más rápido
- Rendimiento de aprendizaje automático hasta tres veces mejor

Además, Graviton3 es la primera instancia en la nube que incluye memoria. DDR5

Las siguientes tablas muestran la diferencia en el ahorro de costes entre las instancias basadas en Graviton y las instancias equivalentes basadas en x86.

En esta tabla se muestran los ahorros de Graviton del 19,20 por ciento.

Tipo de instancia	Arquitectura	vCPU	Memoria (GB)	Coste por hora (bajo demanda)
t4g.xlarge	ARM	4	16	0,1344$
t3.xlarge	x86	4	16	0,1664$

Esta tabla muestra los ahorros de Graviton del 14,99 por ciento.

Tipo de instancia	Arquitectura	vCPU	Memoria (GB)	Coste por hora (bajo demanda)
c7g.4xlarge	ARM	16	32	0,5781$
c6i.4xlarge	x86	16	32	0,6800$

Es importante probar el perfil de rendimiento de la aplicación al considerar Graviton. Graviton no sustituye a las prácticas sólidas de desarrollo de software. Puede utilizar las pruebas para comprobar si está aprovechando al máximo sus recursos informáticos subyacentes.

Recomendaciones de optimización de costos

Hay varias formas de aprovechar los procesadores/instancias Graviton. En esta sección, se explican los cambios necesarios para pasar de utilizar una máquina con arquitectura x86 a instancias Graviton (ARM).

Cambiar la configuración del tiempo de ejecución en Lambda

Se recomienda cambiar la configuración del tiempo de ejecución. AWS Lambda Para obtener más información, consulte Modificación del entorno de ejecución en la documentación de Lambda. Dado

que .NET es un lenguaje compilado, debe seguir un proceso de compilación para que funcione. Para ver un ejemplo de cómo hacerlo, consulte. NET en Graviton in GitHub.

Contenedores

Para una carga de trabajo contenerizada, cree una imagen de contenedor con varias arquitecturas. Para ello, especifique varias arquitecturas en el comando build de Docker. Por ejemplo:

```
docker buildx build -t "myImageName:latest" --platform linux/amd64,linux/arm64 --push .
```

También puedes usar una herramienta como la que te ayude AWS Cloud Development Kit (AWS CDK) a organizar la compilación. Para ver ejemplos de Docker, consulta Cómo crear imágenes de varios arcos para Arm y x86 con escritorios Docker en la documentación de Docker.

Amazon EC2

Para migrar a ARM desde x86/x64, diríjase a la arquitectura ARM en el paso de compilación. En Visual Studio, puede crear una CPU. ARM64 Para obtener instrucciones, consulte Para configurar un proyecto para que se dirija a Arm64 y otras plataformas en la documentación de Microsoft.

Si utiliza la CLI de.NET, al ejecutar la compilación en una máquina ARM se obtiene una compilación compatible con Graviton. Para ver una demostración, vea Acelerar el rendimiento de .NET 6 con Arm64 en AWS Graviton2 o más. YouTube Los problemas de dependencia provocarán errores en el tiempo de compilación que luego se pueden solucionar de forma individual. Mientras haya bibliotecas ARM para cualquier dependencia, la transición debería ser relativamente sencilla.

Recursos adicionales

- Cómo crear sus contenedores para ARM y ahorrar con las instancias Graviton y Spot en Amazon ECS (AWS blog)
- AWS Lambda Funciones impulsadas por el procesador AWS Graviton2: ejecute sus funciones en Arm y obtenga una relación precio-rendimiento hasta un 34% superior (blog)AWS
- Migración de AWS Lambda funciones a procesadores AWS Graviton2 basados en ARM (blog)AWS
- Cree e implemente aplicaciones web.NET en clústeres Amazon ECS de AWS Graviton 2 con tecnología ARM mediante AWS CDK (blog)AWS
- Graviton Fast Start: un nuevo programa que le ayudará a trasladar sus cargas de trabajo a Graviton (blog) AWS AWS

- Cómo potenciar .NET 5 con AWS Graviton2: puntos de referencia (blog)AWS

Support dinámico para aplicaciones estáticas de.NET Framework

Descripción general

Una de las principales ventajas del uso de la nube para las aplicaciones es la elasticidad, es decir, la capacidad de ampliar o reducir el cómputo en función de la demanda. Esto le permite pagar solo por la capacidad de cómputo que necesita, en lugar de aprovisionarla para los picos de uso. El Cyber Monday, en el que los minoristas en línea pueden obtener rápidamente muchas veces más tráfico de lo normal (por ejemplo, miles de veces en cuestión de minutos), es un buen ejemplo de elasticidad.

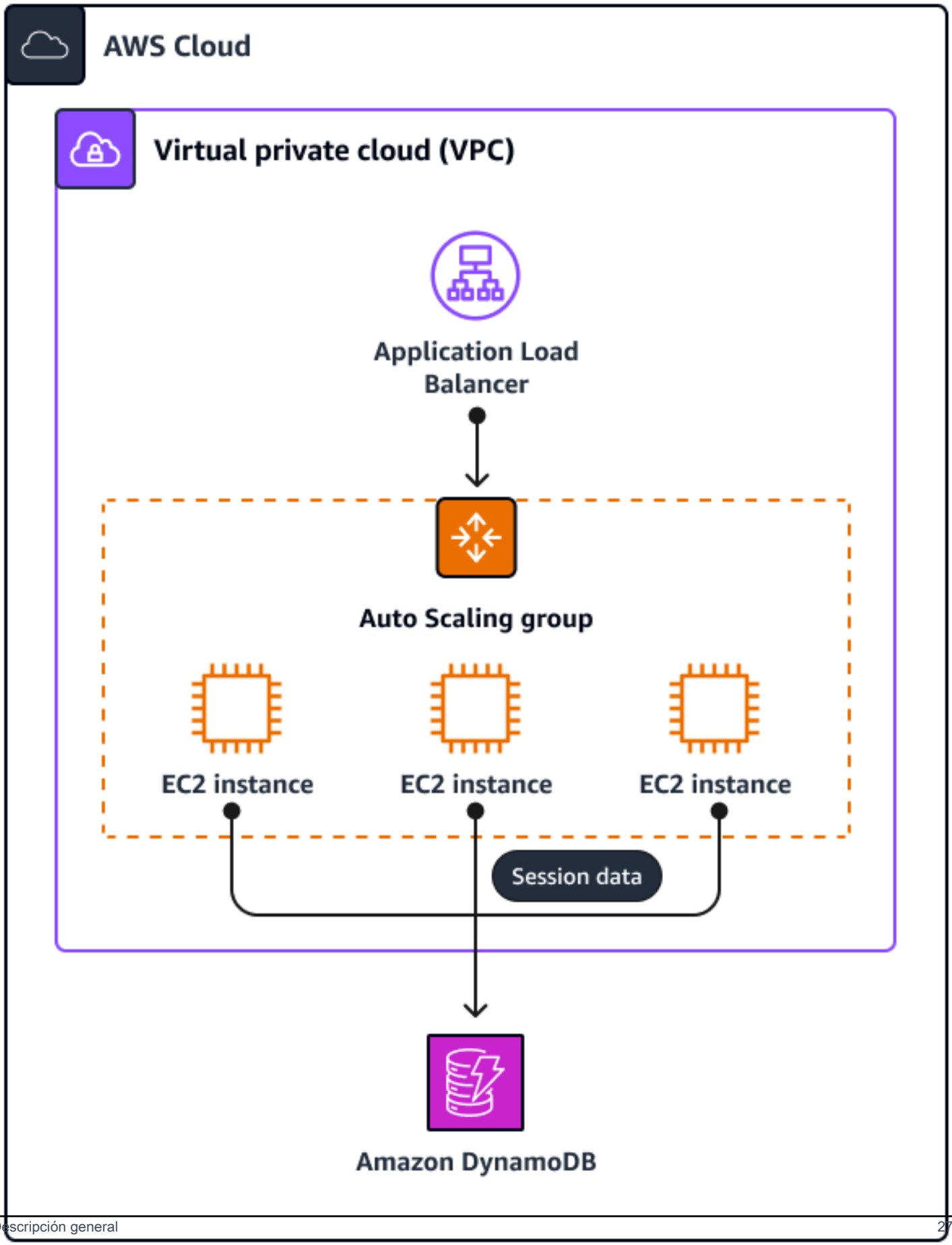
Si va a llevar aplicaciones web antiguas de .NET a la nube (por ejemplo, aplicaciones de ASP.NET Framework que se ejecutan en IIS), la capacidad de escalar rápidamente los conjuntos de servidores con equilibrio de carga puede resultar difícil o imposible debido a la naturaleza activa de la aplicación. Los datos de las sesiones de usuario se almacenan en la memoria de la aplicación, normalmente con variables estáticas o de estado de sesión de ASP.NET que contienen datos de solicitudes cruzadas que deben conservarse. La afinidad entre las sesiones de los usuarios se suele mantener mediante sesiones fijas del equilibrador de carga.

Esto resulta ser un desafío desde el punto de vista operativo. Cuando se requiere una mayor capacidad, debe aprovisionar y añadir servidores de forma intencionada. Este proceso puede ser lento. Dejar los nodos fuera de servicio en caso de que se les aplique un parche o por fallos inesperados puede resultar problemático para la experiencia del usuario final, ya que todos los usuarios asociados a los nodos afectados pierden su estado. En el mejor de los casos, esto requeriría que los usuarios volvieran a iniciar sesión.

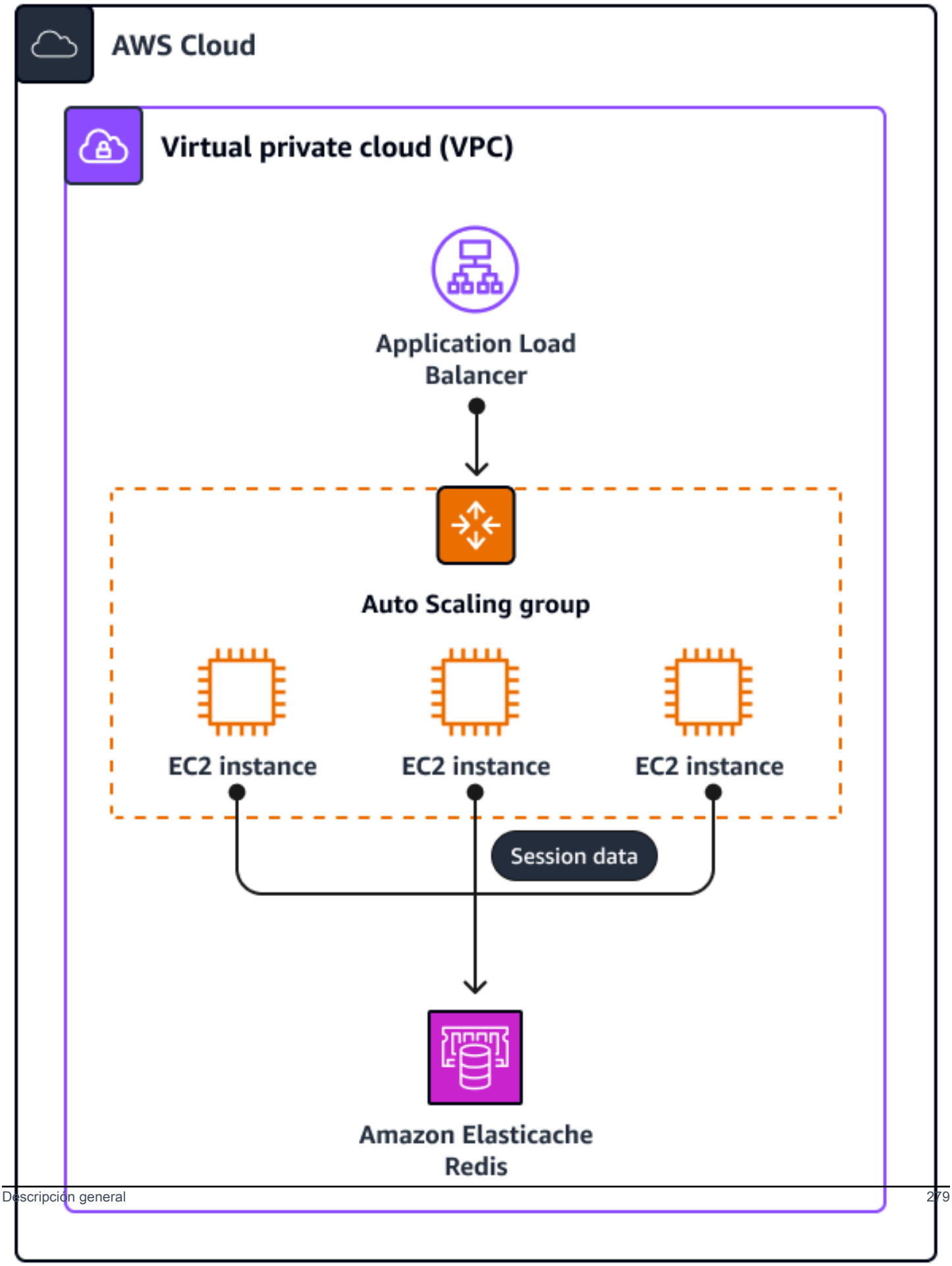
Al centralizar el estado de la sesión de las aplicaciones de ASP.NET y aplicar reglas de escalado automático a las aplicaciones de ASP.NET antiguas, puede aprovechar la elasticidad de la nube y, potencialmente, aprovechar los ahorros de costos al ejecutar las aplicaciones. Por ejemplo, obtiene reducciones de costos gracias a la escalabilidad del procesamiento, pero también puede elegir entre los diferentes modelos de precios disponibles, como reducir el uso de instancias reservadas y utilizar los precios de Amazon Spot Instance.

Dos técnicas habituales incluyen el uso de Amazon DynamoDB como proveedor de estado de sesión y el uso de ElastiCache Amazon (Redis OSS) como almacén de sesiones de ASP.NET.

En el siguiente diagrama se muestra una arquitectura que utiliza DynamoDB como proveedor de estado de sesión.


AWS Cloud
Virtual private cloud (VPC)
Application Load Balancer
Auto Scaling group
EC2 instance
EC2 instance
EC2 instance
Session data
Amazon DynamoDB

El siguiente diagrama muestra una arquitectura que utiliza ElastiCache (Redis OSS) como proveedor de estado de sesión.

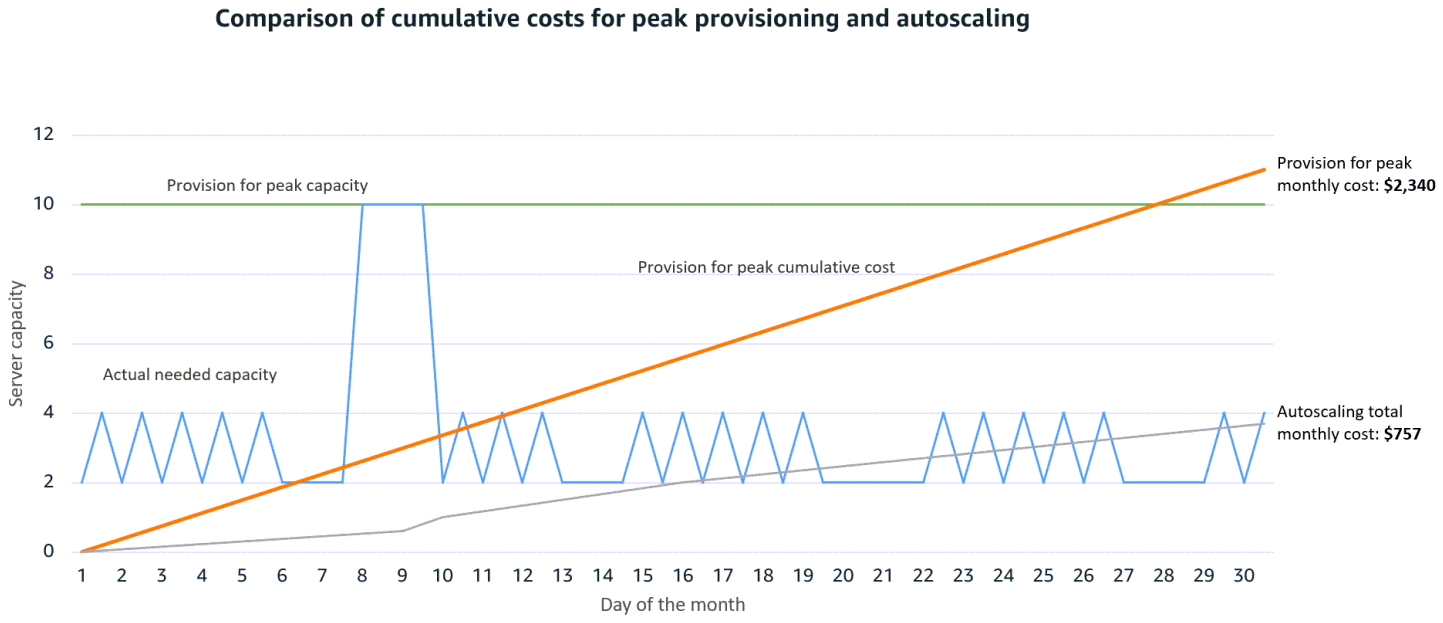

AWS Cloud
Virtual private cloud (VPC)
Application Load Balancer
Auto Scaling group
EC2 instance
EC2 instance
EC2 instance
Session data
Amazon Elasticache Redis

Impacto del costo

Para determinar las ventajas del escalado para una aplicación de producción, le recomendamos que modele su demanda real. En esta sección se hacen las siguientes suposiciones para modelar una aplicación de muestra:

- Las instancias que se añaden y se quitan de la rotación son idénticas y no se introduce ninguna variación en el tamaño de las instancias.
- La utilización del servidor nunca cae por debajo de dos servidores activos para mantener una alta disponibilidad de la aplicación.
- La cantidad de servidores se amplía linealmente con el tráfico (es decir, el doble de tráfico requerirá el doble de procesamiento).
- El tráfico se modela a lo largo de un mes en incrementos de seis horas, con una variación intradía y un pico de tráfico anormal (por ejemplo, una venta promocional) durante un día en el que el tráfico se multiplica por 10. El tráfico de fin de semana se modela en función de la utilización básica.
- El tráfico nocturno se modela en función de la utilización básica, mientras que el tráfico de los días laborables se modela en función del cuádruple de la utilización.
- Los precios de las instancias reservadas se basan en un año, sin necesidad de anticipar. Los precios diurnos normales utilizan los precios bajo demanda, mientras que para la demanda puntual se utilizan los precios de las instancias puntuales.

El siguiente diagrama ilustra cómo este modelo aprovecha la elasticidad de una aplicación.NET en lugar de aprovisionarla para los picos de uso. Esto se traduce en un ahorro de aproximadamente el 68 por ciento.



Si utiliza DynamoDB como mecanismo de almacenamiento del estado de la sesión, utilice los siguientes parámetros:

```
Storage: 20GB
Session Reads: 40 million
Session Writes: 20 million
Pricing Model: On demand
```

El costo mensual estimado de este servicio es de aproximadamente 35 USD al mes.

Si utiliza ElastiCache (Redis OSS) como mecanismo de almacenamiento del estado de la sesión, utilice los siguientes parámetros:

```
Number of Nodes: 3
Node size: cache.t4g.medium
Pricing Model: 1y reserved
```

El costo mensual estimado de este servicio es de aproximadamente 91,00$ al mes.

Recomendaciones de optimización de costos

El primer paso es implementar el estado de la sesión en una aplicación .NET antigua. Si lo utiliza ElastiCache como mecanismo de almacenamiento de estado, siga las instrucciones que aparecen en el blog de herramientas para AWS desarrolladores ElastiCache como almacén de sesiones de

ASP.NET. Si utiliza DynamoDB, siga las instrucciones de ¿Qué es? de AWS SDK para .NET la documentación. SDK para .NET

Si la aplicación utiliza la InProcsesión para empezar, asegúrese de que todos los objetos que planea almacenar en la sesión se puedan serializar. Para ello, utilice el `SerializableAttribute` atributo para decorar las clases cuyas instancias se almacenarán en la sesión. Por ejemplo:

```
[Serializable()]
public class TestSimpleObject {
public string SessionProperty {get;set;}
}
```

Además, el.NET `MachineKey` debe ser el mismo en todos los servidores en uso. Este suele ser el caso cuando las instancias se crean a partir de una Amazon Machine Image (AMI) común. Por ejemplo:

```
<machineKey
validationKey="some long hashed value"
decryptionKey="another long hashed value"
validation="SHA1"/>
```

Sin embargo, es importante asegurarse de que, si se cambia una imagen base, se configure con la misma imagen de máquina.NET (ya sea a nivel de IIS o de servidor). Para obtener más información, consulte SystemWebSectionGroup. MachineKey Propiedad en la documentación de Microsoft.

Por último, debe determinar el mecanismo para agregar servidores a un grupo de Auto Scaling en respuesta a un evento de escalado. Existen varias formas de lograrlo. Recomendamos los siguientes métodos para implementar sin problemas las aplicaciones.NET Framework en una EC2 instancia de un grupo de Auto Scaling:

- Use EC2 Image Builder para configurar una AMI que contenga el servidor y la aplicación completamente configurados. A continuación, puede usar esta AMI para configurar la plantilla de lanzamiento de su grupo de Auto Scaling.
- AWS CodeDeployUtilícela para implementar la aplicación. CodeDeploy permite la integración directa con Amazon EC2 Auto Scaling. Esto proporciona una alternativa a la creación de una AMI nueva para cada versión de la aplicación.

Recursos adicionales

- Creación de imágenes con EC2 Image Builder (documentación EC2 de Image Builder)
- Implementación de aplicaciones web.NET AWS CodeDeploy mediante Visual Studio Team Services (blog AWS sobre herramientas para desarrolladores)

Utilice el almacenamiento en caché para reducir la demanda de bases de datos

Descripción general

Puede utilizar el almacenamiento en caché como una estrategia eficaz para reducir los costes de sus aplicaciones.NET. Muchas aplicaciones utilizan bases de datos de fondo, como SQL Server, cuando las aplicaciones requieren un acceso frecuente a los datos. El costo de mantener estos servicios de backend para hacer frente a la demanda puede ser elevado, pero puede utilizar una estrategia de almacenamiento en caché eficaz para reducir la carga de las bases de datos de backend reduciendo los requisitos de tamaño y escalado. Esto puede ayudarlo a reducir los costos y mejorar el rendimiento de sus aplicaciones.

El almacenamiento en caché es una técnica útil para ahorrar costes relacionados con las cargas de trabajo de lectura intensiva que utilizan recursos más caros, como SQL Server. Es importante utilizar la técnica adecuada para la carga de trabajo. Por ejemplo, el almacenamiento en caché local no es escalable y requiere que mantengas una caché local para cada instancia de una aplicación. Debe sopesar el impacto en el rendimiento en comparación con los costos potenciales, de modo que el menor costo de la fuente de datos subyacente compense cualquier costo adicional relacionado con el mecanismo de almacenamiento en caché.

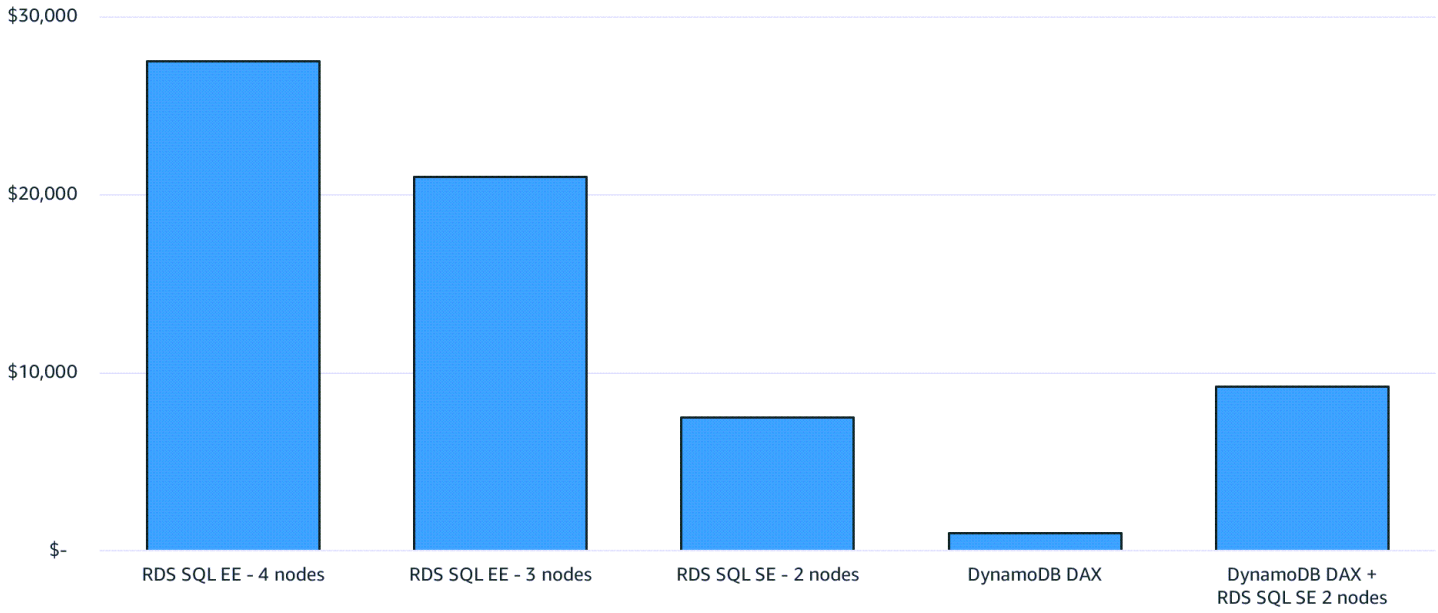
Impacto del costo

SQL Server requiere que tenga en cuenta las solicitudes de lectura al dimensionar la base de datos. Esto podría afectar a los costes, ya que es posible que tenga que introducir réplicas de lectura para adaptarse a la carga. Si utiliza réplicas de lectura, es importante que comprenda que solo están disponibles en la edición Enterprise de SQL Server. Esta edición requiere una licencia más cara que la edición SQL Server Standard.

El siguiente diagrama está diseñado para ayudarle a comprender la eficacia del almacenamiento en caché. Muestra Amazon RDS for SQL Server con cuatro nodos db.m4.2xlarge que ejecutan la

edición SQL Server Enterprise. Se implementa en una configuración Multi-AZ con una réplica de lectura. El tráfico de lectura exclusivo (por ejemplo, las consultas SELECT) se dirige a las réplicas de lectura. En comparación, Amazon DynamoDB utiliza un clúster DynamoDB Accelerator (DAX) r4.2xlarge de dos nodos.

En el siguiente gráfico se muestran los resultados obtenidos al eliminar la necesidad de réplicas de lectura dedicadas que gestionen un tráfico de lectura elevado.



Puede conseguir importantes ahorros de costes mediante el uso del almacenamiento en caché local sin réplicas de lectura o mediante la introducción de DAX junto con SQL Server en Amazon RDS como capa de almacenamiento en caché. Esta capa se descarga de SQL Server y reduce el tamaño del SQL Server necesario para ejecutar la base de datos.

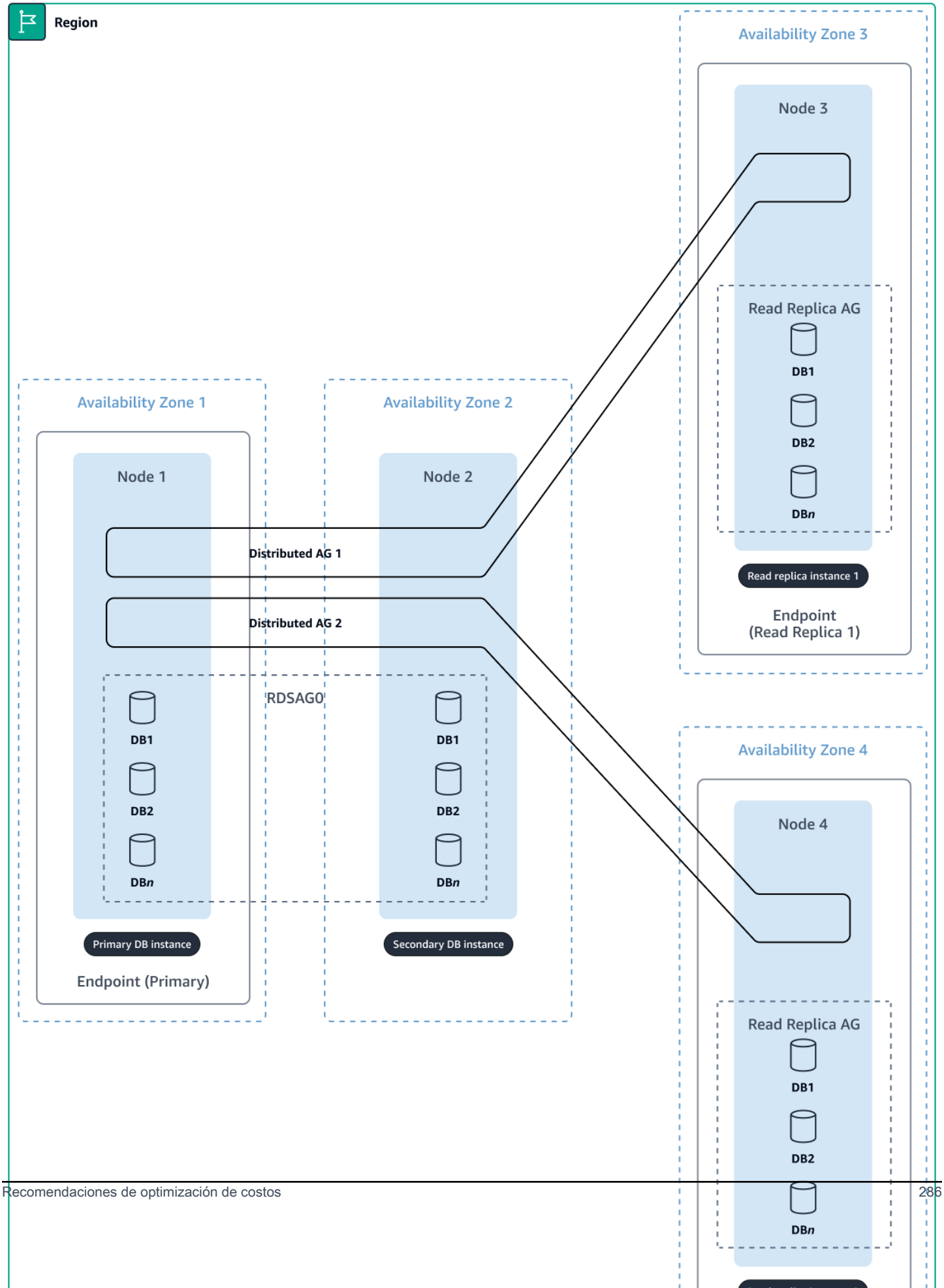
Recomendaciones de optimización de costos

Almacenamiento en la caché local

El almacenamiento en caché local es una de las formas más utilizadas para almacenar en caché el contenido de las aplicaciones alojadas tanto en entornos locales como en la nube. Esto se debe a que su implementación es relativamente fácil e intuitiva. El almacenamiento en caché local implica tomar contenido de una base de datos u otra fuente y almacenarlo en caché local en la memoria o en el disco para un acceso más rápido. Este enfoque, aunque es fácil de implementar, no es ideal para algunos casos de uso. Por ejemplo, esto incluye casos de uso en los que el contenido almacenado en caché debe persistir en el tiempo, como conservar el estado de la aplicación o el estado del

usuario. Otro caso de uso es cuando es necesario acceder al contenido almacenado en caché desde otras instancias de la aplicación.

El siguiente diagrama ilustra un clúster de SQL Server de alta disponibilidad con cuatro nodos y dos réplicas de lectura.


Region
Availability Zone 3
Node 3
Read Replica AG
DB1
DB2
DBn
Read replica instance 1
Endpoint
(Read Replica 1)
Availability Zone 1
Node 1
Distributed AG 1
Distributed AG 2
RDSAG0
DB1
DB2
DBn
Primary DB instance
Endpoint (Primary)
Availability Zone 2
Node 2
DB1
DB2
DBn
Secondary DB instance
Availability Zone 4
Node 4
Read Replica AG
DB1
DB2
DBn

Con el almacenamiento en caché local, es posible que necesite equilibrar la carga del tráfico entre varias EC2 instancias. Cada instancia debe mantener su propia caché local. Si la caché almacena información con estado, es necesario que se realicen confirmaciones periódicas en la base de datos y es posible que sea necesario reenviar a los usuarios a la misma instancia para cada solicitud posterior (sesión fija). Esto supone un desafío cuando se intenta escalar las aplicaciones, ya que algunas instancias pueden estar sobreutilizadas, mientras que otras están infrautilizadas debido a la distribución desigual del tráfico.

Puede utilizar el almacenamiento en caché local, ya sea en memoria o mediante almacenamiento local, para las aplicaciones.NET. Para ello, puede añadir la funcionalidad de almacenar objetos en el disco y recuperarlos cuando sea necesario, o bien consultar datos de la base de datos y conservarlos en la memoria. Para realizar el almacenamiento en caché local en memoria y en el almacenamiento local de datos de un SQL Server en C#, por ejemplo, puede utilizar una combinación de `MemoryCache` bibliotecas y. `LiteDB` `MemoryCache`proporciona almacenamiento en caché en memoria, a la vez que `LiteDB` es una base de datos NoSQL integrada basada en disco que es rápida y ligera.

Para realizar el almacenamiento en caché en memoria, utilice la biblioteca.NET. `System.Runtime.MemoryCache` El siguiente ejemplo de código muestra cómo usar la `System.Runtime.Caching.MemoryCache` clase para almacenar en caché los datos en la memoria. Esta clase proporciona una forma de almacenar datos temporalmente en la memoria de la aplicación. Esto puede ayudar a mejorar el rendimiento de una aplicación al reducir la necesidad de obtener datos de un recurso más caro, como una base de datos o una API.

Así es como funciona el código:

1. Se crea una instancia estática privada `_memoryCache` de `MemoryCache` named. La caché recibe un nombre (`dataCache`) para identificarla. A continuación, la memoria caché almacena y recupera los datos.
2. El `GetData` método es un método genérico que utiliza dos argumentos: una `string` clave y un `Func<T>` delegado llamado`getData`. La clave se usa para identificar los datos en caché, mientras que el `getData` delegado representa la lógica de recuperación de datos que se ejecuta cuando los datos no están presentes en la caché.
3. El método primero comprueba si los datos están presentes en la memoria caché mediante el `_memoryCache.Contains(key)` método. Si los datos están en la memoria caché, el método los recupera `_memoryCache.Get(key)` y convierte los datos en el tipo T esperado.

4. Si los datos no están en la caché, el método llama al `getData` delegado para que los busque. A continuación, agrega los datos a la memoria caché mediante`_memoryCache.Add(key, data, DateTimeOffset.Now.AddMinutes(10))`. Esta llamada especifica que la entrada de la caché debe caducar después de 10 minutos, momento en el que los datos se eliminan automáticamente de la memoria caché.
5. El `ClearCache` método toma una `string` clave como argumento y elimina los datos asociados a esa clave de la memoria caché mediante`_memoryCache.Remove(key)`.

```
using System;
using System.Runtime.Caching;

public class InMemoryCache
{
    private static MemoryCache _memoryCache = new MemoryCache("dataCache");

    public static T GetData<T>(string key, Func<T> getData)
    {
        if (_memoryCache.Contains(key))
        {
            return (T)_memoryCache.Get(key);
        }

        T data = getData();
        _memoryCache.Add(key, data, DateTimeOffset.Now.AddMinutes(10));

        return data;
    }

    public static void ClearCache(string key)
    {
        _memoryCache.Remove(key);
    }
}
```

Puede usar el siguiente código:

```
public class Program
{
    public static void Main()
    {
        string cacheKey = "sample_data";
```

```
        Func<string> getSampleData = () =>
        {
            // Replace this with your data retrieval logic
            return "Sample data";
        };

        string data = InMemoryCache.GetData(cacheKey, getSampleData);
        Console.WriteLine("Data: " + data);
    }
}
```

El siguiente ejemplo muestra cómo usar LiteDB para almacenar datos en caché en un almacenamiento local. Puede usar LiteDB como alternativa o complemento al almacenamiento en caché en memoria. El siguiente código muestra cómo utilizar la biblioteca LiteDB para almacenar datos en caché en el almacenamiento local. La `LocalStorageCache` clase contiene las funciones principales para administrar la memoria caché.

```
using System;
using LiteDB;

public class LocalStorageCache
{
    private static string _liteDbPath = @"Filename=LocalCache.db";

    public static T GetData<T>(string key, Func<T> getData)
    {
        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection<T>("cache");
            var item = collection.FindOne(Query.EQ("_id", key));

            if (item != null)
            {
                return item;
            }
        }

        T data = getData();

        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection<T>("cache");
```

```
            collection.Upsert(new BsonValue(key), data);
        }

        return data;
    }

    public static void ClearCache(string key)
    {
        using (var db = new LiteDatabase(_liteDbPath))
        {
            var collection = db.GetCollection("cache");
            collection.Delete(key);
        }
    }
}

public class Program
{
    public static void Main()
    {
        string cacheKey = "sample_data";

        Func<string> getSampleData = () =>
        {
            // Replace this with your data retrieval logic
            return "Sample data";
        };

        string data = LocalStorageCache.GetData(cacheKey, getSampleData);
        Console.WriteLine("Data: " + data);
    }
}
```

Si tiene una caché estática o archivos estáticos que no cambian con frecuencia, también puede almacenar estos archivos en el almacenamiento de objetos de Amazon Simple Storage Service (Amazon S3). La aplicación puede recuperar el archivo de caché estática al iniciarse para usarlo localmente. Para obtener más información sobre cómo recuperar archivos de Amazon S3 mediante .NET, consulte Descarga de objetos en la documentación de Amazon S3.

Almacenamiento en caché con DAX

Puede utilizar una capa de almacenamiento en caché que se pueda compartir entre todas las instancias de la aplicación. DynamoDB Accelerator (DAX) es una memoria caché en memoria

totalmente gestionada y de alta disponibilidad para DynamoDB que puede multiplicar por diez el rendimiento. Puede usar DAX para reducir los costos al reducir la necesidad de aprovisionar en exceso las unidades de capacidad de lectura en las tablas de DynamoDB. Esto resulta especialmente útil para cargas de trabajo que requieren un uso intensivo de lecturas y que requieren lecturas repetidas de claves individuales.

DynamoDB tiene un precio bajo demanda o con capacidad aprovisionada, por lo que el número de lecturas y escrituras al mes contribuye al coste. Si tiene cargas de trabajo de lectura pesadas, los clústeres de DAX pueden ayudarle a reducir los costes al reducir el número de lecturas en las tablas de DynamoDB. Para obtener instrucciones sobre cómo configurar DAX, consulte Aceleración en memoria con DynamoDB Accelerator (DAX) en la documentación de DynamoDB. Para obtener información sobre la integración de aplicaciones de.NET, consulte Integrating Amazon DynamoDB DAX en su aplicación ASP.NET en. YouTube

Recursos adicionales

- Aceleración en memoria con DynamoDB Accelerator (DAX) - Amazon DynamoDB (documentación de DynamoDB)
- Integración de Amazon DynamoDB DAX en su aplicación ASP.NET () YouTube
- Descarga de objetos (documentación de Amazon S3)

Considere la posibilidad de utilizar .NET sin servidor

Descripción general

La informática sin servidor se ha convertido en un enfoque popular para crear e implementar aplicaciones. Esto se debe principalmente a la escalabilidad y la agilidad que ofrece el enfoque sin servidor a la hora de crear una arquitectura moderna. Sin embargo, es importante tener en cuenta el impacto en los costes de la informática sin servidor en algunos escenarios.

Lambda es una plataforma informática sin servidor que permite a los desarrolladores ejecutar código sin necesidad de servidores dedicados. Lambda es una opción particularmente atractiva para los desarrolladores de.NET que buscan reducir los costos de infraestructura. Con Lambda, los desarrolladores de.NET pueden desarrollar e implementar aplicaciones que son altamente escalables y potencialmente rentables. Al utilizar un enfoque sin servidores, los desarrolladores ya no aprovisionan servidores para gestionar las solicitudes de aplicaciones. En su lugar, los desarrolladores pueden crear funciones que se ejecuten bajo demanda. Esto hace que un enfoque

sin servidor sea más escalable, administrable y potencialmente más rentable que ejecutar, administrar y escalar máquinas virtuales. Como resultado, solo paga por los recursos utilizados por la aplicación, sin tener que preocuparse por los recursos infrautilizados o los costos de mantenimiento del servidor.

Los desarrolladores pueden usar versiones modernas de .NET multiplataforma para crear aplicaciones sin servidor que sean rápidas, eficientes y rentables. .NET Core y las versiones más recientes son un marco gratuito y de código abierto que se adapta mejor a la ejecución en plataformas sin servidor que las versiones anteriores de .NET Framework. Esto permite a los desarrolladores reducir el tiempo de desarrollo y aumentar el rendimiento de las aplicaciones. El dominio.NET moderno también es compatible con una variedad de lenguajes de programación, incluidos C# y F#. Por este motivo, es una opción atractiva para los desarrolladores que desean crear arquitecturas modernas en la nube.

En esta sección se explica cómo puede ahorrar costes mediante el uso de Lambda como opción sin servidor. Puede optimizar aún más los costos ajustando los perfiles de ejecución de sus funciones Lambda, ajustando el tamaño correcto de la asignación de memoria de sus funciones Lambda, utilizando el AOT nativo y pasando a funciones basadas en Graviton.

Impacto del costo

La medida en que pueda reducir los costes depende de varios factores, como el número de ejecuciones que ejecutarán sus funciones sin servidor, además de la cantidad de memoria asignada y la duración de cada función. AWS Lambda ofrece una capa gratuita, que incluye un millón de solicitudes gratuitas al mes y 400 000 GB por segundo de tiempo de procesamiento al mes. Puede reducir considerablemente los costes mensuales de las cargas de trabajo que se encuentren dentro o alrededor de estos límites de la capa gratuita.

También puede haber costes adicionales al utilizar un balanceador de carga con funciones Lambda como objetivo. Esto se calcula como la cantidad de datos procesados por el balanceador de carga para los objetivos Lambda.

Recomendaciones de optimización de costos

Ajuste el tamaño correcto de sus funciones Lambda

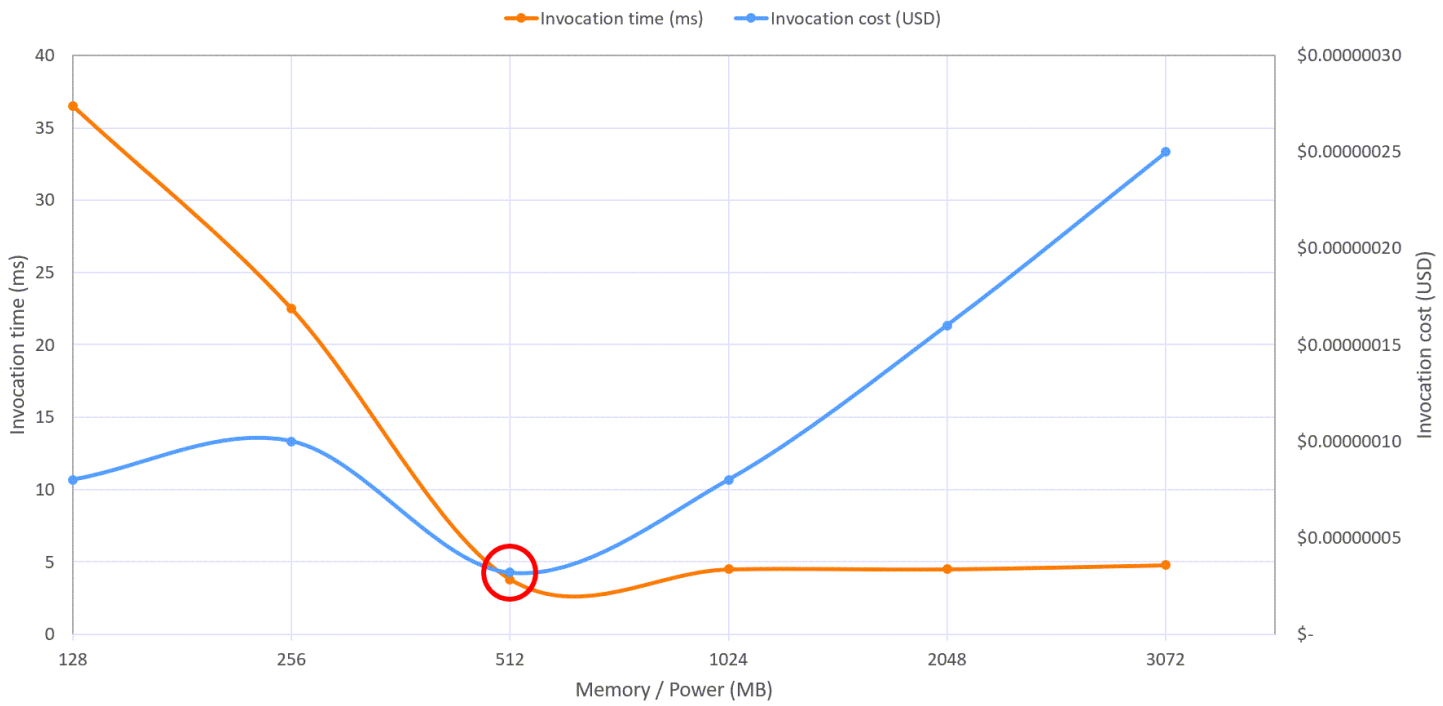
El dimensionamiento correcto es una práctica esencial para la optimización de costos en las funciones Lambda basadas en .NET. Este proceso implica identificar la configuración de memoria óptima que equilibre el rendimiento con la rentabilidad, sin necesidad de cambiar el código.

Al configurar la memoria para una función Lambda, que va desde 128 MB hasta 10 240 MB, también se ajusta la cantidad de vCPU disponible durante la invocación. Esto permite que las aplicaciones vinculadas a la memoria o a la CPU accedan a recursos adicionales durante la ejecución, lo que se traduce en una posible reducción de la duración de la invocación y del coste total.

Sin embargo, identificar la configuración óptima para las funciones de Lambda basadas en .NET puede ser un proceso manual y laborioso, especialmente si los cambios son frecuentes. La herramienta AWS Lambda de ajuste de potencia puede ayudarle a identificar la configuración adecuada mediante el análisis de un conjunto de configuraciones de memoria comparándolas con un ejemplo de carga útil.

Por ejemplo, aumentar la memoria de una función Lambda basada en .NET puede mejorar el tiempo total de invocación y reducir los costes sin afectar al rendimiento. La configuración de memoria óptima para una función puede variar. La herramienta AWS Lambda de ajuste de potencia puede ayudar a identificar la configuración más rentable para cada función.

En el siguiente gráfico de ejemplo, el tiempo total de invocación mejora a medida que aumenta la memoria de esta función Lambda. Esto se traduce en una reducción del coste total de la ejecución sin que ello afecte al rendimiento original de la función. Para esta función, la configuración de memoria óptima para la función es de 512 MB, ya que aquí es donde la utilización de los recursos es más eficiente en relación con el coste total de cada invocación. Esto varía según la función y, al utilizar la herramienta en las funciones de Lambda, se puede identificar si se benefician del tamaño correcto.



Le recomendamos que complete este ejercicio con regularidad, como parte de cualquier prueba de integración cuando se publiquen nuevas actualizaciones. Si se actualiza con poca frecuencia, realice este ejercicio periódicamente para asegurarse de que las funciones estén ajustadas y tengan el tamaño adecuado. Una vez que haya identificado la configuración de memoria adecuada para las funciones de Lambda, puede añadir el tamaño correcto a sus procesos. La herramienta AWS Lambda Power Tuning genera resultados programáticos que los flujos de trabajo de CI/CD pueden utilizar durante la publicación del código nuevo. Esto le permite automatizar la configuración de la memoria.

Puede descargar la herramienta AWS Lambda Power Tuning de forma gratuita. Para obtener instrucciones sobre cómo utilizar la herramienta, consulte Cómo ejecutar la máquina de estados en GitHub.

Lambda también admite el AOT nativo, lo que permite precompilar las aplicaciones.NET. Esto puede ayudar a reducir los costos al reducir los tiempos de ejecución de las funciones.NET. Para obtener más información sobre la creación de funciones de AOT nativas, consulte Funciones de.NET con compilación de AOT nativa en la documentación de Lambda.

Evite el tiempo de espera por inactividad

La duración de la función Lambda es una dimensión que se utiliza para calcular la facturación. Cuando el código de función realiza una llamada de bloqueo, se le facturará el tiempo que espere

para recibir una respuesta. Este tiempo de espera puede aumentar cuando las funciones de Lambda están encadenadas o una función actúa como orquestadora de otras funciones. Si tiene flujos de trabajo como operaciones por lotes o sistemas de entrega de pedidos, esto añade una sobrecarga de administración. Además, es posible que no sea posible completar toda la lógica del flujo de trabajo y la gestión de errores dentro del tiempo de espera máximo de Lambda de 15 minutos.

En lugar de utilizar esta lógica en el código de la función, le recomendamos que rediseñe la solución para utilizarla AWS Step Functionscomo orquestador del flujo de trabajo. Cuando utilice un flujo de trabajo estándar, se le facturará por cada transición de estado del flujo de trabajo y no por la duración total del flujo de trabajo. Además, puede cambiar la compatibilidad con los reintentos, las condiciones de espera, los flujos de trabajo con errores y las devoluciones de llamada a la condición de estado para permitir que las funciones de Lambda se centren en la lógica empresarial. Para obtener más información, consulte Optimización de AWS Lambda los costes, segunda parte, en el AWS blog de informática.

Pase a las funciones basadas en la gravedad

Las funciones Lambda impulsadas por los procesadores Graviton2 de próxima generación ya están disponibles de forma general. Las funciones Graviton2, que utilizan una arquitectura de procesador basada en ARM, están diseñadas para ofrecer hasta un 19 por ciento más de rendimiento a un costo un 20 por ciento menor para una variedad de cargas de trabajo sin servidor. Con una latencia más baja y un mejor rendimiento, las funciones impulsadas por los procesadores Graviton2 son ideales para impulsar aplicaciones esenciales sin servidor.

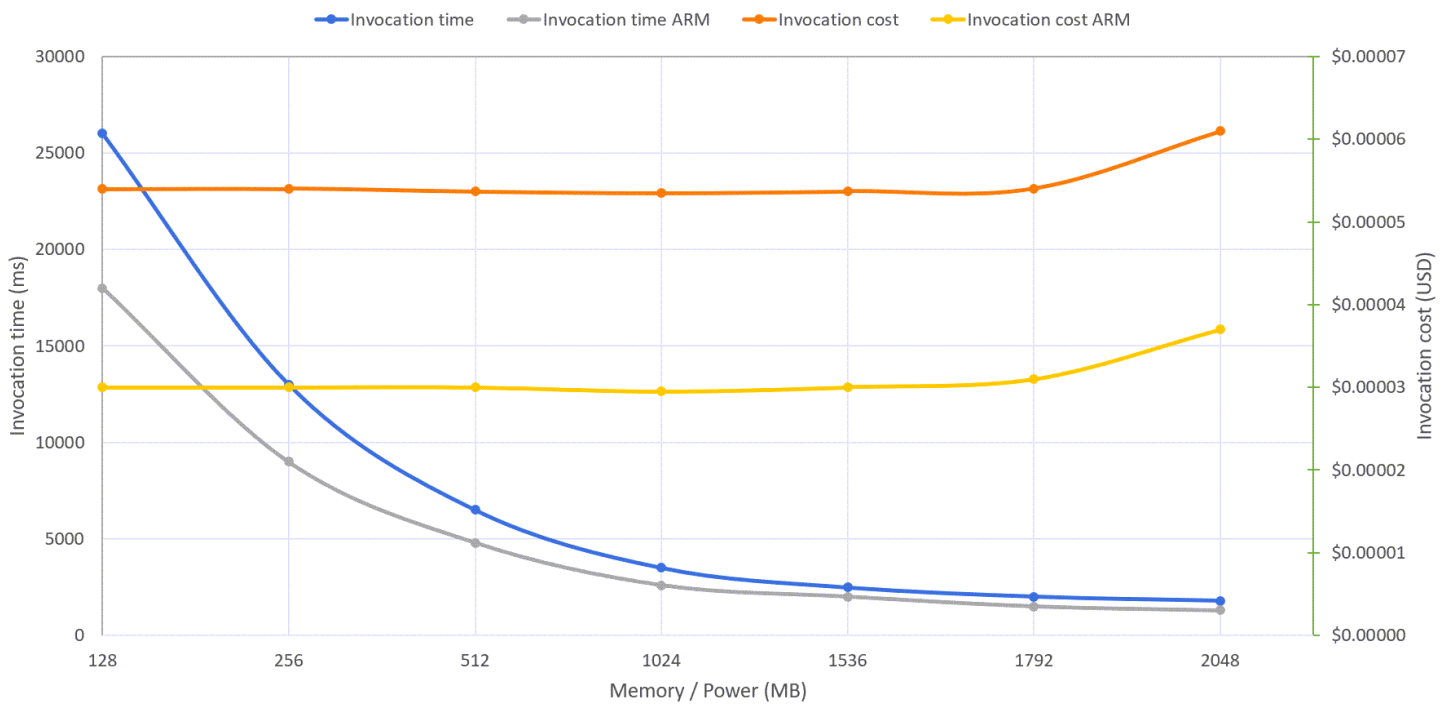
La migración a funciones de Lambda basadas en Graviton puede ser una opción rentable para los desarrolladores de.NET que buscan optimizar sus costos de Lambda. Las funciones basadas en Graviton utilizan procesadores basados en ARM en lugar de los procesadores x86 tradicionales. Esto puede suponer un importante ahorro de costes sin sacrificar el rendimiento.

Si bien pasar a funciones basadas en Gravitón tiene varias ventajas, también hay varios desafíos y consideraciones que le recomendamos que tenga en cuenta. Por ejemplo, las funciones basadas en Graviton requieren el uso de Amazon Linux 2, que puede no ser compatible con todas las aplicaciones.NET. Además, es posible que haya problemas de compatibilidad con bibliotecas o dependencias de terceros que no sean compatibles con los procesadores basados en ARM.

Si ejecuta aplicaciones.NET Framework y desea aprovechar las ventajas de la tecnología sin servidor con Lambda, puede considerar la posibilidad de migrar las aplicaciones a la versión moderna de.NET mediante el Asistente de portabilidad para.NET. Esto puede ayudarle a acelerar

la migración de aplicaciones de.NET antiguas a la versión moderna de.NET, lo que permitirá que la aplicación se ejecute en Linux.

En la siguiente tabla se comparan los resultados de las arquitecturas x86 y ARM/Graviton2 de una función que calcula números primos.



La función utiliza un único hilo. La duración más baja para ambas arquitecturas se indica cuando la memoria está configurada con 1,8 GB. Por encima de eso, las funciones de Lambda tienen acceso a más de 1 vCPU, pero en este caso, la función no puede utilizar la potencia adicional. Por la misma razón, los costes se mantienen estables con una memoria de hasta 1,8 GB. Con más memoria, los costes aumentan porque esta carga de trabajo no supone ningún beneficio de rendimiento adicional. Es evidente que el procesador Graviton2 proporciona un mejor rendimiento y unos costes más bajos para esta función que requiere un uso intensivo de recursos informáticos.

Para configurar la función para utilizar un procesador basado en ARM con Graviton, haga lo siguiente:

1. Inicie sesión en la AWS Management Console consola Lambda y ábrala.
2. Seleccione Crear función.
3. En Function name (Nombre de función), escriba un nombre.
4. Para Runtime, elija .NET 6 (C#/PowerShell).

5. En Arquitectura, seleccione arm64.
6. Realice las configuraciones adicionales que necesite y, a continuación, elija Crear función.

Recursos adicionales

- Funciones Lambda como objetivos (documentación)AWS
- Optimización de los AWS Lambda costes y el rendimiento mediante AWS Compute Optimizer (AWS Compute Blog)
- Optimización de AWS Lambda los costes: primera parte (blog AWS sobre informática)
- Optimización de AWS Lambda los costes: segunda parte (blog AWS sobre informática)
- Creación de aplicaciones.NET sin servidor AWS Lambda mediante .NET 7 (AWS Compute Blog)

Considere la posibilidad de crear bases de datos especialmente diseñadas

Descripción general

Uno de los aspectos más costosos de la ejecución de cargas de trabajo basadas en Microsoft proviene de la concesión de licencias de bases de datos comerciales, como SQL Server. Las empresas suelen estandarizar SQL Server como la plataforma de base de datos preferida y esto se arraiga en la cultura de desarrollo de la organización. Los desarrolladores suelen elegir un modelo relacional basado en SQL Server, independientemente del caso de uso. Los motivos pueden ser:

- La empresa ya tiene instancias o licencias de SQL Server disponibles.
- Los equipos se han acostumbrado al modelo de programación de SQL mediante el uso de bibliotecas compartidas y la lógica empresarial. ORMs
- La gerencia no conoce las alternativas.
- Los desarrolladores no conocen las alternativas.

Las bases de datos diseñadas específicamente pueden adaptarse a los patrones de acceso a los datos de su caso de uso. Las empresas adoptan cada vez más estas bases de datos a medida que adoptan arquitecturas más modernas (como los microservicios) y a medida que se reduce el alcance de las aplicaciones individuales.

Una base de datos creada específicamente no excluye un modelo relacional ni requiere un modelo NoSQL (no relacional). De hecho, una base de datos relacional se considera diseñada específicamente cuando se selecciona en respuesta a las necesidades específicas de una carga de trabajo. El uso de bases de datos diseñadas específicamente puede ayudar a los equipos a reducir los costos de bases de datos asociados a sus aplicaciones.NET y, al mismo tiempo, obtener las ventajas estándar de la nube, como la escalabilidad, la resiliencia y la reducción del trabajo pesado indiferenciado.

En la siguiente tabla se muestran las bases de datos especialmente diseñadas que ofrece. AWS

Base de datos	Tipo	Características
Amazon Aurora PostgreSQL o Amazon Aurora MySQL	Relacional	Casos de uso en los que los datos tienen una estructura fija Las bases de datos relaciona les mantienen naturalmente la coherencia de los datos a través de las transacciones ACID
Amazon DynamoDB	Par clave-valor	Base de datos NoSQL que almacena datos mediante una estructura de datos de tabla hash Almacenamiento y recuperac ión de datos no estructurados de alto rendimiento Los casos de uso incluyen los perfiles de usuario, el estado de la sesión y los datos del carrito de compras
Amazon ElastiCache	En memoria	Base de datos NoSQL de alto rendimiento que almacena datos no estructurados en la memoria con un tiempo de

Base de datos	Tipo	Características
		acceso inferior a un milisegundo Se utiliza para datos efímeros a los que se accede con frecuencia, como sesiones de usuario, y como capa de almacenamiento en caché delante de otros almacenes de datos más lentos Incluye soporte tanto para ElastiCache (Redis OSS) como para (Memcached) ElastiCache
Amazon MemoryDB	Duradero en memoria	Base de datos diseñada específicamente compatible con Redis con almacenamiento duradero
Amazon Timestream	Serie temporal	Base de datos diseñada para la ingesta de datos de alto rendimiento en orden temporal Los casos de uso incluyen aplicaciones de Internet de las cosas (IoT) y almacenamiento de métricas o datos de telemetría

Base de datos	Tipo	Características
Amazon DocumentDB	Documento	Base de datos NoSQL que almacena datos sin una estructura prescrita ni relaciones forzadas con otros datos Suele utilizarse para cargas de trabajo de lectura intensiva, como catálogos de productos
Amazon Neptune	Gráfico	Base de datos NoSQL que contiene datos y una representación de las conexiones entre los elementos de datos Los casos de uso incluyen la detección de fraudes, los motores de recomendación y las aplicaciones sociales
Amazon Keyspaces	Columna ancha	Base de datos distribuida de alto rendimiento basada en Apache Cassandra Los casos de uso incluyen aplicaciones de IoT, procesamiento de eventos y aplicaciones de juegos

Uno de los principales impulsores de la adopción de bases de datos diseñadas específicamente puede atribuirse a la eliminación de las licencias comerciales. Sin embargo, la capacidad de autoscalamiento de bases de datos como DynamoDB (incluido el modo bajo demanda), Aurora, Amazon Neptune y Amazon Keyspaces le permite aprovisionar capacidad para casos normales, en lugar de para picos de uso. Las bases de datos diseñadas específicamente, como Timestream,

no tienen servidores y se escalan automáticamente para satisfacer la demanda sin necesidad de aprovisionamiento previo.

AWS ofrece Babelfish para Aurora PostgreSQL si desea utilizar una base de datos relacional especialmente diseñada y compatible con código abierto, pero no puede o no quiere realizar cambios significativos en el código de su aplicación. En algunos casos, Babelfish le permite utilizar un código de acceso de SQL Server existente, prácticamente sin cambios.

Al elegir una base de datos relacional diseñada específicamente para aplicaciones, es importante conservar las mismas características (o funcionalmente equivalentes) que necesita para sus aplicaciones. Esta recomendación se refiere a las bases de datos diseñadas específicamente como almacén de datos principal para las aplicaciones. Otras recomendaciones abordan aplicaciones específicas (como el almacenamiento en caché).

Impacto del costo

Si bien es poco probable que la adopción de bases de datos diseñadas específicamente para las cargas de trabajo de.NET afecte directamente al consumo/costo de cómputo, puede influir directamente en el costo de los servicios de bases de datos que consumen las aplicaciones.NET. De hecho, el ahorro de costes puede ser un objetivo secundario si se compara con los beneficios adicionales de agilidad, escalabilidad, resiliencia y durabilidad de los datos.

Está fuera del ámbito de esta guía explicar todo el proceso de elegir una base de datos diseñada específicamente para las aplicaciones y rediseñar una estrategia de datos para utilizarlas de forma eficaz. Para obtener más información, consulte Bases de datos diseñadas específicamente en el directorio de tutoriales. AWS

En las tablas siguientes se muestran varios ejemplos de cómo la sustitución de SQL Server por una base de datos diseñada específicamente puede modificar los costes de las aplicaciones. Tenga en cuenta que se trata simplemente de estimaciones aproximadas. Se requieren puntos de referencia y la optimización de las cargas de trabajo reales para calcular el costo de producción exacto.

Estas son algunas estimaciones de bases de datos especialmente diseñadas y utilizadas con frecuencia, que incluyen procesamiento bajo demanda y SSD de 100 GB, incluidas bases de datos de instancia única. `us-east-1` Los costos de licencia incluyen la licencia de SQL Server más la garantía del software.

La siguiente tabla muestra los costos estimados de los ejemplos de bases de datos comerciales.

Motor de base de datos	Modelo de licencia	Tipo y especificaciones de la instancia	AWS cómputo más costo de almacenamiento	Coste de la licencia	Coste mensual total
Edición estándar de SQL Server en Amazon EC2	Licencia incluida	r6i.2xlarge (8 CPU/64 GB RAM)	1.345,36 DÓLARES	0,00$	1.345,36 DÓLARES
Edición SQL Server Enterprise en Amazon EC2	Licencia incluida	r6i.2xlarge (8 CPU/64 GB RAM)	2.834,56 DÓLARES	0,00$	2.834,56 DÓLARES
Edición estándar de SQL Server en Amazon EC2	BYOL	r6i.2xlarge (8 CPU/64 GB RAM)	644,56 DÓLARES	456,00 DÓLARES	1.100,56 DÓLARES
Edición SQL Server Enterprise en Amazon EC2	BYOL	r6i.2xlarge (8 CPU/64 GB RAM)	644,56 DÓLARES	1.750,00 DÓLARES	2.394,56 DÓLARES
Edición estándar de SQL Server en Amazon RDS		db.r6i.2xlarge (8 CPU/64 GB RAM)	2.318,30 DÓLARES	0,00$	2.318,30 DÓLARES
Edición SQL Server		db.r6i.2xlarge (8 CPU/64 GB RAM)	3.750,56 DÓLARES	0,00$	3.750,56 DÓLARES

Motor de base de datos	Modelo de licencia	Tipo y especific aciones de la instancia	AWS cómputo más costo de almacenam iento	Coste de la licencia	Coste mensual total
Enterprise en Amazon RDS					

La siguiente tabla muestra los costos estimados de los ejemplos diseñados específicamente.

Motor de base de datos	Tipo y especific aciones de la instancia	AWS cómputo más costo de almacenamiento	Coste de la licencia	Coste mensual total
PostgreSQL de Amazon Aurora	r6g.2xlarge (8 CPU/64 GB RAM)	855,87 DÓLARES	0,00$	855,87 DÓLARES
DynamoDB	Base aprovisio nada: 100 WCU/400 RCU	72,00 DÓLARES		72,00 DÓLARES
Amazon DocumentDB	db.r6i.2xlarge (8 CPU/64 GB RAM)	778,60 DÓLARES		778,60 DÓLARES

> **Important**
>
> La tabla se basa en los costos estimados de licencia de SQL Server with Software Assurance durante los tres primeros años de la compra. Para la edición estándar de SQL Server: 4.100 dólares, paquete de 2 núcleos, 3 años. Para la edición SQL Server Enterprise: 15.700 dólares, paquete de 2 núcleos, 3 años.

Le recomendamos que considere las implicaciones de costo antes de adoptar bases de datos diseñadas específicamente. Por ejemplo, el costo de actualizar las aplicaciones para usar una base de datos diseñada específicamente está relacionado con la complejidad de la aplicación y de la base de datos de origen. Asegúrese de tener en cuenta el coste total de propiedad al planificar este cambio de arquitectura. Esto incluye refactorizar las aplicaciones, capacitar al personal en nuevas tecnologías y planificar cuidadosamente el rendimiento y el consumo previstos para cada carga de trabajo. A partir de ahí, puede determinar si la inversión compensa el ahorro de costes. En la mayoría de los casos, el mantenimiento de un end-of-support producto supone un riesgo para la seguridad y la conformidad, y el coste de remediarlo vale la pena tanto el esfuerzo como la inversión inicial.

Recomendaciones de optimización de costos

Para las aplicaciones.NET que acceden a SQL Server, existen bibliotecas que sustituyen a las bases de datos relacionales diseñadas específicamente. Puede implementar estas bibliotecas en su aplicación para reemplazar una funcionalidad similar de una aplicación de SQL Server.

En la siguiente tabla, se muestran algunas bibliotecas que se pueden usar en muchos escenarios comunes.

Library	Base de datos	Sustitución para	Compatibilidad con marcos
Proveedor principal de Npgsql Entity Framework	PostgreSQL de Amazon Aurora	Proveedor principal de servidores SQL de Entity Framework	.NET moderno
Proveedor de Npgsql Entity Framework 6	PostgreSQL de Amazon Aurora	Proveedor de servidores SQL de Entity Framework 6.0	.NET Framework
Npgsql (biblioteca PostgreSQL compatible con ADO.NET)	PostgreSQL de Amazon Aurora	ADO.NET	.NET Framework /.NET moderno
Proveedor principal de MySQL Entity Framework	Amazon Aurora MySQL	Proveedor principal de servidores SQL de Entity Framework	.NET moderno

Library	Base de datos	Sustitución para	Compatibilidad con marcos
Pomelo. EntityFrameworkCore. MySql	Amazon Aurora MySQL	Proveedor principal de SQL Server de Entity Framework	.NET moderno

La conexión a Amazon Aurora PostgreSQL mediante Babelfish no requiere ningún código especial para conectarse. Sin embargo, todo el código debe probarse minuciosamente antes de su uso.

Otras bases de datos diseñadas específicamente tienen bibliotecas para acceder a bibliotecas compatibles con .NET que le permiten acceder a bases de datos diseñadas específicamente. Entre los ejemplos se incluyen:

- Uso de bases de datos NoSQL de Amazon DynamoDB (documentación)AWS SDK para .NET
- Controlador MongoDB C# (documentación de MongoDB)
- .NET (documentación de Timestream)
- Uso de un controlador de cliente Cassandra .NET Core para acceder a Amazon Keyspaces mediante programación (documentación de Amazon Keyspaces)
- Utilización de.NET para conectarse a una instancia de base de datos de Neptune (documentación de Neptune)

Si migra a bases de datos diseñadas específicamente, puede utilizar estas herramientas AWS para facilitar el proceso de migración:

- AWS Schema Conversion Tool (AWS SCT) puede ayudarle a transformar los esquemas de SQL Server en Amazon Aurora y Amazon DynamoDB.
- AWS Database Migration Service (AWS DMS) puede ayudarlo a migrar datos, una vez o de forma continua, de SQL Server a Aurora o DynamoDB.
- Babelfish Compass puede ayudarle a comprobar la compatibilidad de su base de datos de SQL Server para utilizarla con Babelfish para Aurora PostgreSQL.

Recursos adicionales

- Guía para migrar SQL Server a Amazon Aurora AWS PostgreSQL (blog sobre bases de datos)

- Taller de modernización de .NET (Workshop Studio)AWS
- Jornada de inmersión en la modernización de la aplicación Babelfish (AWS Workshop Studio)
- Día de inmersión en .NET (AWS taller de estudio)
- Cómo empezar a usar Amazon Timestream con .NET () GitHub
- Bases de datos especialmente diseñadas para aplicaciones de.NET modernas en (presentación) AWSAWS

Pasos a seguir a continuación

Cuando termine de revisar esta guía, le recomendamos que siga los siguientes pasos para implementar el MACO:

1. Póngase en contacto con un experto en MACO. Un experto de MACO puede ayudarlo a responder sus preguntas y abordar sus inquietudes. Si ya trabaja con un equipo de AWS cuentas, póngase en contacto con el equipo y solicite la ayuda de un experto de MACO. Si no tienes un equipo de cuentas, ponte en contacto con optimize-microsoft@amazon.com.
2. Aplica las recomendaciones. Aplique las recomendaciones, las mejores prácticas y las estrategias que aprendió en esta guía y al hablar con un experto de MACO.
3. Realice un seguimiento de los cambios de costos.Etiquete sus cargas de trabajo y utilice servicios similares AWS Cost Explorer y AWS Budgets para realizar un seguimiento, una supervisión y un control detallados de los costes.

Historial de documentos

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

Cambio	Descripción	Fecha
Actualizaciones de SQL Server	Hemos actualizado la sección Optimizar la CPU para las cargas de trabajo de SQL Server.	25 de octubre de 2024
Actualizaciones de SQL Server y Container	Añadimos las secciones Optimizar el tamaño de SQL Server mediante Compute Optimizer, Trusted Advisor Revisar las recomendaciones para las cargas de trabajo de SQL Server y Cambiar la plataforma de las aplicaciones de Windows con App2Container.	29 de junio de 2024
Optimización de licencias de SQL Server	Añadimos la sección Optimizar las licencias de SQL Server mediante Compute Optimizer.	22 de mayo de 2024
Publicación inicial	—	21 de diciembre de 2023

AWS Glosario de orientación prescriptiva

Los siguientes son términos de uso común en las estrategias, guías y patrones proporcionados por la Guía AWS prescriptiva. Para sugerir entradas, utilice el enlace Enviar comentarios al final del glosario.

Números

Las 7 R

Siete estrategias de migración comunes para trasladar aplicaciones a la nube. Estas estrategias se basan en las 5 R que Gartner identificó en 2011 y consisten en lo siguiente:

- Refactorizar/rediseñar: traslade una aplicación y modifique su arquitectura mediante el máximo aprovechamiento de las características nativas en la nube para mejorar la agilidad, el rendimiento y la escalabilidad. Por lo general, esto implica trasladar el sistema operativo y la base de datos. Ejemplo: migre su base de datos Oracle local a la edición compatible con PostgreSQL de Amazon Aurora.
- Redefinir la plataforma (transportar y redefinir): traslade una aplicación a la nube e introduzca algún nivel de optimización para aprovechar las capacidades de la nube. Ejemplo: migre su base de datos Oracle local a Amazon Relational Database Service (Amazon RDS) para Oracle en el. Nube de AWS
- Recomprar (readquirir): cambie a un producto diferente, lo cual se suele llevar a cabo al pasar de una licencia tradicional a un modelo SaaS. Ejemplo: migre su sistema de gestión de relaciones con los clientes (CRM) a Salesforce.com.
- Volver a alojar (migrar mediante lift-and-shift): traslade una aplicación a la nube sin realizar cambios para aprovechar las capacidades de la nube. Ejemplo: migre su base de datos Oracle local a Oracle en una EC2 instancia del. Nube de AWS
- Reubicar: (migrar el hipervisor mediante lift and shift): traslade la infraestructura a la nube sin comprar equipo nuevo, reescribir aplicaciones o modificar las operaciones actuales. Los servidores se migran de una plataforma local a un servicio en la nube para la misma plataforma. Ejemplo: migrar una Microsoft Hyper-V aplicación a AWS.
- Retener (revisitar): conserve las aplicaciones en el entorno de origen. Estas pueden incluir las aplicaciones que requieren una refactorización importante, que desee posponer para más adelante, y las aplicaciones heredadas que desee retener, ya que no hay ninguna justificación empresarial para migrarlas.

- Retirar: retire o elimine las aplicaciones que ya no sean necesarias en un entorno de origen.

A

ABAC

Consulte control de acceso basado en atributos.

servicios abstractos

Consulte servicios gestionados.

ACID

Consulte atomicidad, consistencia, aislamiento y durabilidad.

migración activa-activa

Método de migración de bases de datos en el que las bases de datos de origen y destino se mantienen sincronizadas (mediante una herramienta de replicación bidireccional o mediante operaciones de escritura doble) y ambas bases de datos gestionan las transacciones de las aplicaciones conectadas durante la migración. Este método permite la migración en lotes pequeños y controlados, en lugar de requerir una transición única. Es más flexible, pero requiere más trabajo que la migración activa-pasiva.

migración activa-pasiva

Método de migración de bases de datos en el que las bases de datos de origen y destino se mantienen sincronizadas, pero solo la base de datos de origen gestiona las transacciones de las aplicaciones conectadas, mientras los datos se replican en la base de datos de destino. La base de datos de destino no acepta ninguna transacción durante la migración.

función agregada

Función SQL que opera en un grupo de filas y calcula un único valor de retorno para el grupo. Algunos ejemplos de funciones agregadas incluyen `SUM` y`MAX`.

IA

Véase inteligencia artificial.

AIOps

Consulte las operaciones de inteligencia artificial.

anonimización

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

antipatrones

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

control de aplicaciones

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

cartera de aplicaciones

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

inteligencia artificial (IA)

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

operaciones de inteligencia artificial (AIOps)

El proceso de utilizar técnicas de machine learning para resolver problemas operativos, reducir los incidentes operativos y la intervención humana, y mejorar la calidad del servicio. Para obtener más información sobre cómo AIOps se utiliza en la estrategia de AWS migración, consulte la guía de integración de operaciones.

cifrado asimétrico

Algoritmo de cifrado que utiliza un par de claves, una clave pública para el cifrado y una clave privada para el descifrado. Puede compartir la clave pública porque no se utiliza para el descifrado, pero el acceso a la clave privada debe estar sumamente restringido.

atomicidad, consistencia, aislamiento, durabilidad (ACID)

Conjunto de propiedades de software que garantizan la validez de los datos y la fiabilidad operativa de una base de datos, incluso en caso de errores, cortes de energía u otros problemas.

control de acceso basado en atributos (ABAC)

La práctica de crear permisos detallados basados en los atributos del usuario, como el departamento, el puesto de trabajo y el nombre del equipo. Para obtener más información, consulte ABAC AWS en la documentación AWS Identity and Access Management (IAM).

origen de datos fidedigno

Ubicación en la que se almacena la versión principal de los datos, que se considera la fuente de información más fiable. Puede copiar los datos del origen de datos autorizado a otras ubicaciones con el fin de procesarlos o modificarlos, por ejemplo, anonimizarlos, redactarlos o seudonimizarlos.

Zona de disponibilidad

Una ubicación distinta dentro de una Región de AWS que está aislada de los fallos en otras zonas de disponibilidad y que proporciona una conectividad de red económica y de baja latencia a otras zonas de disponibilidad de la misma región.

AWS Marco de adopción de la nube (AWS CAF)

Un marco de directrices y mejores prácticas AWS para ayudar a las organizaciones a desarrollar un plan eficiente y eficaz para migrar con éxito a la nube. AWS CAF organiza la orientación en seis áreas de enfoque denominadas perspectivas: negocios, personas, gobierno, plataforma, seguridad y operaciones. Las perspectivas empresariales, humanas y de gobernanza se centran en las habilidades y los procesos empresariales; las perspectivas de plataforma, seguridad y operaciones se centran en las habilidades y los procesos técnicos. Por ejemplo, la perspectiva humana se dirige a las partes interesadas que se ocupan de los Recursos Humanos (RR. HH.), las funciones del personal y la administración de las personas. Desde esta perspectiva, AWS CAF proporciona orientación para el desarrollo, la formación y la comunicación de las personas a fin de preparar a la organización para una adopción exitosa de la nube. Para obtener más información, consulte la Página web de AWS CAF y el Documento técnico de AWS CAF.

AWS Marco de calificación de la carga de trabajo (AWS WQF)

Herramienta que evalúa las cargas de trabajo de migración de bases de datos, recomienda estrategias de migración y proporciona estimaciones de trabajo. AWS WQF se incluye con AWS

Schema Conversion Tool ().AWS SCT Analiza los esquemas de bases de datos y los objetos de código, el código de las aplicaciones, las dependencias y las características de rendimiento y proporciona informes de evaluación.

B

Un bot malo

Un bot destinado a interrumpir o causar daño a personas u organizaciones.

BCP

Consulte la planificación de la continuidad del negocio.

gráfico de comportamiento

Una vista unificada e interactiva del comportamiento de los recursos y de las interacciones a lo largo del tiempo. Puede utilizar un gráfico de comportamiento con Amazon Detective para examinar los intentos de inicio de sesión fallidos, las llamadas sospechosas a la API y acciones similares. Para obtener más información, consulte Datos en un gráfico de comportamiento en la documentación de Detective.

sistema big-endian

Un sistema que almacena primero el byte más significativo. Véase también endianness.

clasificación binaria

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

filtro de floración

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

implementación azul/verde

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

bot

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

botnet

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

branch

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

acceso con cristales rotos

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

estrategia de implementación sobre infraestructura existente

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

caché de búfer

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

capacidad empresarial

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

planificación de la continuidad del negocio (BCP)

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

C

CAF

Consulte el marco AWS de adopción de la nube.

despliegue canario

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

CCoE

Consulte Cloud Center of Excellence.

CDC

Consulte la captura de datos de cambios.

captura de datos de cambio (CDC)

Proceso de seguimiento de los cambios en un origen de datos, como una tabla de base de datos, y registro de los metadatos relacionados con el cambio. Puede utilizar los CDC para diversos fines, como auditar o replicar los cambios en un sistema de destino para mantener la sincronización.

ingeniería del caos

Introducir intencionalmente fallos o eventos disruptivos para poner a prueba la resiliencia de un sistema. Puedes usar AWS Fault Injection Service (AWS FIS) para realizar experimentos que estresen tus AWS cargas de trabajo y evalúen su respuesta.

CI/CD

Consulte la integración continua y la entrega continua.

clasificación

Un proceso de categorización que permite generar predicciones. Los modelos de ML para problemas de clasificación predicen un valor discreto. Los valores discretos siempre son distintos entre sí. Por ejemplo, es posible que un modelo necesite evaluar si hay o no un automóvil en una imagen.

cifrado del cliente

Cifrado de datos localmente, antes de que el objetivo los Servicio de AWS reciba.

Centro de excelencia en la nube (CCoE)

Equipo multidisciplinario que impulsa los esfuerzos de adopción de la nube en toda la organización, incluido el desarrollo de las prácticas recomendadas en la nube, la movilización de recursos, el establecimiento de plazos de migración y la dirección de la organización durante las transformaciones a gran escala. Para obtener más información, consulte las publicaciones de CCo E en el blog de estrategia Nube de AWS empresarial.

computación en la nube

La tecnología en la nube que se utiliza normalmente para la administración de dispositivos de IoT y el almacenamiento de datos de forma remota. La computación en la nube suele estar conectada a la tecnología de computación perimetral.

modelo operativo en la nube

En una organización de TI, el modelo operativo que se utiliza para crear, madurar y optimizar uno o más entornos de nube. Para obtener más información, consulte Creación de su modelo operativo de nube.

etapas de adopción de la nube

Las cuatro fases por las que suelen pasar las organizaciones cuando migran a Nube de AWS:

- Proyecto: ejecución de algunos proyectos relacionados con la nube con fines de prueba de concepto y aprendizaje
- Fundamento: realizar inversiones fundamentales para escalar su adopción de la nube (p. ej., crear una landing zone, definir una CCo E, establecer un modelo de operaciones)
- Migración: migración de aplicaciones individuales
- Reinvención: optimización de productos y servicios e innovación en la nube

Stephen Orban definió estas etapas en la entrada del blog The Journey Toward Cloud-First & the Stages of Adoption en el blog Nube de AWS Enterprise Strategy. Para obtener información sobre su relación con la estrategia de AWS migración, consulte la guía de preparación para la migración.

CMDB

Consulte la base de datos de administración de la configuración.

repositorio de código

Una ubicación donde el código fuente y otros activos, como documentación, muestras y scripts, se almacenan y actualizan mediante procesos de control de versiones. Los repositorios en la nube más comunes incluyen GitHub oBitbucket Cloud. Cada versión del código se denomina rama. En una estructura de microservicios, cada repositorio se encuentra dedicado a una única funcionalidad. Una sola canalización de CI/CD puede utilizar varios repositorios.

caché en frío

Una caché de búfer que está vacía no está bien poblada o contiene datos obsoletos o irrelevantes. Esto afecta al rendimiento, ya que la instancia de la base de datos debe leer desde la memoria principal o el disco, lo que es más lento que leer desde la memoria caché del búfer.

datos fríos

Datos a los que se accede con poca frecuencia y que suelen ser históricos. Al consultar este tipo de datos, normalmente se aceptan consultas lentas. Trasladar estos datos a niveles o clases de almacenamiento de menor rendimiento y menos costosos puede reducir los costos.

visión artificial (CV)

Campo de la IA que utiliza el aprendizaje automático para analizar y extraer información de formatos visuales, como imágenes y vídeos digitales. Por ejemplo, Amazon SageMaker AI proporciona algoritmos de procesamiento de imágenes para CV.

desviación de configuración

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

base de datos de administración de configuración (CMDB)

Repositorio que almacena y administra información sobre una base de datos y su entorno de TI, incluidos los componentes de hardware y software y sus configuraciones. Por lo general, los

datos de una CMDB se utilizan en la etapa de detección y análisis de la cartera de productos durante la migración.

paquete de conformidad

Conjunto de AWS Config reglas y medidas correctivas que puede reunir para personalizar sus comprobaciones de conformidad y seguridad. Puede implementar un paquete de conformidad como una entidad única en una región Cuenta de AWS y, o en una organización, mediante una plantilla YAML. Para obtener más información, consulta los paquetes de conformidad en la documentación. AWS Config

integración y entrega continuas (CI/CD)

El proceso de automatización de las etapas de origen, compilación, prueba, puesta en escena y producción del proceso de publicación del software. CI/CD is commonly described as a pipeline. CI/CDpuede ayudarlo a automatizar los procesos, mejorar la productividad, mejorar la calidad del código y entregar con mayor rapidez. Para obtener más información, consulte Beneficios de la entrega continua. CD también puede significar implementación continua. Para obtener más información, consulte Entrega continua frente a implementación continua.

CV

Vea la visión artificial.

D

datos en reposo

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

clasificación de datos

Un proceso para identificar y clasificar los datos de su red en función de su importancia y sensibilidad. Es un componente fundamental de cualquier estrategia de administración de riesgos de ciberseguridad porque lo ayuda a determinar los controles de protección y retención adecuados para los datos. La clasificación de datos es un componente del pilar de seguridad del AWS Well-Architected Framework. Para obtener más información, consulte Clasificación de datos.

desviación de datos

Una variación significativa entre los datos de producción y los datos que se utilizaron para entrenar un modelo de machine learning, o un cambio significativo en los datos de entrada

a lo largo del tiempo. La desviación de los datos puede reducir la calidad, la precisión y la imparcialidad generales de las predicciones de los modelos de machine learning.

datos en tránsito

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

malla de datos

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

minimización de datos

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

perímetro de datos

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

preprocesamiento de datos

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

procedencia de los datos

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

titular de los datos

Persona cuyos datos se recopilan y procesan.

almacenamiento de datos

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

lenguaje de definición de datos (DDL)

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

lenguaje de manipulación de datos (DML)

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

DDL

Consulte el lenguaje de definición de bases de datos.

conjunto profundo

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

aprendizaje profundo

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

defense-in-depth

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

administrador delegado

En AWS Organizations, un servicio compatible puede registrar una cuenta de AWS miembro para administrar las cuentas de la organización y gestionar los permisos de ese servicio. Esta cuenta se denomina administrador delegado para ese servicio. Para obtener más información y una lista de servicios compatibles, consulte Servicios que funcionan con AWS Organizations en la documentación de AWS Organizations .

Implementación

El proceso de hacer que una aplicación, características nuevas o correcciones de código se encuentren disponibles en el entorno de destino. La implementación abarca implementar

cambios en una base de código y, a continuación, crear y ejecutar esa base en los entornos de la aplicación.

entorno de desarrollo

Consulte entorno.

control de detección

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

asignación de flujos de valor para el desarrollo (DVSM)

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

gemelo digital

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

tabla de dimensiones

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

desastre

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

recuperación de desastres (DR)

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

DML

Consulte el lenguaje de manipulación de bases de datos.

diseño basado en el dominio

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

DR

Consulte recuperación ante desastres.

detección de deriva

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

DVSM

Consulte el mapeo del flujo de valor del desarrollo.

E

EDA

Consulte el análisis exploratorio de datos.

EDI

Véase intercambio electrónico de datos.

computación en la periferia

La tecnología que aumenta la potencia de cálculo de los dispositivos inteligentes en la periferia de una red de IoT. En comparación con la computación en nube, la computación perimetral puede reducir la latencia de la comunicación y mejorar el tiempo de respuesta.

intercambio electrónico de datos (EDI)

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

cifrado

Proceso informático que transforma datos de texto plano, legibles por humanos, en texto cifrado.

clave de cifrado

Cadena criptográfica de bits aleatorios que se genera mediante un algoritmo de cifrado. Las claves pueden variar en longitud y cada una se ha diseñado para ser impredecible y única.

endianidad

El orden en el que se almacenan los bytes en la memoria del ordenador. Los sistemas big-endianos almacenan primero el byte más significativo. Los sistemas Little-Endian almacenan primero el byte menos significativo.

punto de conexión

Consulte el punto final del servicio.

servicio de punto de conexión

Servicio que puede alojar en una nube privada virtual (VPC) para compartir con otros usuarios. Puede crear un servicio de punto final AWS PrivateLink y conceder permisos a otros directores Cuentas de AWS o a AWS Identity and Access Management (IAM). Estas cuentas o entidades principales pueden conectarse a su servicio de punto de conexión de forma privada mediante la creación de puntos de conexión de VPC de interfaz. Para obtener más información, consulte Creación de un servicio de punto de conexión en la documentación de Amazon Virtual Private Cloud (Amazon VPC).

planificación de recursos empresariales (ERP)

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

cifrado de sobre

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

entorno

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

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

epopeya

En las metodologías ágiles, son categorías funcionales que ayudan a organizar y priorizar el trabajo. Las epopeyas brindan una descripción detallada de los requisitos y las tareas de implementación. Por ejemplo, las epopeyas AWS de seguridad de CAF incluyen la gestión de identidades y accesos, los controles de detección, la seguridad de la infraestructura, la protección de datos y la respuesta a incidentes. Para obtener más información sobre las epopeyas en la estrategia de migración de AWS , consulte la Guía de implementación del programa.

ERP

Consulte planificación de recursos empresariales.

análisis de datos de tipo exploratorio (EDA)

El proceso de analizar un conjunto de datos para comprender sus características principales. Se recopilan o agregan datos y, a continuación, se realizan las investigaciones iniciales para

encontrar patrones, detectar anomalías y comprobar las suposiciones. El EDA se realiza mediante el cálculo de estadísticas resumidas y la creación de visualizaciones de datos.

F

tabla de datos

La tabla central de un esquema en forma de estrella. Almacena datos cuantitativos sobre las operaciones comerciales. Normalmente, una tabla de hechos contiene dos tipos de columnas: las que contienen medidas y las que contienen una clave externa para una tabla de dimensiones.

fallan rápidamente

Una filosofía que utiliza pruebas frecuentes e incrementales para reducir el ciclo de vida del desarrollo. Es una parte fundamental de un enfoque ágil.

límite de aislamiento de fallas

En el Nube de AWS, un límite, como una zona de disponibilidad Región de AWS, un plano de control o un plano de datos, que limita el efecto de una falla y ayuda a mejorar la resiliencia de las cargas de trabajo. Para obtener más información, consulte Límites de AWS aislamiento de errores.

rama de característica

Consulte la sucursal.

características

Los datos de entrada que se utilizan para hacer una predicción. Por ejemplo, en un contexto de fabricación, las características pueden ser imágenes que se capturan periódicamente desde la línea de fabricación.

importancia de las características

La importancia que tiene una característica para las predicciones de un modelo. Por lo general, esto se expresa como una puntuación numérica que se puede calcular mediante diversas técnicas, como las explicaciones aditivas de Shapley (SHAP) y los gradientes integrados. Para obtener más información, consulte Interpretabilidad del modelo de aprendizaje automático con AWS.

transformación de funciones

Optimizar los datos para el proceso de ML, lo que incluye enriquecer los datos con fuentes adicionales, escalar los valores o extraer varios conjuntos de información de un solo campo de datos. Esto permite que el modelo de ML se beneficie de los datos. Por ejemplo, si divide la fecha del “27 de mayo de 2021 00:15:37” en “jueves”, “mayo”, “2021” y “15”, puede ayudar al algoritmo de aprendizaje a aprender patrones matizados asociados a los diferentes componentes de los datos.

indicaciones de unos pocos pasos

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

FGAC

Consulte el control de acceso detallado.

control de acceso preciso (FGAC)

El uso de varias condiciones que tienen por objetivo permitir o denegar una solicitud de acceso.

migración relámpago

Método de migración de bases de datos que utiliza la replicación continua de datos mediante la captura de datos modificados para migrar los datos en el menor tiempo posible, en lugar de utilizar un enfoque gradual. El objetivo es reducir al mínimo el tiempo de inactividad.

FM

Consulte el modelo básico.

modelo de base (FM)

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

G

IA generativa

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

bloqueo geográfico

Consulta las restricciones geográficas.

restricciones geográficas (bloqueo geográfico)

En Amazon CloudFront, una opción para impedir que los usuarios de países específicos accedan a las distribuciones de contenido. Puede utilizar una lista de permitidos o bloqueados para especificar los países aprobados y prohibidos. Para obtener más información, consulta Restringir la distribución geográfica del contenido en la CloudFront documentación.

Flujo de trabajo de Gitflow

Un enfoque en el que los entornos inferiores y superiores utilizan diferentes ramas en un repositorio de código fuente. El flujo de trabajo de Gitflow se considera heredado, y el flujo de trabajo basado en enlaces troncales es el enfoque moderno preferido.

imagen dorada

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

estrategia de implementación desde cero

La ausencia de infraestructura existente en un entorno nuevo. Al adoptar una estrategia de implementación desde cero para una arquitectura de sistemas, puede seleccionar todas las tecnologías nuevas sin que estas deban ser compatibles con una infraestructura existente, lo que también se conoce como implementación sobre infraestructura existente. Si está ampliando la infraestructura existente, puede combinar las estrategias de implementación sobre infraestructuras existentes y de implementación desde cero.

barrera de protección

Una regla de alto nivel que ayuda a regular los recursos, las políticas y el cumplimiento en todas las unidades organizativas (OUs). Las barreras de protección preventivas aplican políticas para garantizar la alineación con los estándares de conformidad. Se implementan mediante políticas de control de servicios y límites de permisos de IAM. Las barreras de protección de detección detectan las vulneraciones de las políticas y los problemas de conformidad, y generan alertas para su corrección. Se implementan mediante Amazon AWS Config AWS Security Hub GuardDuty AWS Trusted Advisor, Amazon Inspector y AWS Lambda cheques personalizados.

H

HA

Consulte la alta disponibilidad.

migración heterogénea de bases de datos

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

alta disponibilidad (HA)

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

modernización histórica

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

datos retenidos

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

migración homogénea de bases de datos

Migración de la base de datos de origen a una base de datos de destino que comparte el mismo motor de base de datos (por ejemplo, Microsoft SQL Server a Amazon RDS para SQL Server). La migración homogénea suele formar parte de un esfuerzo para volver a alojar o redefinir la plataforma. Puede utilizar las utilidades de bases de datos nativas para migrar el esquema.

datos recientes

Datos a los que se accede con frecuencia, como datos en tiempo real o datos traslacionales recientes. Por lo general, estos datos requieren un nivel o una clase de almacenamiento de alto rendimiento para proporcionar respuestas rápidas a las consultas.

hotfix

Una solución urgente para un problema crítico en un entorno de producción. Debido a su urgencia, las revisiones suelen realizarse fuera del flujo de trabajo habitual de las versiones. DevOps

periodo de hiperatención

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

I

IaC

Vea la infraestructura como código.

políticas basadas en identidad

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

aplicación inactiva

Aplicación que utiliza un promedio de CPU y memoria de entre 5 y 20 por ciento durante un periodo de 90 días. En un proyecto de migración, es habitual retirar estas aplicaciones o mantenerlas en las instalaciones.

IIoT

Consulte Internet de las cosas industrial.

infraestructura inmutable

Un modelo que implementa una nueva infraestructura para las cargas de trabajo de producción en lugar de actualizar, aplicar parches o modificar la infraestructura existente. Las infraestructuras inmutables son intrínsecamente más consistentes, fiables y predecibles que las infraestructuras mutables. Para obtener más información, consulte las prácticas recomendadas para implementar con una infraestructura inmutable en Well-Architected Framework AWS .

VPC entrante (de entrada)

En una arquitectura de AWS cuentas múltiples, una VPC que acepta, inspecciona y enruta las conexiones de red desde fuera de una aplicación. La arquitectura AWS de referencia de seguridad recomienda configurar la cuenta de red con entradas, salidas e inspección VPCs para proteger la interfaz bidireccional entre la aplicación y el resto de Internet.

migración gradual

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

Industria 4.0

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

infraestructura

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

infraestructura como código (IaC)

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

Internet de las cosas industrial (T) IIo

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

VPC de inspección

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

Internet de las cosas (IoT)

Red de objetos físicos conectados con sensores o procesadores integrados que se comunican con otros dispositivos y sistemas a través de Internet o de una red de comunicación local. Para obtener más información, consulte ¿Qué es IoT?.

interpretabilidad

Característica de un modelo de machine learning que describe el grado en que un ser humano puede entender cómo las predicciones del modelo dependen de sus entradas. Para obtener más información, consulte Interpretabilidad del modelo de aprendizaje automático con. AWS

IoT

Consulte Internet de las cosas.

biblioteca de información de TI (ITIL)

Conjunto de prácticas recomendadas para ofrecer servicios de TI y alinearlos con los requisitos empresariales. La ITIL proporciona la base para la ITSM.

administración de servicios de TI (ITSM)

Actividades asociadas con el diseño, la implementación, la administración y el soporte de los servicios de TI para una organización. Para obtener información sobre la integración de las operaciones en la nube con las herramientas de ITSM, consulte la Guía de integración de operaciones.

ITIL

Consulte la biblioteca de información de TI.

ITSM

Consulte Administración de servicios de TI.

L

control de acceso basado en etiquetas (LBAC)

Una implementación del control de acceso obligatorio (MAC) en la que a los usuarios y a los propios datos se les asigna explícitamente un valor de etiqueta de seguridad. La intersección entre la etiqueta de seguridad del usuario y la etiqueta de seguridad de los datos determina qué filas y columnas puede ver el usuario.

zona de aterrizaje

Una landing zone es un AWS entorno multicuenta bien diseñado, escalable y seguro. Este es un punto de partida desde el cual las empresas pueden lanzar e implementar rápidamente cargas de trabajo y aplicaciones con confianza en su entorno de seguridad e infraestructura. Para obtener más información sobre las zonas de aterrizaje, consulte Configuración de un entorno de AWS seguro y escalable con varias cuentas.

modelo de lenguaje grande (LLM)

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

migración grande

Migración de 300 servidores o más.

LBAC

Consulte control de acceso basado en etiquetas.

privilegio mínimo

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

migrar mediante lift-and-shift

Ver 7 Rs.

sistema little-endian

Un sistema que almacena primero el byte menos significativo. Véase también endianness.

LLM

Véase un modelo de lenguaje amplio.

entornos inferiores

Véase entorno.

M

machine learning (ML)

Un tipo de inteligencia artificial que utiliza algoritmos y técnicas para el reconocimiento y el aprendizaje de patrones. El ML analiza y aprende de los datos registrados, como los datos del Internet de las cosas (IoT), para generar un modelo estadístico basado en patrones. Para más información, consulte Machine learning.

rama principal

Ver sucursal.

malware

Software diseñado para comprometer la seguridad o la privacidad de la computadora. El malware puede interrumpir los sistemas informáticos, filtrar información confidencial u obtener acceso no autorizado. Algunos ejemplos de malware son los virus, los gusanos, el ransomware, los troyanos, el spyware y los registradores de pulsaciones de teclas.

servicios gestionados

Servicios de AWS para los que AWS opera la capa de infraestructura, el sistema operativo y las plataformas, y usted accede a los puntos finales para almacenar y recuperar datos. Amazon Simple Storage Service (Amazon S3) y Amazon DynamoDB son ejemplos de servicios gestionados. También se conocen como servicios abstractos.

sistema de ejecución de fabricación (MES)

Un sistema de software para rastrear, monitorear, documentar y controlar los procesos de producción que convierten las materias primas en productos terminados en el taller.

MAP

Consulte Migration Acceleration Program.

mecanismo

Un proceso completo en el que se crea una herramienta, se impulsa su adopción y, a continuación, se inspeccionan los resultados para realizar ajustes. Un mecanismo es un ciclo que se refuerza y mejora a sí mismo a medida que funciona. Para obtener más información, consulte Creación de mecanismos en el AWS Well-Architected Framework.

cuenta de miembro

Todas las Cuentas de AWS demás cuentas, excepto la de administración, que forman parte de una organización. AWS Organizations Una cuenta no puede pertenecer a más de una organización a la vez.

MES

Consulte el sistema de ejecución de la fabricación.

Transporte telemétrico de Message Queue Queue (MQTT)

Un protocolo de comunicación ligero machine-to-machine (M2M), basado en el patrón de publicación/suscripción, para dispositivos de IoT con recursos limitados.

microservicio

Un servicio pequeño e independiente que se comunica a través de una red bien definida APIs y que, por lo general, es propiedad de equipos pequeños e independientes. Por ejemplo, un sistema de seguros puede incluir microservicios que se adapten a las capacidades empresariales, como las de ventas o marketing, o a subdominios, como las de compras, reclamaciones o análisis. Los beneficios de los microservicios incluyen la agilidad, la escalabilidad flexible, la facilidad de implementación, el código reutilizable y la resiliencia. Para obtener más información, consulte Integrar microservicios mediante AWS servicios sin servidor.

arquitectura de microservicios

Un enfoque para crear una aplicación con componentes independientes que ejecutan cada proceso de la aplicación como un microservicio. Estos microservicios se comunican a través de una interfaz bien definida mediante un uso ligero. APIs Cada microservicio de esta arquitectura se puede actualizar, implementar y escalar para satisfacer la demanda de funciones específicas de una aplicación. Para obtener más información, consulte Implementación de microservicios en. AWS

Programa de aceleración de la migración (MAP)

Un AWS programa que proporciona soporte de consultoría, formación y servicios para ayudar a las organizaciones a crear una base operativa sólida para migrar a la nube y para ayudar a compensar el costo inicial de las migraciones. El MAP incluye una metodología de migración para ejecutar las migraciones antiguas de forma metódica y un conjunto de herramientas para automatizar y acelerar los escenarios de migración más comunes.

migración a escala

Proceso de transferencia de la mayoría de la cartera de aplicaciones a la nube en oleadas, con más aplicaciones desplazadas a un ritmo más rápido en cada oleada. En esta fase, se utilizan las prácticas recomendadas y las lecciones aprendidas en las fases anteriores para implementar una fábrica de migración de equipos, herramientas y procesos con el fin de agilizar la migración de las cargas de trabajo mediante la automatización y la entrega ágil. Esta es la tercera fase de la estrategia de migración de AWS.

fábrica de migración

Equipos multifuncionales que agilizan la migración de las cargas de trabajo mediante enfoques automatizados y ágiles. Los equipos de las fábricas de migración suelen incluir a analistas y propietarios de operaciones, empresas, ingenieros de migración, desarrolladores y DevOps profesionales que trabajan a pasos agigantados. Entre el 20 y el 50 por ciento de la cartera de aplicaciones empresariales se compone de patrones repetidos que pueden optimizarse mediante un enfoque de fábrica. Para obtener más información, consulte la discusión sobre las fábricas de migración y la Guía de fábricas de migración a la nube en este contenido.

metadatos de migración

Información sobre la aplicación y el servidor que se necesita para completar la migración. Cada patrón de migración requiere un conjunto diferente de metadatos de migración. Algunos ejemplos de metadatos de migración son la subred de destino, el grupo de seguridad y AWS la cuenta.

patrón de migración

Tarea de migración repetible que detalla la estrategia de migración, el destino de la migración y la aplicación o el servicio de migración utilizados. Ejemplo: realoje la migración a Amazon EC2 con AWS Application Migration Service.

Migration Portfolio Assessment (MPA)

Una herramienta en línea que proporciona información para validar el modelo de negocio para migrar a. Nube de AWS La MPA ofrece una evaluación detallada de la cartera (adecuación del

tamaño de los servidores, precios, comparaciones del costo total de propiedad, análisis de los costos de migración), así como una planificación de la migración (análisis y recopilación de datos de aplicaciones, agrupación de aplicaciones, priorización de la migración y planificación de oleadas). La herramienta MPA (requiere iniciar sesión) está disponible de forma gratuita para todos los AWS consultores y consultores asociados de APN.

Evaluación de la preparación para la migración (MRA)

Proceso que consiste en obtener información sobre el estado de preparación de una organización para la nube, identificar sus puntos fuertes y débiles y elaborar un plan de acción para cerrar las brechas identificadas mediante el AWS CAF. Para obtener más información, consulte la Guía de preparación para la migración. La MRA es la primera fase de la estrategia de migración de AWS.

estrategia de migración

El enfoque utilizado para migrar una carga de trabajo a. Nube de AWS Para obtener más información, consulte la entrada de las 7 R de este glosario y consulte Movilice a su organización para acelerar las migraciones a gran escala.

ML

Consulte el aprendizaje automático.

modernización

Transformar una aplicación obsoleta (antigua o monolítica) y su infraestructura en un sistema ágil, elástico y de alta disponibilidad en la nube para reducir los gastos, aumentar la eficiencia y aprovechar las innovaciones. Para obtener más información, consulte Estrategia para modernizar las aplicaciones en el Nube de AWS.

evaluación de la preparación para la modernización

Evaluación que ayuda a determinar la preparación para la modernización de las aplicaciones de una organización; identifica los beneficios, los riesgos y las dependencias; y determina qué tan bien la organización puede soportar el estado futuro de esas aplicaciones. El resultado de la evaluación es un esquema de la arquitectura objetivo, una hoja de ruta que detalla las fases de desarrollo y los hitos del proceso de modernización y un plan de acción para abordar las brechas identificadas. Para obtener más información, consulte Evaluación de la preparación para la modernización de las aplicaciones en el Nube de AWS.

aplicaciones monolíticas (monolitos)

Aplicaciones que se ejecutan como un único servicio con procesos estrechamente acoplados. Las aplicaciones monolíticas presentan varios inconvenientes. Si una característica de la

aplicación experimenta un aumento en la demanda, se debe escalar toda la arquitectura. Agregar o mejorar las características de una aplicación monolítica también se vuelve más complejo a medida que crece la base de código. Para solucionar problemas con la aplicación, puede utilizar una arquitectura de microservicios. Para obtener más información, consulte Descomposición de monolitos en microservicios.

MAPA

Consulte la evaluación de la cartera de migración.

MQTT

Consulte Message Queue Queue Telemetría y Transporte.

clasificación multiclase

Un proceso que ayuda a generar predicciones para varias clases (predice uno de más de dos resultados). Por ejemplo, un modelo de ML podría preguntar “¿Este producto es un libro, un automóvil o un teléfono?” o “¿Qué categoría de productos es más interesante para este cliente?”.

infraestructura mutable

Un modelo que actualiza y modifica la infraestructura existente para las cargas de trabajo de producción. Para mejorar la coherencia, la fiabilidad y la previsibilidad, el AWS Well-Architected Framework recomienda el uso de una infraestructura inmutable como práctica recomendada.

O

OAC

Consulte el control de acceso de origen.

OAI

Consulte la identidad de acceso de origen.

OCM

Consulte gestión del cambio organizacional.

migración fuera de línea

Método de migración en el que la carga de trabajo de origen se elimina durante el proceso de migración. Este método implica un tiempo de inactividad prolongado y, por lo general, se utiliza para cargas de trabajo pequeñas y no críticas.

OI

Consulte integración de operaciones.

OLA

Véase el acuerdo a nivel operativo.

migración en línea

Método de migración en el que la carga de trabajo de origen se copia al sistema de destino sin que se desconecte. Las aplicaciones que están conectadas a la carga de trabajo pueden seguir funcionando durante la migración. Este método implica un tiempo de inactividad nulo o mínimo y, por lo general, se utiliza para cargas de trabajo de producción críticas.

OPC-UA

Consulte Open Process Communications: arquitectura unificada.

Comunicaciones de proceso abierto: arquitectura unificada (OPC-UA)

Un protocolo de comunicación machine-to-machine (M2M) para la automatización industrial. El OPC-UA proporciona un estándar de interoperabilidad con esquemas de cifrado, autenticación y autorización de datos.

acuerdo de nivel operativo (OLA)

Acuerdo que aclara lo que los grupos de TI operativos se comprometen a ofrecerse entre sí, para respaldar un acuerdo de nivel de servicio (SLA).

revisión de la preparación operativa (ORR)

Una lista de preguntas y las mejores prácticas asociadas que le ayudan a comprender, evaluar, prevenir o reducir el alcance de los incidentes y posibles fallos. Para obtener más información, consulte Operational Readiness Reviews (ORR) en AWS Well-Architected Framework.

tecnología operativa (OT)

Sistemas de hardware y software que funcionan con el entorno físico para controlar las operaciones, los equipos y la infraestructura industriales. En la industria manufacturera, la integración de los sistemas de TO y tecnología de la información (TI) es un enfoque clave para las transformaciones de la industria 4.0.

integración de operaciones (OI)

Proceso de modernización de las operaciones en la nube, que implica la planificación de la preparación, la automatización y la integración. Para obtener más información, consulte la Guía de integración de las operaciones.

registro de seguimiento organizativo

Un registro creado por el AWS CloudTrail que se registran todos los eventos para todos Cuentas de AWS los miembros de una organización AWS Organizations. Este registro de seguimiento se crea en cada Cuenta de AWS que forma parte de la organización y realiza un seguimiento de la actividad en cada cuenta. Para obtener más información, consulte Crear un registro para una organización en la CloudTrail documentación.

administración del cambio organizacional (OCM)

Marco para administrar las transformaciones empresariales importantes y disruptivas desde la perspectiva de las personas, la cultura y el liderazgo. La OCM ayuda a las empresas a prepararse para nuevos sistemas y estrategias y a realizar la transición a ellos, al acelerar la adopción de cambios, abordar los problemas de transición e impulsar cambios culturales y organizacionales. En la estrategia de AWS migración, este marco se denomina aceleración de personal, debido a la velocidad de cambio que requieren los proyectos de adopción de la nube. Para obtener más información, consulte la Guía de OCM.

control de acceso de origen (OAC)

En CloudFront, una opción mejorada para restringir el acceso y proteger el contenido del Amazon Simple Storage Service (Amazon S3). El OAC admite todos los buckets de S3 Regiones de AWS, el cifrado del lado del servidor AWS KMS (SSE-KMS) y las solicitudes dinámicas `PUT` y `DELETE` dirigidas al bucket de S3.

identidad de acceso de origen (OAI)

En CloudFront, una opción para restringir el acceso y proteger el contenido de Amazon S3. Cuando utiliza OAI, CloudFront crea un principal con el que Amazon S3 puede autenticarse. Los directores autenticados solo pueden acceder al contenido de un bucket de S3 a través de una distribución específica. CloudFront Consulte también el OAC, que proporciona un control de acceso más detallado y mejorado.

ORR

Consulte la revisión de la preparación operativa.

OT

Consulte la tecnología operativa.

VPC saliente (de salida)

En una arquitectura de AWS cuentas múltiples, una VPC que gestiona las conexiones de red que se inician desde una aplicación. La arquitectura AWS de referencia de seguridad recomienda configurar la cuenta de red con entradas, salidas e inspección VPCs para proteger la interfaz bidireccional entre la aplicación e Internet en general.

P

límite de permisos

Una política de administración de IAM que se adjunta a las entidades principales de IAM para establecer los permisos máximos que puede tener el usuario o el rol. Para obtener más información, consulte Límites de permisos en la documentación de IAM.

información de identificación personal (PII)

Información que, vista directamente o combinada con otros datos relacionados, puede utilizarse para deducir de manera razonable la identidad de una persona. Algunos ejemplos de información de identificación personal son los nombres, las direcciones y la información de contacto.

PII

Consulte la información de identificación personal.

manual de estrategias

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

PLC

Consulte controlador lógico programable.

PLM

Consulte la gestión del ciclo de vida del producto.

policy

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

persistencia políglota

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

evaluación de cartera

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

predicate

Una condición de consulta que devuelve `true` o`false`, por lo general, se encuentra en una cláusula. `WHERE`

pulsar un predicado

Técnica de optimización de consultas de bases de datos que filtra los datos de la consulta antes de transferirlos. Esto reduce la cantidad de datos que se deben recuperar y procesar de la base de datos relacional y mejora el rendimiento de las consultas.

control preventivo

Un control de seguridad diseñado para evitar que ocurra un evento. Estos controles son la primera línea de defensa para evitar el acceso no autorizado o los cambios no deseados en la red. Para obtener más información, consulte Controles preventivos en Implementación de controles de seguridad en AWS.

entidad principal

Una entidad AWS que puede realizar acciones y acceder a los recursos. Esta entidad suele ser un usuario raíz para un Cuenta de AWS rol de IAM o un usuario. Para obtener más información, consulte Entidad principal en Términos y conceptos de roles en la documentación de IAM.

privacidad desde el diseño

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

zonas alojadas privadas

Un contenedor que contiene información sobre cómo desea que Amazon Route 53 responda a las consultas de DNS de un dominio y sus subdominios dentro de uno o más VPCs. Para obtener más información, consulte Uso de zonas alojadas privadas en la documentación de Route 53.

control proactivo

Un control de seguridad diseñado para evitar el despliegue de recursos no conformes. Estos controles escanean los recursos antes de aprovisionarlos. Si el recurso no cumple con el control, significa que no está aprovisionado. Para obtener más información, consulte la guía de referencia de controles en la AWS Control Tower documentación y consulte Controles proactivos en Implementación de controles de seguridad en AWS.

gestión del ciclo de vida del producto (PLM)

La gestión de los datos y los procesos de un producto a lo largo de todo su ciclo de vida, desde el diseño, el desarrollo y el lanzamiento, pasando por el crecimiento y la madurez, hasta el rechazo y la retirada.

entorno de producción

Consulte el entorno.

controlador lógico programable (PLC)

En la fabricación, una computadora adaptable y altamente confiable que monitorea las máquinas y automatiza los procesos de fabricación.

encadenamiento rápido

Utilizar la salida de una solicitud de LLM como entrada para la siguiente solicitud para generar mejores respuestas. Esta técnica se utiliza para dividir una tarea compleja en subtareas o para

refinar o ampliar de forma iterativa una respuesta preliminar. Ayuda a mejorar la precisión y la relevancia de las respuestas de un modelo y permite obtener resultados más detallados y personalizados.

seudonimización

El proceso de reemplazar los identificadores personales de un conjunto de datos por valores de marcadores de posición. La seudonimización puede ayudar a proteger la privacidad personal. Los datos seudonimizados siguen considerándose datos personales.

publish/subscribe (pub/sub)

Un patrón que permite las comunicaciones asíncronas entre microservicios para mejorar la escalabilidad y la capacidad de respuesta. Por ejemplo, en un MES basado en microservicios, un microservicio puede publicar mensajes de eventos en un canal al que se puedan suscribir otros microservicios. El sistema puede añadir nuevos microservicios sin cambiar el servicio de publicación.

Q

plan de consulta

Serie de pasos, como instrucciones, que se utilizan para acceder a los datos de un sistema de base de datos relacional SQL.

regresión del plan de consulta

El optimizador de servicios de la base de datos elige un plan menos óptimo que antes de un cambio determinado en el entorno de la base de datos. Los cambios en estadísticas, restricciones, configuración del entorno, enlaces de parámetros de consultas y actualizaciones del motor de base de datos PostgreSQL pueden provocar una regresión del plan.

R

Matriz RACI

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

RAG

Consulte Retrieval Augmented Generation.

ransomware

Software malicioso que se ha diseñado para bloquear el acceso a un sistema informático o a los datos hasta que se efectúe un pago.

Matriz RASCI

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

RCAC

Consulte control de acceso por filas y columnas.

réplica de lectura

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

rediseñar

Ver 7 Rs.

objetivo de punto de recuperación (RPO)

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

objetivo de tiempo de recuperación (RTO)

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

refactorizar

Ver 7 Rs.

Región

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

regresión

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

volver a alojar

Consulte 7 Rs.

versión

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

trasladarse

Ver 7 Rs.

redefinir la plataforma

Ver 7 Rs.

recompra

Ver 7 Rs.

resiliencia

La capacidad de una aplicación para resistir las interrupciones o recuperarse de ellas. La alta disponibilidad y la recuperación ante desastres son consideraciones comunes a la hora de planificar la resiliencia en el. Nube de AWS Para obtener más información, consulte Nube de AWS Resiliencia.

política basada en recursos

Una política asociada a un recurso, como un bucket de Amazon S3, un punto de conexión o una clave de cifrado. Este tipo de política especifica a qué entidades principales se les permite el acceso, las acciones compatibles y cualquier otra condición que deba cumplirse.

matriz responsable, confiable, consultada e informada (RACI)

Una matriz que define las funciones y responsabilidades de todas las partes involucradas en las actividades de migración y las operaciones de la nube. El nombre de la matriz se deriva de los tipos de responsabilidad definidos en la matriz: responsable (R), contable (A), consultado (C) e informado (I). El tipo de soporte (S) es opcional. Si incluye el soporte, la matriz se denomina matriz RASCI y, si la excluye, se denomina matriz RACI.

control receptivo

Un control de seguridad que se ha diseñado para corregir los eventos adversos o las desviaciones con respecto a su base de seguridad. Para obtener más información, consulte Controles receptivos en Implementación de controles de seguridad en AWS.

retain

Consulte 7 Rs.

jubilarse

Ver 7 Rs.

Generación aumentada de recuperación (RAG)

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

rotación

Proceso de actualizar periódicamente un secreto para dificultar el acceso de un atacante a las credenciales.

control de acceso por filas y columnas (RCAC)

El uso de expresiones SQL básicas y flexibles que tienen reglas de acceso definidas. El RCAC consta de permisos de fila y máscaras de columnas.

RPO

Consulte el objetivo del punto de recuperación.

RTO

Consulte el objetivo de tiempo de recuperación.

manual de procedimientos

Conjunto de procedimientos manuales o automatizados necesarios para realizar una tarea específica. Por lo general, se diseñan para agilizar las operaciones o los procedimientos repetitivos con altas tasas de error.

S

SAML 2.0

Un estándar abierto que utilizan muchos proveedores de identidad (IdPs). Esta función permite el inicio de sesión único (SSO) federado, de modo que los usuarios pueden iniciar sesión AWS

Management Console o llamar a las operaciones de la AWS API sin tener que crear un usuario en IAM para todos los miembros de la organización. Para obtener más información sobre la federación basada en SAML 2.0, consulte Acerca de la federación basada en SAML 2.0 en la documentación de IAM.

SCADA

Consulte el control de supervisión y la adquisición de datos.

SCP

Consulte la política de control de servicios.

secreta

Información confidencial o restringida, como una contraseña o credenciales de usuario, que almacene de forma cifrada. AWS Secrets Manager Se compone del valor secreto y sus metadatos. El valor secreto puede ser binario, una sola cadena o varias cadenas. Para obtener más información, consulta ¿Qué hay en un secreto de Secrets Manager? en la documentación de Secrets Manager.

seguridad desde el diseño

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

control de seguridad

Barrera de protección técnica o administrativa que impide, detecta o reduce la capacidad de un agente de amenazas para aprovechar una vulnerabilidad de seguridad. Existen cuatro tipos principales de controles de seguridad: preventivos, de detección, con capacidad de respuesta y proactivos.

refuerzo de la seguridad

Proceso de reducir la superficie expuesta a ataques para hacerla más resistente a los ataques. Esto puede incluir acciones, como la eliminación de los recursos que ya no se necesitan, la implementación de prácticas recomendadas de seguridad consistente en conceder privilegios mínimos o la desactivación de características innecesarias en los archivos de configuración.

sistema de información sobre seguridad y administración de eventos (SIEM)

Herramientas y servicios que combinan sistemas de administración de información sobre seguridad (SIM) y de administración de eventos de seguridad (SEM). Un sistema de SIEM

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

automatización de la respuesta de seguridad

Una acción predefinida y programada que está diseñada para responder automáticamente a un evento de seguridad o remediarlo. Estas automatizaciones sirven como controles de seguridad detectables o adaptables que le ayudan a implementar las mejores prácticas AWS de seguridad. Algunos ejemplos de acciones de respuesta automatizadas incluyen la modificación de un grupo de seguridad de VPC, la aplicación de parches a una EC2 instancia de Amazon o la rotación de credenciales.

cifrado del servidor

Cifrado de los datos en su destino, por parte de quien Servicio de AWS los recibe.

política de control de servicio (SCP)

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

punto de enlace de servicio

La URL del punto de entrada de un Servicio de AWS. Para conectarse mediante programación a un servicio de destino, puede utilizar un punto de conexión. Para obtener más información, consulte Puntos de conexión de Servicio de AWS en Referencia general de AWS.

acuerdo de nivel de servicio (SLA)

Acuerdo que aclara lo que un equipo de TI se compromete a ofrecer a los clientes, como el tiempo de actividad y el rendimiento del servicio.

indicador de nivel de servicio (SLI)

Medición de un aspecto del rendimiento de un servicio, como la tasa de errores, la disponibilidad o el rendimiento.

objetivo de nivel de servicio (SLO)

Una métrica objetivo que representa el estado de un servicio, medido mediante un indicador de nivel de servicio.

modelo de responsabilidad compartida

Un modelo que describe la responsabilidad que compartes con respecto a la seguridad y AWS el cumplimiento de la nube. AWS es responsable de la seguridad de la nube, mientras que usted es responsable de la seguridad en la nube. Para obtener más información, consulte el Modelo de responsabilidad compartida.

SIEM

Consulte la información de seguridad y el sistema de gestión de eventos.

punto único de fallo (SPOF)

Una falla en un único componente crítico de una aplicación que puede interrumpir el sistema.

SLA

Consulte el acuerdo de nivel de servicio.

SLI

Consulte el indicador de nivel de servicio.

SLO

Consulte el objetivo de nivel de servicio.

split-and-seed modelo

Un patrón para escalar y acelerar los proyectos de modernización. A medida que se definen las nuevas funciones y los lanzamientos de los productos, el equipo principal se divide para crear nuevos equipos de productos. Esto ayuda a ampliar las capacidades y los servicios de su organización, mejora la productividad de los desarrolladores y apoya la innovación rápida. Para obtener más información, consulte Enfoque gradual para modernizar las aplicaciones en el. Nube de AWS

SPOF

Consulte el punto único de falla.

esquema en forma de estrella

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

patrón de higo estrangulador

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

subred

Un intervalo de direcciones IP en la VPC. Una subred debe residir en una sola zona de disponibilidad.

supervisión, control y adquisición de datos (SCADA)

En la industria manufacturera, un sistema que utiliza hardware y software para monitorear los activos físicos y las operaciones de producción.

cifrado simétrico

Un algoritmo de cifrado que utiliza la misma clave para cifrar y descifrar los datos.

pruebas sintéticas

Probar un sistema de manera que simule las interacciones de los usuarios para detectar posibles problemas o monitorear el rendimiento. Puede usar Amazon CloudWatch Synthetics para crear estas pruebas.

indicador del sistema

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

T

etiquetas

Pares clave-valor que actúan como metadatos para organizar los recursos. AWS Las etiquetas pueden ayudarle a administrar, identificar, organizar, buscar y filtrar recursos. Para obtener más información, consulte Etiquetado de los recursos de AWS.

variable de destino

El valor que intenta predecir en el ML supervisado. Esto también se conoce como variable de resultado. Por ejemplo, en un entorno de fabricación, la variable objetivo podría ser un defecto del producto.

lista de tareas

Herramienta que se utiliza para hacer un seguimiento del progreso mediante un manual de procedimientos. La lista de tareas contiene una descripción general del manual de procedimientos y una lista de las tareas generales que deben completarse. Para cada tarea general, se incluye la cantidad estimada de tiempo necesario, el propietario y el progreso.

entorno de prueba

Consulte entorno.

entrenamiento

Proporcionar datos de los que pueda aprender su modelo de ML. Los datos de entrenamiento deben contener la respuesta correcta. El algoritmo de aprendizaje encuentra patrones en los datos de entrenamiento que asignan los atributos de los datos de entrada al destino (la respuesta que desea predecir). Genera un modelo de ML que captura estos patrones. Luego, el modelo de ML se puede utilizar para obtener predicciones sobre datos nuevos para los que no se conoce el destino.

puerta de enlace de tránsito

Un centro de tránsito de red que puede usar para interconectar sus VPCs redes con las locales. Para obtener más información, consulte Qué es una pasarela de tránsito en la AWS Transit Gateway documentación.

flujo de trabajo basado en enlaces troncales

Un enfoque en el que los desarrolladores crean y prueban características de forma local en una rama de característica y, a continuación, combinan esos cambios en la rama principal. Luego, la rama principal se adapta a los entornos de desarrollo, preproducción y producción, de forma secuencial.

acceso de confianza

Otorgar permisos a un servicio que especifique para realizar tareas en su organización AWS Organizations y en sus cuentas en su nombre. El servicio de confianza crea un rol vinculado al servicio en cada cuenta, cuando ese rol es necesario, para realizar las tareas de administración

por usted. Para obtener más información, consulte AWS Organizations Utilización con otros AWS servicios en la AWS Organizations documentación.

ajuste

Cambiar aspectos de su proceso de formación a fin de mejorar la precisión del modelo de ML. Por ejemplo, puede entrenar el modelo de ML al generar un conjunto de etiquetas, incorporar etiquetas y, luego, repetir estos pasos varias veces con diferentes ajustes para optimizar el modelo.

equipo de dos pizzas

Un DevOps equipo pequeño al que puedes alimentar con dos pizzas. Un equipo formado por dos integrantes garantiza la mejor oportunidad posible de colaboración en el desarrollo de software.

U

incertidumbre

Un concepto que hace referencia a información imprecisa, incompleta o desconocida que puede socavar la fiabilidad de los modelos predictivos de ML. Hay dos tipos de incertidumbre: la incertidumbre epistémica se debe a datos limitados e incompletos, mientras que la incertidumbre aleatoria se debe al ruido y la aleatoriedad inherentes a los datos. Para más información, consulte la guía Cuantificación de la incertidumbre en los sistemas de aprendizaje profundo.

tareas indiferenciadas

También conocido como tareas arduas, es el trabajo que es necesario para crear y operar una aplicación, pero que no proporciona un valor directo al usuario final ni proporciona una ventaja competitiva. Algunos ejemplos de tareas indiferenciadas son la adquisición, el mantenimiento y la planificación de la capacidad.

entornos superiores

Ver entorno.

V

succión

Una operación de mantenimiento de bases de datos que implica limpiar después de las actualizaciones incrementales para recuperar espacio de almacenamiento y mejorar el rendimiento.

control de versión

Procesos y herramientas que realizan un seguimiento de los cambios, como los cambios en el código fuente de un repositorio.

Emparejamiento de VPC

Una conexión entre dos VPCs que le permite enrutar el tráfico mediante direcciones IP privadas. Para obtener más información, consulte ¿Qué es una interconexión de VPC? en la documentación de Amazon VPC.

vulnerabilidad

Defecto de software o hardware que pone en peligro la seguridad del sistema.

W

caché caliente

Un búfer caché que contiene datos actuales y relevantes a los que se accede con frecuencia. La instancia de base de datos puede leer desde la caché del búfer, lo que es más rápido que leer desde la memoria principal o el disco.

datos templados

Datos a los que el acceso es infrecuente. Al consultar este tipo de datos, normalmente se aceptan consultas moderadamente lentas.

función de ventana

Función SQL que realiza un cálculo en un grupo de filas que se relacionan de alguna manera con el registro actual. Las funciones de ventana son útiles para procesar tareas, como calcular una media móvil o acceder al valor de las filas en función de la posición relativa de la fila actual.

carga de trabajo

Conjunto de recursos y código que ofrece valor comercial, como una aplicación orientada al cliente o un proceso de backend.

flujo de trabajo

Grupos funcionales de un proyecto de migración que son responsables de un conjunto específico de tareas. Cada flujo de trabajo es independiente, pero respalda a los demás flujos de trabajo del proyecto. Por ejemplo, el flujo de trabajo de la cartera es responsable de priorizar las aplicaciones, planificar las oleadas y recopilar los metadatos de migración. El flujo de trabajo de la cartera entrega estos recursos al flujo de trabajo de migración, que luego migra los servidores y las aplicaciones.

GUSANO

Mira, escribe una vez, lee muchas.

WQF

Consulte el marco AWS de calificación de la carga de trabajo.

escribe una vez, lee muchas (WORM)

Un modelo de almacenamiento que escribe los datos una sola vez y evita que los datos se eliminen o modifiquen. Los usuarios autorizados pueden leer los datos tantas veces como sea necesario, pero no pueden cambiarlos. Esta infraestructura de almacenamiento de datos se considera inmutable.

Z

ataque de día cero

Un ataque, normalmente de malware, que aprovecha una vulnerabilidad de día cero.

vulnerabilidad de día cero

Un defecto o una vulnerabilidad sin mitigación en un sistema de producción. Los agentes de amenazas pueden usar este tipo de vulnerabilidad para atacar el sistema. Los desarrolladores suelen darse cuenta de la vulnerabilidad a raíz del ataque.

aviso de tiro cero

Proporcionar a un LLM instrucciones para realizar una tarea, pero sin ejemplos (imágenes) que puedan ayudar a guiarla. El LLM debe utilizar sus conocimientos previamente entrenados para

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

aplicación zombi

Aplicación que utiliza un promedio de CPU y memoria menor al 5 por ciento. En un proyecto de migración, es habitual retirar estas aplicaciones.

Las traducciones son generadas a través de traducción automática. En caso de conflicto entre la traducción y la version original de inglés, prevalecerá la version en inglés.