



Opciones de almacenamiento en nube de AWS

Joseph Baron, Amazon Web Services

Robert Schneider, Think88

Diciembre de 2010

Introducción

Amazon Web Services (AWS) es una plataforma informática en nube fácil de utilizar, rentable y flexible que incluye varias opciones de almacenamiento de datos basadas en la nube. Estas alternativas ofrecen una amplia variedad de opciones a los arquitectos y desarrolladores.

Este documento técnico le ayuda a comprender las principales opciones de almacenamiento de datos disponibles con la plataforma de informática en nube de AWS. Ofrecemos una descripción general de cada opción de almacenamiento, describimos los escenarios de uso ideales y analizamos características importantes específicas de la nube como la elasticidad y escalabilidad. Prestamos especial atención a la identificación del nivel de durabilidad que ofrece cada opción de almacenamiento. El presente documento concluye con una tabla de referencias rápidas que compara las opciones de almacenamiento que se presentan aquí. Consulte la sección "Referencias y más documentación" para conocer más recursos.

Por último, en un [documento adjunto e independiente presentamos varios casos prácticos de almacenamiento](#) en los que se muestra cómo utilizar de forma conjunta varias opciones de almacenamiento de la nube de AWS. Puede utilizar estos casos prácticos como orientación a la hora de diseñar su propia arquitectura de almacenamiento.

Alternativas de almacenamiento tradicionales frente a las basadas en la nube

Los arquitectos de las aplicaciones y la infraestructura de TI in situ y tradicionales disponen de numerosas posibilidades para el almacenamiento de datos, entre las que destacan las siguientes:

- **Memoria:** el almacenamiento en memoria como las cachés de archivos, las cachés de objetos y los discos RAM, ofrece un acceso muy rápido a los datos.
- **Cola de mensajes:** almacenamiento duradero temporal para los datos enviados de forma asíncrona entre sistemas informáticos o componentes de aplicaciones.
- **Red de área de almacenamiento (SAN):** los dispositivos de bloque (LUN de disco virtual) de SAN dedicadas suelen ofrecer el máximo nivel de rendimiento y durabilidad del disco para el almacenamiento de bases de datos y de datos de archivo importantes para la empresa, pero sin duda alguna se trata de uno de los sistemas más caros.
- **Almacenamiento conectado directo (DAS):** las matrices o las unidades de disco duro locales que residen en cada servidor ofrecen mayor rendimiento que una SAN, pero menos durabilidad para archivos temporales y persistentes, almacenamiento de base de datos y almacenamiento de inicio de sistemas operativos en comparación con una SAN.
- **Almacenamiento conectado a la red (NAS):** el almacenamiento NAS ofrece una interfaz de almacenamiento de nivel de archivo que puede compartirse entre varios sistemas. NAS suele ser un tipo de almacenamiento más lento que SAN o DAS.
- **Bases de datos:** el acceso a los datos estructurados y su mantenimiento se suele realizar a través de una base de datos relacional como MySQL, PostgreSQL, Oracle, Microsoft SQL Server y DB2, o bien en repositorios de bases de datos no relacionales. Los volúmenes de almacenamiento de base de datos normalmente residen en dispositivos SAN o DAS.

- **Sin conexión:** los datos almacenados para copias de seguridad y fines de archivado normalmente se alojan en soportes que no son discos, como cintas, CD, DVD y otros, que suelen almacenarse en ubicaciones remotas seguras para casos de recuperación ante desastres.

Cada una de estas opciones de almacenamiento tradicionales difiere en términos de rendimiento, durabilidad y coste, así como en sus interfaces. Los arquitectos consideran todos estos factores a la hora de identificar la solución de almacenamiento correcta para la tarea que están desarrollando. Cabe destacar que la mayoría de las infraestructuras de TI y arquitecturas de aplicaciones utilizan varias tecnologías de almacenamiento de forma conjunta, cada una de las cuales se ha seleccionado para satisfacer las necesidades de una subclase particular de almacenamiento de datos. Estas combinaciones forman una jerarquía de niveles de almacenamiento de datos.

Como se observa en este documento, AWS ofrece varias opciones de almacenamiento basadas en la nube. Cada una tiene una combinación exclusiva de rendimiento, durabilidad, coste e interfaz, y se mejora más con factores adicionales como la elasticidad, la disponibilidad y la escalabilidad. Estos factores adicionales son importantes para las soluciones basadas en la nube y de escala web. Como bien ocurre con las aplicaciones in situ tradicionales, puede utilizar varias opciones de almacenamiento en nube de forma conjunta para formar una jerarquía completa de almacenamiento de datos.

En este documento se analizan las siguientes opciones de almacenamiento en nube de AWS:

- **Volúmenes Elastic Block Storage (EBS) de Amazon EC2**
- **Volúmenes de almacenamiento de instancias locales de Amazon EC2 (efímeros)**
- **Amazon Simple Storage Service (Amazon S3)**
- **Amazon Simple Queue Service (SQS)**
- **Amazon SimpleDB**
- **Bases de datos relacionales de Amazon EC2**
- **Amazon Relational Database Service (RDS)**

Vamos a analizar cada una de las opciones de almacenamiento de AWS de forma más detallada. Para cada candidato, presentaremos la siguiente información:

- Nombre y descripción
- Escenario ideal de uso
- Rendimiento
- Durabilidad y disponibilidad
- Coste
- Elasticidad y escalabilidad

- Interfaces
- Los antipatrones son situaciones en las que otras opciones de almacenamiento constituirían una opción mejor.

Para conocer otras categorías comparativas entre el abanico de opciones de almacenamiento de AWS, consulte [Referencia rápida sobre el almacenamiento de AWS en la página 36](#).

Volúmenes de Amazon Elastic Block Store (EBS)

Los volúmenes de Amazon Elastic Block Store (EBS) ofrecen almacenamiento de nivel de bloque duradero para utilizarlo con las instancias de Amazon EC2 (máquinas virtuales). Los volúmenes de Amazon EBS constituyen almacenamiento fuera de la instancia y conectado a la red que persiste con independencia de la vida de ejecución de una única instancia de Amazon EC2. Después de asociar un volumen de EBS a una instancia de Amazon EC2, puede interactuar con él como lo haría con una unidad de disco duro física, normalmente mediante el formateo del mismo con un sistema de archivos de su elección. Puede utilizar un volumen de EBS para iniciar una instancia de Amazon EC2 (solo AMI de EBS) y asociar varios volúmenes de EBS a una única instancia de Amazon EC2. Tenga en cuenta, sin embargo, que cualquier volumen exclusivo de EBS puede asociarse solo a una instancia de Amazon EC2 en cualquier momento. Un volumen de EBS no se puede compartir con otros usuarios, a menos que cree una instantánea de EBS (consulte la siguiente sección "Durabilidad y disponibilidad"). Los tamaños para los volúmenes de EBS oscilan entre 1 GB y 1 TB, y se asignan en incrementos de 1 GB.

Escenario ideal de uso

Amazon EBS está pensado para los datos que cambian con relativa frecuencia y que precisan de una persistencia a largo plazo. EBS ofrece almacenamiento persistente en bloque y de carácter virtual para los servidores virtuales de Amazon EC2, de forma que puede utilizarlo de la misma forma en que utilizaría una unidad de disco duro en un servidor físico. Amazon EBS es particularmente apropiado para utilizarse como la solución de almacenamiento principal para un sistema de archivos, una base de datos o para cualquier aplicación que requiera actualizaciones muy pormenorizadas y acceso a almacenamiento de nivel de bloque sin formatear y sin procesar.

Rendimiento

Por lo general, se puede prever que los volúmenes individuales de EBS ofrezcan un rendimiento, un tiempo medio entre errores (MTTF) y una fiabilidad comparables con una unidad USB externa. Tenga en cuenta que los volúmenes de EBS se parecen a las unidades de discos locales, aunque realmente están conectados mediante la red a una instancia de Amazon EC2. Por tanto, otra E/S de red ejecutada por la instancia, así como la carga total de la red compartida, puede afectar al rendimiento de los volúmenes individuales de EBS.

Aunque cada aplicación (y su rendimiento asociado) es exclusiva, puede diseñar e implementar muchas técnicas tradicionales de optimización del rendimiento del disco con volúmenes de EBS. La combinación de Amazon EC2 y EBS le permite utilizar muchas técnicas idénticas de optimización del rendimiento que utiliza con el almacenamiento y los servidores in situ. Por ejemplo, podría crear varios volúmenes y asociarlos posteriormente a una única instancia de Amazon EC2. Con varios volúmenes de EBS asociados puede dividir la carga total de E/S de la aplicación mediante la asignación de un volumen para los datos de registro, un volumen para la base de datos y otro volumen para los datos de archivo. De forma alternativa, puede dividir los datos entre varios volúmenes de EBS utilizando un controlador de dispositivo del software RAID 0, agregando las IOP disponibles, el rendimiento de volumen total y el tamaño total del volumen.

Durabilidad y disponibilidad

Cada volumen de Amazon EBS se replica automáticamente dentro de la misma zona de disponibilidad para impedir la pérdida de datos debido al fallo de cualquier componente de hardware exclusivo. Amazon EBS también ofrece la posibilidad de crear instantáneas de volúmenes en un punto en el tiempo, que persisten en Amazon S3 (consulte la información que se facilita a continuación). Estas instantáneas se pueden utilizar como punto de partida para nuevos volúmenes de Amazon EBS y para proteger los datos que se desea conservar a largo plazo.

La durabilidad del volumen de EBS depende del tamaño del mismo y de la cantidad de datos modificados desde la última instantánea. Las instantáneas de EBS son copias de seguridad incrementales y de algún punto en el tiempo que contienen solamente bloques de los datos modificados desde la última instantánea. Los volúmenes de EBS que operan con 20 GB o menos de datos modificados desde la instantánea más reciente pueden prever un porcentaje de error anual (AFR) comprendido entre el 0,1% y el 0,5%. A fin de maximizar la durabilidad y la disponibilidad de los datos almacenados en los volúmenes de EBS, los usuarios deben realizar instantáneas de sus volúmenes de EBS con frecuencia. En el caso de que el volumen de Amazon EBS falle, todas las instantáneas de dicho volumen permanecen intactas, lo que le permite reproducir el volumen desde el último punto de instantánea.

Los volúmenes de Amazon EBS están diseñados para ofrecer una alta disponibilidad. No obstante, habida cuenta de que los volúmenes de EBS se crean en una zona de disponibilidad concreta, no se encontrarán disponibles si la zona de disponibilidad tampoco se encuentra disponible. Tenga en cuenta que mientras que un único volumen de EBS está limitado a una única zona de disponibilidad, una instantánea de EBS de un volumen se encuentra disponible en todas las zonas de disponibilidad de una región, y puede utilizar una instantánea de EBS para crear uno o varios volúmenes nuevos de EBS en cualquier zona de disponibilidad. Las instantáneas de EBS también se pueden compartir con otras cuentas de usuario. Esto ofrece un mecanismo para compartir y una copia de seguridad como "clon de disco" e "imagen de disco" fácil de utilizar.

A fin de maximizar la durabilidad y la disponibilidad de los datos de EBS, los usuarios deben realizar instantáneas de sus volúmenes de EBS con frecuencia.

Coste

Al igual que ocurre con todos los servicios de Amazon Web Services, con Amazon Elastic Block Store solo ha de pagar por lo que use, sin necesidad de pagar tarifas mínimas ni de asumir contratos a largo plazo. El precio de Amazon EBS se aplica por GB mensual de almacenamiento aprovisionado y por millón de solicitudes de E/S. El almacenamiento de volumen se factura según la cantidad de asignación hasta que esta se libera. Los precios de las instantáneas de Amazon EBS se establecen por GB mensual de datos almacenados, así como por miles de solicitudes PUT y por decenas de miles de solicitudes GET cuando se guardan y cargan instantáneas. En el caso de las instantáneas de EBS, solo se le facturará por el almacenamiento que consuma *realmente*. Tenga en cuenta que las instantáneas de EBS son incrementales y comprimidas, por lo que el almacenamiento utilizado en cualquier instantánea suele ser mucho menos que el almacenamiento consumido en un volumen de EBS. La lista detallada de los precios se encuentra en <http://aws.amazon.com/ec2/pricing/>.

Tenga en cuenta que no se aplica ningún cargo por la transferencia de información entre las distintas soluciones de almacenamiento de AWS (por ejemplo, la instancia de Amazon EC2 con EBS, Amazon S3, Amazon RDS, etc.), siempre que estas se encuentren dentro de la misma región de AWS.

Elasticidad y escalabilidad

AWS le facilita la ampliación del espacio de almacenamiento disponible para la instancia de Amazon EC2. Puede crear un nuevo volumen de EBS y asociarlo a la instancia y empezar a utilizarlo junto con los existentes. Si no desea crear y mantener nuevos volúmenes, le indicamos cómo ampliar el tamaño de un único volumen:

1. Desactive la aplicación o el sistema de archivos.
2. Realice una instantánea de los datos del volumen de EBS en Amazon S3 (usando la opción “Create Snapshot from Volume”).
3. Cree un nuevo volumen de EBS a partir de la instantánea, pero especifique un tamaño superior al del volumen original.
4. Enlace el nuevo volumen más grande a la instancia de Amazon EC2.
5. Desconecte y suprima el volumen de EBS original.

Interfaces

Para crear, eliminar, describir, enlazar y disociar volúmenes de EBS para las instancias de Amazon EC2, Amazon ofrece API de control en formatos SOAP y REST. Puede utilizar las API para crear, eliminar y describir instantáneas de EBS en Amazon S3, así como sus atributos asociados. Si prefiere trabajar con herramientas gráficas, AWS Management Console y la extensión ElasticFox de Firefox le ofrece todas las capacidades de la API en una interfaz de explorador de fácil acceso. Independientemente de cómo cree el volumen de EBS, tenga en cuenta que todo el almacenamiento se asigna en el momento en que se crea el volumen y que se le facturará por este almacenamiento, incluso aunque no lo utilice.

EBS presenta una interfaz de dispositivo de bloques para la instancia de Amazon EC2. Es decir, para la instancia de Amazon EC2, un volumen de EBS aparece justo como una unidad de disco local. Para grabar datos en volúmenes de EBS y para leerlos, debe utilizar interfaces del sistema de archivos nativo del sistema operativo que elija.

Antipatrones de Amazon EBS

Como se ha descrito previamente, EBS es ideal para la información que necesita conservarse más allá del tiempo que dura una única instancia de Amazon EC2. Sin embargo, en determinadas situaciones, es posible que otras opciones de almacenamiento de AWS sean más apropiadas:

- **Almacenamiento temporal:** si no le preocupa el hecho de que los datos sigan siendo accesibles después de que se haya terminado una instancia de Amazon EC2 (archivos temporales, disco de prueba, búferes y demás), considere la opción de beneficiarse del volumen de almacenamiento que se ofrece automáticamente con la mayoría de instancias de Amazon EC2. Los volúmenes efímeros se ofrecen sin coste adicional al coste estándar de la instancia de Amazon EC2. Consulte la siguiente sección para obtener información adicional.
- **Almacenamiento de larga duración:** si necesita almacenamiento muy duradero, utilice Amazon S3. El almacenamiento estándar de Amazon S3 está diseñado para ofrecer una durabilidad anual por objeto del 99,999999999%. En cambio, los volúmenes de EBS con menos de 20 GB de datos modificados desde la última instantánea están diseñados para ofrecer una duración anual comprendida entre el 99,5% y el 99,9%; se prevé que los volúmenes con más datos modificados ofrezcan una durabilidad proporcionalmente inferior.

- **Datos estáticos o contenido web:** si sus datos no cambian con tanta frecuencia, Amazon S3 puede representar una solución escalable y más eficaz para el almacenamiento de información fija. Además, el contenido web distribuido por EBS requiere que un servidor web se ejecute en Amazon EC2, al mismo tiempo que puede entregar contenido web directamente fuera de Amazon S3.
- **Información sobre los pares de valores de claves:** tal como se describe en la sección “Amazon SimpleDB”, en la página 15, Amazon SimpleDB ofrece recursos escalables y elásticos para el almacenamiento de datos “sin esquema”. Si intenta utilizar tecnología relacional (incluidos volúmenes asociados de EBS) para mantener información de valor de claves, cabe estudiar la posibilidad de utilizar SimpleDB como una alternativa.

Volúmenes de almacenamiento de instancias locales de Amazon Elastic Compute Cloud (EC2)

Los volúmenes de almacenamiento de instancias locales de Amazon EC2 (denominados también unidades efímeras) ofrecen almacenamiento temporal de nivel de bloque para las instancias de Amazon EC2. Cuando crea una instancia de Amazon EC2 a partir de una imagen de máquina de Amazon (AMI), en la mayoría de los casos incorpora un bloque preconfigurado de almacenamiento en disco enlazado previamente.¹ A diferencia de los volúmenes de EBS, los datos sobre volúmenes de almacenamiento de instancias se conservan solo durante el tiempo que dure la instancia asociada de Amazon EC2. La cantidad de almacenamiento del disco oscila entre 160 GB y 1,7 TB, y varía en función del tipo de instancia de Amazon EC2. Las instancias más grandes de Amazon EC2 tienen más volúmenes de almacenamiento de instancia y más grandes. No obstante, aunque estas cantidades parecen ser flexibles y pueden resultar útiles a menudo, este almacenamiento es temporal y resulta de mayor utilidad como volumen de prueba o disco RAM.

Escenario ideal de uso

Los volúmenes de almacenamiento de instancias locales son ideales para el almacenamiento temporal de información que cambia constantemente, como los búferes, las cachés, los datos de pruebas y otro contenido temporal, o para los datos que se replican a través de un conjunto de instancias, como un conjunto equilibrado de carga de servidores web. El almacenamiento de instancias de Amazon EC2 está diseñado para este propósito. Se compone del dispositivo de inicio de la máquina virtual (por ejemplo, solo las AMI de almacenamiento), además de contener uno o varios volúmenes adicionales dedicados a la instancia de Amazon EC2 (para las AMI de EBS y las AMI de almacenamiento de instancias). Este almacenamiento se utiliza solo desde una única instancia de Amazon EC2 durante su vida útil. A diferencia de los volúmenes de EBS, los volúmenes de almacenamiento de instancias no se pueden asociar a otra instancia ni tampoco disociarse de ella.

Rendimiento

Habida cuenta de que la máquina virtual de la instancia de Amazon EC2 y los volúmenes de almacenamiento de instancias locales están ubicados en el mismo servidor físico, la interacción con este almacenamiento es muy rápida, en particular, para el acceso secuencial. Para aumentar las operaciones de E/S por segundo (IOPS) o para mejorar el rendimiento del disco, se pueden agrupar varios volúmenes de almacenamiento de instancias usando el software RAID 0 (fragmentación de discos).

¹ La mayoría de los tipos de instancias de Amazon EC2 ofrecen volúmenes de almacenamiento de instancias locales; no obstante, las microinstancias, como el tipo t1.micro, solo ofrecen almacenamiento de EBS. Además, las instancias que utiliza Amazon EBS para el dispositivo raíz (“inicio desde EBS0”) no exponen los volúmenes de almacenamiento de instancias efímeras de forma predeterminada. Si lo desea, puede exponer volúmenes de almacenamiento de instancias en el momento de iniciar la instancia mediante la especificación de una asignación de dispositivos de bloques. Consulte Amazon EC2 User Guide para obtener información detallada.

Durabilidad y disponibilidad

Los volúmenes de almacenamiento de instancias locales de Amazon EC2 no están previstos para utilizarse como almacenamiento de disco duradero. Los datos almacenados en volúmenes de almacenamiento de instancias locales se conservan solo durante la vida útil de la instancia de Amazon EC2. Los datos de los volúmenes de almacenamiento de instancias locales de Amazon EC2 son persistentes a través de reinicios ordenados de instancias, pero no en situaciones en las que la instancia de Amazon EC2 termina o pasa por un ciclo de fallo/reinicio.

No debe utilizar los volúmenes de almacenamiento de instancias locales para los datos que se conserven con el tiempo, como en el caso del almacenamiento en archivos permanentes o en bases de datos. No obstante, aunque los volúmenes de almacenamiento de instancias locales no son persistentes, puede conservar los datos mediante la copia o copia de seguridad periódica en EBS o Amazon S3.

Coste

El coste de la instancia de Amazon EC2 incluye todos los volúmenes de almacenamiento (efímeros) de instancias locales. Aunque no se aplican cargos por el almacenamiento de datos en volúmenes de almacenamiento de instancias locales, tenga en cuenta que los datos transferidos a y desde los volúmenes de almacenamiento de instancias de Amazon EC2 desde fuera de una región de Amazon EC2 generarán cargos por la transferencia de datos, y se aplicarán cargos adicionales por el uso de almacenamiento persistente, como Amazon S3, volúmenes de EBS e instantáneas de EBS. Puede encontrar información detallada acerca de los precios establecidos para Amazon EC2, EBS y la transferencia de datos en el sitio web <http://aws.amazon.com/ec2/pricing/>.

Elasticidad y escalabilidad

Los volúmenes de almacenamiento de instancias locales tienen un tamaño fijo para un tipo de instancia determinado de Amazon EC2 y están conectados a una instancia concreta, por lo que este tipo de almacenamiento es relativamente poco elástico. No obstante, puede conseguir la elasticidad total de almacenamiento mediante la incorporación de otras opciones de almacenamiento adecuadas como Amazon S3 o Elastic Block Storage (EBS) en su estrategia de almacenamiento de Amazon EC2.

Interfaces

La instancia de Amazon EC2 funciona con volúmenes de almacenamiento de instancias locales como si se tratara de unidades de disco locales. Esto significa que puede interactuar con los datos alojados en el almacenamiento de instancia local usando mecanismos de sistemas de archivos nativos, como NTFS en Windows o XFS en Linux. Tenga en cuenta que, en algunos casos, un dispositivo de volumen de almacenamiento de instancia local se asociará a la instancia de Amazon EC2 al iniciarla, pero debe formatearse con un sistema adecuado de archivos y montarse antes de utilizarlo.

Antipatrones de volúmenes de almacenamiento de instancias locales de Amazon EC2

Los volúmenes de almacenamiento de instancias locales de Amazon EC2 son volúmenes de pruebas rápidos y gratuitos (es decir, que se incluyen en el precio de la instancia de Amazon EC2) más adecuados para el almacenamiento de datos temporales que pueden regenerarse con facilidad. Sin embargo, en numerosas situaciones, es posible que otras opciones de almacenamiento de AWS sean más apropiadas:

- **Almacenamiento persistente:** si precisa de un almacenamiento de disco virtual persistente similar a una unidad de disco físico para archivos u otros datos que deben persistir más allá de la vida útil de una única instancia de Amazon EC2, los volúmenes de EBS o Amazon S3 son soluciones más apropiadas.
- **Almacenamiento en la base de datos:** en la mayoría de los casos, las bases de datos requieren almacenamiento que persista más allá de la vida útil de una única instancia de Amazon EC2, por lo que los volúmenes de EBS resultan la opción más adecuada.
- **Almacenamiento compartido:** los volúmenes de almacenamiento de instancias están diseñados para una única instancia de Amazon EC2, y no se pueden compartir con otros sistemas ni usuarios. Si precisa de almacenamiento que pueda disociarse de una instancia y asociarse a otra distinta, o si necesita tener la capacidad de compartir datos con facilidad, Amazon S3 o los volúmenes de EBS son la opción más adecuada.
- **Instantáneas:** si precisa de comodidad, durabilidad a largo plazo, disponibilidad y posibilidad de compartir en relación con las instantáneas del disco de un punto en el tiempo, los volúmenes de EBS son la opción más adecuada.

Amazon Simple Storage Service (Amazon S3)

Amazon S3 es una solución de almacenamiento de objetos distribuido de alta escalabilidad, durabilidad y disponibilidad diseñada para un almacenamiento de datos primarios y de vital importancia con una interfaz de servicio web fácil de utilizar. En aplicaciones in situ tradicionales, este tipo de datos, por lo general, podría mantenerse en SAN o NAS. Sin embargo, un mecanismo basado en la nube como Amazon S3 es mucho más ágil, flexible y georredundante. Amazon S3 proporciona una sencilla interfaz de servicios web que puede utilizarse para almacenar y recuperar la cantidad de datos que desee, cuando desee, ya sea desde Amazon EC2 o desde cualquier parte de la Web. Puede grabar, leer y eliminar objetos que contengan de 1 byte a 5 terabytes de datos, y el número de objetos que puede almacenar en un depósito de Amazon S3 es ilimitado. Amazon S3 también es muy escalable, permitiendo que muchos clientes independientes o subprocesos de aplicaciones puedan tener acceso simultáneo de lectura o escritura a los datos de Amazon S3.

Escenario ideal de uso

Un uso muy común de Amazon S3 consiste en el almacenamiento de contenido web estático. Este contenido puede entregarse directamente fuera de Amazon S3 a través de un servidor web, ya que cada objeto de Amazon S3 tiene una dirección URL HTTP exclusiva, o bien puede entregarse a través de la red de entrega de contenido (CDN), como Amazon CloudFront. Gracias a la elasticidad de Amazon S3, esta solución resulta particularmente apta para el alojamiento de contenido web con demandas de ancho de banda realmente altas. Además, habida cuenta de que no hay aprovisionamiento de almacenamiento, Amazon S3 funciona bien para sitios web de crecimiento rápido que alojen contenido generado por el usuario con muchos datos, como en los casos en que se comparten vídeos y fotos.

Amazon S3 se suele utilizar como almacén de datos para informática de gran escala como el análisis de transacciones financieras o datos de secuencia de clics y transcodificación de soportes. Habida cuenta de la escalabilidad horizontal de Amazon S3, puede acceder a los datos desde varios nodos informáticos simultáneamente sin estar limitado por una única conexión.

Por último, Amazon S3 suele utilizarse como el almacén de origen para datos de vital importancia, almacenamiento de instantáneas de alta fiabilidad para los volúmenes de Amazon EBS, aplicaciones de almacenamiento de copia de seguridad y soluciones principales para la recuperación ante desastres para favorecer la continuidad del negocio. Teniendo en cuenta que Amazon S3 almacena objetos de forma redundante en varios dispositivos distribuidos por varias instalaciones, ofrece la infraestructura de almacenamiento de alta durabilidad que resulta necesaria para estos escenarios.

Rendimiento

Resulta rápido acceder a Amazon S3 desde Amazon Elastic Compute Cloud (Amazon EC2) dentro de la misma región. Si accede a Amazon S3 utilizando varios subprocesos, varias aplicaciones o varios clientes de forma simultánea, el rendimiento agregado total de Amazon S3 se escalará con normalidad a índices que superen significativamente lo que un único servidor puede generar o consumir.

Para agilizar el acceso a datos relevantes, muchos desarrolladores asocian Amazon S3 con Amazon SimpleDB. Amazon S3 almacena la información real y SimpleDB funciona como el repositorio de los metadatos asociados (por ejemplo, el nombre de objetos, el tamaño, las palabras clave, etc.). SimpleDB ofrece indexación automática, por lo que resulta muy eficaz la localización de una referencia a los objetos a través de la búsqueda de metadatos. Posteriormente, este resultado puede utilizarse para identificar y a continuación recuperar el objeto de Amazon S3.

Durabilidad y disponibilidad

Mediante el almacenamiento automático y sincrónico de los datos en varios dispositivos y en varias instalaciones dentro de la región geográfica seleccionada, el almacenamiento de Amazon S3 ofrece el máximo nivel de durabilidad y disponibilidad de los datos en la plataforma de AWS. La corrección de errores está integrada y no existen puntos aislados de fallos. Amazon S3 es un servicio diseñado para soportar la pérdida simultánea de datos en dos instalaciones, por lo que resulta muy adecuado para funcionar como el almacenamiento de datos principal para los datos de vital importancia. De hecho, Amazon S3 está diseñado para ofrecer una durabilidad del 99,999999999% (11 nueves) por cada objeto y una disponibilidad del 99,99% durante un período de un año. Además de su redundancia integrada, los datos de Amazon S3 también pueden protegerse frente a fallos de aplicaciones y eliminaciones no intencionadas a través del uso del control de versiones de Amazon S3. También puede activar el control de versiones de Amazon S3 con MFA Delete. Con esta opción activada en un depósito, se precisan dos formas de autenticación para eliminar una versión de un objeto de Amazon S3: credenciales válidas de la cuenta de AWS más un código de seis dígitos (una contraseña de un solo uso basada en el tiempo) desde un dispositivo de token físico.

En el caso de los datos no importantes que pueden reproducirse fácilmente si fuera necesario, como los medios transcodificados o las miniaturas de imágenes, puede utilizar la opción del almacenamiento de redundancia reducida (RRS) en Amazon S3, que ofrece un nivel inferior de durabilidad a un coste inferior de almacenamiento. Los objetos almacenados con la utilización de la opción RRS tienen menos redundancia que los objetos almacenados con el almacenamiento estándar de Amazon S3. En cualquier caso, los datos aún siguen almacenados en varios dispositivos distribuidos en varias ubicaciones. RRS está diseñado para ofrecer una durabilidad por cada objeto del 99,99% durante un año concreto. Mientras que RRS ofrece menos durabilidad que el almacenamiento estándar de Amazon S3, es 400 veces más duradero que una unidad de disco duro típica.

Coste

Al igual que ocurre con todos los servicios de Amazon Web Services, con Amazon S3 solo ha de pagar por lo que use, sin necesidad de pagar tarifas mínimas ni de asumir contratos a largo plazo. Amazon S3 tiene tres componentes de precios: almacenamiento (por GB mensuales), transferencia de datos interna y externa (por GB mensuales) y solicitudes

(por miles de solicitudes realizadas al mes). Para los nuevos clientes, AWS ofrece una capa de uso gratuito que incluye hasta 5 GB de almacenamiento de Amazon S3. Puede encontrar información acerca de los precios totales de Amazon S3 en <http://aws.amazon.com/s3/pricing/>.

Elasticidad y escalabilidad

Amazon S3 se ha diseñado para ofrecer automáticamente un nivel muy alto de elasticidad y escalabilidad. A diferencia de un sistema de archivos habitual que experimenta problemas para almacenar una gran volumen de archivos en un directorio, Amazon S3 admite un volumen ilimitado de archivos en cualquier depósito. Además, a diferencia de cualquier unidad de disco que tiene un límite sobre el volumen total de datos que pueden almacenarse antes de que deba particionar los datos entre las unidades o los servidores, un depósito de Amazon S3 puede almacenar un volumen ilimitado de bytes. Puede almacenar cualquier cantidad de objetos, y Amazon S3 gestionará el escalado y la distribución de copias redundantes de la información en otros servidores ubicados en otras ubicaciones de la misma región, pero todas han de utilizar la infraestructura de alto rendimiento de Amazon.

Interfaces

Amazon S3 ofrece las API de servicios web SOAP y RESTful. Estas API permiten que los objetos de Amazon S3 (archivos) se almacenen en depósitos con nombres exclusivos (carpetas de nivel superior). Cada objeto debe tener una clave de objeto exclusiva (nombre de archivo) que funcione como un identificador del objeto dentro de dicho depósito. Aunque Amazon S3 es un almacén de objetos basado en la Web en lugar de un sistema de archivos tradicional, puede emular fácilmente una jerarquía de sistema de archivos (carpeta1/carpeta2/archivo) en Amazon S3 mediante la creación de nombres clave de objetos que se correspondan con el nombre de la ruta completa de cada archivo.

La mayoría de los desarrolladores que compilan aplicaciones en Amazon S3 utilizan un conjunto de herramientas de nivel superior. AWS y otros terceros han creado bibliotecas y conjuntos de herramientas de la interfaz de Amazon S3 para varios lenguajes y plataformas de desarrollo de software populares, como:

- **Java**, usando el AWS SDK para Java (<http://aws.amazon.com/sdkforjava/>) y AWS Toolkit for Eclipse (<http://aws.amazon.com/eclipse/>).
- **C#**, usando el AWS SDK para .NET (<http://aws.amazon.com/sdkfornet/>) y las plantillas de Visual Studio.
- **PHP**, usando la instalación básica de PHP5 o el AWS SDK para PHP (<http://aws.amazon.com/sdkforphp/>).
- **Perl**, usando los módulos Digest::SHA1, Bundle::LWP y XML::Simple, que se pueden descargar del sitio web de Comprehensive Perl Archive Network, en <http://www.cpan.org>.

Aunque Amazon S3 está diseñado para que se pueda acceder a este servicio principalmente a través de las API, AWS Management Console y las herramientas de terceros también ofrecen interfaces de línea de comandos gráficas para interactuar con los objetos y los depósitos de Amazon S3. Amazon S3 también admite el protocolo BitTorrent, que permite que los clientes recuperen información de forma simultánea de S3 y de otros proveedores.

Para cargar o descargar grandes volúmenes de datos, también puede utilizar el servicio AWS Import/Export (<http://aws.amazon.com/importexport/>). AWS Import/Export acelera la transferencia de grandes cantidades de datos hacia y desde AWS, utilizando dispositivos de almacenamiento portátiles para su transporte. Amazon montará el dispositivo y copiará los datos en o desde un depósito designado de Amazon S3. Este enfoque a menudo resulta más rápido y económico que transferir cantidades masivas de datos a través de Internet.

Tenga en cuenta que RRS utiliza las mismas interfaces que Amazon S3 estándar; la opción de RRS se especifica para un objeto o depósito de Amazon S3 estableciendo la propiedad de la clase de almacenamiento durante una operación PUT, o bien seleccionando la configuración “Use Reduced Redundancy Storage” de AWS Management Console.

Antipatrones de Amazon S3

Amazon S3 es un servicio óptimo para almacenar numerosas clases de información que son relativamente estáticas y que se benefician de sus características de durabilidad, disponibilidad y elasticidad. Sin embargo, en una serie de situaciones, Amazon S3 no resulta una solución óptima:

- **Sistema de archivos:** Amazon S3 utiliza un nombre de espacio sin procesar y no funciona como un sistema de archivos independiente compatible con POSIX. No obstante, mediante la utilización de delimitadores (normalmente los caracteres "/" o "\"), puede crear sus claves para emular la estructura jerárquica de carpetas del sistema de archivos dentro de un depósito determinado.
- **Datos estructurados con consultas:** Amazon S3 no ofrece funciones de consulta para recuperar un objeto específico que necesite para conocer el nombre y la clave del depósito. Por tanto, no puede utilizar Amazon S3 como una base de datos por sí sola. Un patrón de uso común consiste en colocar un objeto en Amazon S3 y utilizar Amazon SimpleDB para almacenar los metadatos de los objetos. Puede utilizar también otras tecnologías de base de datos como Amazon RDS o una base de datos que se ejecute en una instancia de Amazon EC2. También puede buscar la base de datos para localizar el nombre y la clave del depósito del objeto y, posteriormente, recuperar el objeto del depósito de Amazon S3.
- **Datos que cambian con rapidez:** los datos que han de actualizarse con mucha frecuencia pueden gestionarse mejor con una solución de almacenamiento con latencias de lectura y escritura más bajas, como EBS o una base de datos.

Amazon Simple Queue Service

Amazon Simple Queue Service (Amazon SQS) ofrece un servicio de cola de mensajes alojados fiable y muy escalable para el almacenamiento temporal y la entrega de mensajes cortos de datos de texto (hasta 64 KB). Una cola de Amazon SQS es un repositorio de datos temporales para los mensajes que están a la espera de ser procesados (normalmente, un mensaje generado por el componente de una aplicación y a la espera de ser gestionado por otro). Los mensajes de Amazon SQS se pueden enviar y recibir a través de servidores o de componentes de aplicaciones distribuidos dentro del entorno de Amazon EC2 o en todas las ubicaciones de Internet. Amazon SQS admite un número ilimitado de colas y admite al menos una entrega de mensajes desordenada.

Aunque Amazon SQS y otros servicios de cola de mensajes se consideran normalmente como protocolos de comunicación asíncrona, Amazon SQS también puede verse como una solución de almacenamiento apta para ofrecer almacenamiento de datos temporal, pero duradero, para muchas clases de aplicaciones. El uso de Amazon SQS como almacenamiento temporal puede minimizar el uso de otros mecanismos de almacenamiento, como los archivos temporales del disco.

Escenario ideal de uso

Amazon SQS es ideal para cualquier escenario donde varios componentes de aplicaciones deben comunicar y coordinar su trabajo de manera imprecisa. Esto ocurre particularmente en los escenarios productor-consumidor en los que algunos componentes pueden funcionar con mayor rapidez o lentitud que otros, o donde una serie de componentes de

interacción cambian según el tiempo o la carga. Amazon SQS puede funcionar como el "pegamento de software" que permite que los componentes se comuniquen de forma fiable sin estar totalmente acoplados o sin que dependan tanto del funcionamiento sincrónico o de un número fijo de componentes.

Un uso clásico de Amazon SQS consiste en coordinar una línea base de procesamiento de varios pasos, en la que cada mensaje esté asociado con una tarea que deba procesarse. Un mensaje de Amazon SQS describe cada tarea indicando la tarea que ha de desarrollarse y un indicador de los datos de la tarea en Amazon S3.

A modo de ejemplo, suponga que tiene un número de archivos de imágenes para codificar. En una cola de trabajadores de SNS, puede crear un mensaje SNS para cada archivo que especifique el comando (jpeg-encode) y la ubicación del archivo en Amazon S3. Un grupo de instancias de Amazon EC2 que ejecute el software necesario de procesamiento de imágenes hace lo siguiente:

1. Sacar de la cola los mensajes de la tarea de forma asíncrona.
2. Recuperar el archivo guardado.
3. Procesar la conversión.
4. Volver a grabar la imagen en Amazon S3.
5. Escribir un mensaje del tipo "tarea completa" en otra cola.
6. Eliminar el mensaje de la tarea original.
7. Comprobar si hay más mensajes en la cola del trabajador.

El uso de la cola de Amazon SQS permite aumentar o reducir el número de instancias de trabajadores, así como aumentar o disminuir la potencia de procesamiento cada instancia de trabajador para adaptar la carga de trabajo total, sin realizar ningún cambio de la aplicación.

Rendimiento

Amazon SQS es un sistema de cola distribuido optimizado para la escalabilidad horizontal, no para ofrecer velocidades de envío o recepción de un único subproceso. Un único cliente puede enviar o recibir mensajes de Amazon SQS a una velocidad de 5 a 50 mensajes por segundo. Se puede conseguir un rendimiento de recepción superior mediante la solicitud de varios mensajes (hasta 10) en una única llamada. Pueden transcurrir varios segundos antes de que el mensaje que se ha puesto en cola esté disponible para su recepción.

Durabilidad y disponibilidad

Por cuestiones de diseño, los mensajes de Amazon SQS son muy duraderos pero temporales. Para evitar que los mensajes se pierdan o dejen de estar disponibles, todos los mensajes se almacenan en varios servidores y centros de datos. El tiempo de retención del mensaje se configura según la cola, desde un mínimo de una hora hasta un máximo de 14 días. Los mensajes se mantienen en una cola hasta que se eliminan de forma explícita o hasta que se eliminan automáticamente cuando vence el tiempo de retención.

Coste

Al igual que ocurre con todos los servicios de Amazon Web Services, con Amazon SQS solo ha de pagar por lo que use, sin necesidad de pagar tarifas mínimas ni de asumir contratos a largo plazo. A fin de empezar a utilizar y respaldar aplicaciones sencillas, Amazon SQS ofrece una capa gratuita del servicio que proporciona 100.000 solicitudes al mes sin cargo alguno. Más allá de la capa gratuita, los precios de Amazon SQS se basan en una serie de solicitudes (precios por cada 10.000 solicitudes) y la cantidad de datos transferidos dentro y fuera (precios por GB al mes). Puede encontrar información acerca de los precios totales de Amazon SQS en <http://aws.amazon.com/sqs/pricing/>.

Elasticidad y escalabilidad

Amazon SQS ofrece alta elasticidad y escalabilidad. Se trata de una solución diseñada para permitir que un número ilimitado de sistemas puedan leer y grabar un número ilimitado de mensajes en cualquier momento. Admite un número ilimitado de colas y un número ilimitado de mensajes por cola para cada usuario.

Interfaces

Se puede acceder a Amazon SQS a través de las interfaces SOAP y de consulta (HTTP). Existen cinco API que ayudan a los desarrolladores a iniciarse en SQS: CreateQueue, SendMessage, ReceiveMessage, ChangeMessageVisibility y DeleteMessage. Las API adicionales ofrecen una funcionalidad avanzada. En todos los casos, las API de consulta y SOAP se pueden utilizar con Java, C#, Perl y PHP.

Antipatrones de Amazon SQS

- **Datos binarios o grandes:** los mensajes de Amazon SQS deben ser de texto y han de tener una longitud máxima de 64 KB. Si los datos que necesita almacenar en una cola superan esta longitud, o son binarios, es mejor que utilice Amazon S3 o RDS para almacenar datos grandes o binarios, y que almacene un indicador para los datos en Amazon SQS.
- **Almacenamiento a largo plazo:** si los datos de los mensajes han de almacenarse durante más de 14 días, resulta más apropiado utilizar Amazon S3 o algún otro mecanismo de almacenamiento.

Amazon SimpleDB

Amazon SimpleDB representa un enfoque innovador con respecto al almacenamiento y la gestión de datos estructurados en la nube, que difiere de la técnica tradicional que consiste en la implementación de servidores de bases de datos relacionales. SimpleDB es un almacén de datos no relacionales de alta disponibilidad, escalabilidad y flexibilidad que libera gran parte del trabajo que supone la administración de la base de datos y la gestión de sistemas asociados. Con Amazon SimpleDB, dispone de un modelo de datos sin esquema, y puede almacenar elementos de datos comprimidos de un número flexible de pares de nombre/valor.

Escenario ideal de uso

Esta alternativa de almacenamiento es apta para situaciones en las que disponga de datos específicos y estructurados que necesite conservar y posteriormente consultar (interacciones orientadas a lectura principalmente), con alta disponibilidad y durabilidad, pero para lo que no quiere o no necesita incurrir en un exceso de administración de una base de datos relacional completa.

Muchas organizaciones utilizan SimpleDB como parte de una arquitectura de almacenamiento más grande basada en la nube. Un patrón común consiste en utilizar SimpleDB para realizar un seguimiento de los metadatos relativos a la información que está almacenada en otras soluciones de AWS. Por ejemplo, suponga que está compilando una solución basada en la red que almacenará millones de imágenes multimegabyte. En esta situación, puede utilizar Amazon Elastic Block Storage (EBS) o Amazon Simple Storage Service (Amazon S3) para almacenar las imágenes, de forma que SimpleDB sirva como repositorio para los detalles relacionados con cada imagen. Puede realizar consultas a SimpleDB para localizar una imagen concreta y utilizar esta información para recuperar los resultados obtenidos de otras soluciones de almacenamiento de Amazon.

Rendimiento

Gracias a las prestaciones que ofrece la infraestructura probada de alta velocidad de Amazon, el enfoque de SimpleDB crea una base de datos rápida y de alta disponibilidad que además ofrece una gran escalabilidad. Para conseguir un rendimiento superior y la mínima latencia, Amazon SimpleDB ofrece de forma predeterminada lecturas eventualmente coherentes. Para las aplicaciones que necesitan "leer lo último que se ha grabado", también se ofrece una lectura coherente.

Durabilidad y disponibilidad

SimpleDB ofrece una durabilidad muy alta de los datos a través de una replicación georredundante totalmente automática. Varias réplicas de cada elemento de datos de SimpleDB se encuentran almacenadas en diferentes ubicaciones dentro de la región geográfica seleccionada.

Coste

Al igual que ocurre con todos los servicios de Amazon Web Services, con Amazon SimpleDB solo ha de pagar por lo que use, sin necesidad de pagar tarifas mínimas ni de asumir contratos a largo plazo. Los costes de SimpleDB incluyen el almacenamiento de datos estructurados (por GB al mes), transferencia de datos (por GB al mes) y horas de máquina (al mes) en relación con operaciones PUT y GET. El primer GB mensual de almacenamiento de datos estructurados y las primeras 25 horas de máquina mensuales (aproximadamente 2 M de solicitudes GET o de solicitudes de la API SELECT) son gratuitos. Mediante la utilización de esta capa gratuita, muchas aplicaciones no incurrirán nunca en ningún cargo derivado de la utilización de SimpleDB.

Puede encontrar toda la información acerca de los precios de Amazon SimpleDB en <http://aws.amazon.com/simpliedb/pricing/>.

Elasticidad y escalabilidad

Al crear un dominio de SimpleDB, AWS asigna un máximo de 10 GB de almacenamiento por dominio. Habida cuenta de que paga solo por los recursos informáticos y de almacenamiento a medida que los utiliza, puede reducirlo de forma transparente hasta cero o aumentarlo. Amazon SimpleDB gestiona los requisitos de almacenamiento e indización. Esta solución almacena los datos indizados automáticamente de forma redundante en varias ubicaciones dentro de la región que haya seleccionado al crear el dominio de SimpleDB. Esto elimina la carga que supone la administración de la base de datos y le permite simplificar los datos PUT y GET. En el caso de que se trate de escalabilidad horizontal para aumentar el rendimiento o el tamaño del conjunto de datos más allá de los límites de un único dominio, puede crear varios dominios de SimpleDB y ejecutar varias operaciones de forma paralela. Para conjuntos de datos cuyo tamaño sea superior a 10 GB, debe considerar la posibilidad de utilizar la opción de particionar los datos entre varios dominios de SimpleDB.

Interfaces

Dispone de una serie de técnicas diferentes para realizar una interacción de programación con SimpleDB. Amazon ofrece una API de servicio web, disponible para los estilos SOAP y RESTful. La API ofrece actualmente nueve operaciones, que permiten gestionar el dominio (creación, enumeración, eliminación y obtención de metadatos) y trabajar con atributos (obtener, grabar de forma individual o por lotes y eliminar). Aunque el SQL estándar no se encuentra disponible para SimpleDB, puede utilizar la operación SimpleDB Select para crear consultas tipo SQL que permiten recuperar un conjunto de atributos basados en los criterios que usted establezca.

Si prefiere utilizar un conjunto de desarrollo de software, se encuentran disponibles conjuntos de herramientas y bibliotecas para lenguajes y plataformas de desarrollo de software populares, como:

- **Java**, usando el AWS SDK para Java (<http://aws.amazon.com/sdkforjava/>) y AWS Toolkit for Eclipse (<http://aws.amazon.com/eclipse/>).
- **C#**, usando el AWS SDK para .NET (<http://aws.amazon.com/sdkfornet/>) y las plantillas de Visual Studio
- **PHP**, usando la instalación básica de PHP5 o el AWS SDK para PHP (<http://aws.amazon.com/sdkforphp/>).
- **Perl**, usando los módulos Digest::SHA1, Bundle::LWP y XML::Simple, que se pueden descargar del sitio web de Comprehensive Perl Archive Network, en <http://www.cpan.org>.

Antipatrones de Amazon SimpleDB

Es fácil determinar qué escenarios de uso no son adecuados para SimpleDB. Si se reúne algunos de los siguientes criterios, considere la opción de utilizar una de las demás opciones de almacenamiento de AWS:

- **Aplicación grabada previamente y enlazada a una base de datos relacional de carácter tradicional:** si trata de enlazar una aplicación existente a la nube de AWS y necesita seguir utilizando una base de datos relacional, puede optar por utilizar Amazon RDS (si su base de datos es MySQL) o una de las distintas AMI de base de datos de Amazon EC2 configuradas previamente. También puede crear su propia instancia de Amazon EC2 e instalar un motor de base de datos en ella.

- **Uniones y transacciones complejas:** aunque muchas soluciones pueden beneficiarse de SimpleDB para prestar asistencia a sus usuarios, es posible que su aplicación precise de uniones, transacciones complejas, así como de otra infraestructura relacional que ofrezcan las plataformas de bases de datos tradicionales. En este caso, es posible que desee explorar Amazon RDS o Amazon EC2 con una base de datos instalada.
- **Datos de BLOB (objetos binarios grandes):** si pretende almacenar objetos binarios (por ejemplo, vídeo, imágenes o música), es posible que deba tener en cuenta la posibilidad de utilizar Amazon Elastic Block Storage (EBS) o Amazon Simple Storage Service (Amazon S3). Sin embargo, SimpleDB tiene una función que desempeñar en este escenario para realizar un seguimiento de los metadatos (por ejemplo, nombre del elemento, tamaño, fecha de creación, propietario, ubicación, etc.) sobre los objetos binarios.
- **Datos escritos (numéricos):** SimpleDB almacena todos los datos como cadenas de texto, por lo que, si necesita manipular datos escritos o numéricos, RDS o una base de datos tradicional de Amazon EC2 puede ser una solución más apta para ello.
- **Grandes cantidades de datos:** como se ha mencionado anteriormente, SimpleDB ofrece un modelo atractivo de precios y almacenamiento para muchas aplicaciones. Sin embargo, si tiene que administrar varios conjuntos de datos grandes, también debe considerar alternativas de almacenamiento como Amazon S3 o EBS. Una vez más, puede utilizar SimpleDB como un mecanismo eficaz para realizar un seguimiento de los datos relativos a los conjuntos de datos grandes.

Amazon Relational Database Service (Amazon RDS)

Amazon Relational Database Service (RDS) es una base de datos relacional MySQL totalmente funcional que se ofrece como un servicio gestionado basado en la nube. Si su aplicación necesita almacenamiento relacional, pero quiere reducir el tiempo que emplea en tareas de gestión de la base de datos, Amazon RDS automatiza las tareas administrativas más comunes para reducir la complejidad y el coste total de la propiedad. Amazon RDS realiza automáticamente una copia de seguridad de su base de datos y mantiene el software de la misma, lo que le permite emplear más tiempo en el desarrollo de la aplicación. Resulta una solución óptima para datos estructurados tradicionales que requieren más consultas sofisticadas y capacidades de unión en comparación con lo que ofrecen otras soluciones de base de datos de Amazon, como SimpleDB.

Escenario ideal de uso

Amazon RDS resulta una opción ideal para cualquier aplicación que utilice MySQL como su repositorio de información y en el caso de que desee beneficiarse de una base de datos de alta escalabilidad, bajo mantenimiento, rentabilidad y basada en la nube sin la necesidad de cambiar ningún código.

Rendimiento

Amazon RDS ofrece un alto rendimiento a través de una combinación de instancias configurables que se ejecutan en una infraestructura de primera clase probada de Amazon con operaciones de mantenimiento y copia de seguridad totalmente automáticas. Las configuraciones de bases de datos disponibles abarcan desde una instancia pequeña (plataforma de 64 bits con 1,7 GB de RAM y 1 unidad informática elástica (ECU)) hasta una instancia cuádruple extragrande (plataforma de 64 bits con 68 GB de RAM y 26 ECU).

Para conseguir un rendimiento óptimo, los diseñadores y administradores de bases de datos han de seleccionar el perfil de instancia adecuado (incluidos la RAM y el almacenamiento) según las necesidades informáticas concretas. Si necesita realizar ajustes, pueden utilizar las API de Amazon para ajustar la configuración de la base de datos. Amazon CloudWatch ofrece métricas que pueden ayudar a determinar cuándo es necesario ajustar el rendimiento o cuándo se necesita más elasticidad. Por último, los administradores pueden optar por cambiar a una configuración de base de datos más grande en caso de que la demanda supere los recursos de la instancia actual.

Elasticidad y escalabilidad

Los recursos de RDS se pueden escalar de forma elástica en varias dimensiones: tamaño de almacenamiento de la base de datos, capacidad informática de la instancia de base de datos y número de réplicas de lectura.

Para obtener mayor elasticidad informática, puede configurar más instancias de Amazon RDS y utilizar la partición para distribuir la carga de trabajo.

Para escalar el almacenamiento de base de datos de RDS de forma elástica, puede utilizar las herramientas de línea de comandos, las API o AWS Management Console para solicitar más capacidad de almacenamiento. En función de cuáles sean sus necesidades, puede añadir almacenamiento de forma inmediata o durante el próximo período de mantenimiento.

El aumento o la reducción de los recursos informáticos se puede realizar de forma sencilla con una única API o con el comando de AWS Management Console. Por ejemplo, es posible que necesite más potencia informática para crear facturas al final de cada mes. Resulta fácil ampliarla de forma temporal a una instancia extragrande cuádruple, ejecutar cargas de trabajo informático intensivo y posteriormente ejecutar una configuración más pequeña y rentable durante el resto del mes.

Para escalar los recursos de la base de datos de RDS dentro o fuera (escalado horizontal usando varias instancias de RDS), los administradores pueden crear una o varias réplicas de lectura de RDS. Las réplicas de lectura utilizan una capacidad de replicación asíncrona integrada de MySQL para permitir escalar de forma elástica los recursos de RDS para respaldar la lectura de grandes cargas de trabajo.

Por último, los administradores pueden configurar instancias adicionales de Amazon RDS y beneficiarse de la partición de la base de datos para distribuir la carga de trabajo, consiguiendo de esta forma incluso más elasticidad y escalabilidad de la base de datos.

Durabilidad y disponibilidad

Para conseguir mayor durabilidad, RDS ofrece dos tipos de copias de seguridad de base de datos que se replican por varias zonas de disponibilidad, a saber, las copias de seguridad automatizadas de las instancias de base de datos y las instantáneas de base de datos iniciadas por los usuarios. Si habilita las copias de seguridad automáticas de instancias de base de datos, RDS realizará automáticamente una copia de seguridad completa de los datos durante el período de copia de seguridad especificado y, además, capturará los registros de transacción de la base de datos. Estas copias de seguridad automáticas se ofrecen sin coste adicional, pueden conservarse hasta durante 8 días y se pueden utilizar para hacer un restablecimiento a un punto temporal a partir del inicio del período de retención hasta aproximadamente los últimos cinco minutos desde la hora actual. Las instantáneas de las bases de datos las inician los usuarios, se pueden crear en cualquier momento y se pueden conservar hasta que se eliminen de forma explícita. Las instantáneas de base de datos le permiten restablecer la base de datos a un estado conocido.

La función de implementación Multi-AZ de RDS mejora la durabilidad y disponibilidad de la base de datos mediante una replicación sincrónica de los datos entre una instancia de base de datos de RDS primaria y una instancia de espera en otra zona de disponibilidad. En el extraño caso de que se produzca un error de algún componente de la base de datos o de la zona de disponibilidad, RDS conmutará el error automáticamente a modo de espera (operación que suele tardar unos tres minutos) y las transacciones de la base de datos se reanudarán en cuanto se promueva el modo de espera. La replicación sincrónica garantiza que no se pierda ningún dato.

Tenga en cuenta que la replicación sincrónica que ofrece la función de implementación zonas de disponibilidad múltiples (Multi-AZ) de RDS es complementaria a la replicación asíncrona integrada que ofrecen las réplicas de lectura de RDS. Puede utilizar la función de forma aislada o una combinación de ambas funciones.

Coste

Al igual que ocurre con todos los servicios de Amazon Web Services, con Amazon RDS solo ha de pagar por lo que use, sin necesidad de pagar tarifas mínimas ni de asumir contratos a largo plazo. Amazon RDS ofrece una estructura de precios por capas, basada en el tamaño de la instancia de la base de datos, el tipo de implementación (Single-AZ/Multi-AZ) y la región de AWS. Los precios de Amazon RDS se fijan en función de una serie de factores: las horas de la instancia de la base de datos (por hora), la cantidad de almacenamiento ofrecido por la base de datos (por GB al mes y por millones de solicitudes de E/S), el almacenamiento adicional de copia de seguridad (por GB al mes) y la transferencia interna o externa de datos (por GB al mes). Puede encontrar toda la información acerca de los precios de RDS en <http://aws.amazon.com/rds/pricing/>.

Interfaces

Para ayudarle a iniciarse en sus servicios, Amazon ofrece una serie de guiones de línea de comandos fáciles de entender para crear, mantener, controlar y detener la instancia de la base de datos. De forma alternativa, puede utilizar operaciones basadas en servicios web de Amazon para realizar la misma serie de tareas.

Después de crear la instancia, puede configurar el esquema y los datos usando cualquier herramienta que sirva para una base de datos relacional. En función del volumen y la ubicación de los datos existentes localmente, en su lugar puede optar por extraer los datos locales usando la utilidad mysqldump y canalizarlos directamente hacia el ejecutable de MySQL para insertarlos en Amazon RDS. Para conjuntos de datos más grandes, puede resultar más beneficioso compilar primero su esquema en Amazon RDS, extraer los datos localmente en un archivo sin procesar y, a continuación, utilizar la utilidad mysqlimport para cargarlos en Amazon RDS. El tema de la migración a Amazon RDS se tratará de forma más detallada en un próximo documento técnico.

Después de haber colocado el esquema y los datos, puede interactuar con la información a través del SQL estándar de MySQL, así como con JDBC y otras API conocidas, además de con otras herramientas gráficas que puedan utilizarse con datos relacionales. Hay que hacer cambios de códigos para que la aplicación pueda interactuar con RDS: simplemente tiene que reemplazar la dirección del servidor de la base de datos (por ejemplo, dbserver.suempresa.com) con el punto final de DNS público (por ejemplo, miinstancia.c0cafngtpzd2.us-east-1.rds.amazonaws.com) que ofrece AWS cuando se crea la instancia. Este punto final de DNS seguirá siendo el mismo mientras dure la instancia. Aparte de configurar el punto final, todo lo relacionado con la aplicación basada en MySQL permanecerá invariable.

Antipatrones de Amazon RDS

Amazon RDS es una solución ideal para los datos de MySQL basados en la nube, pero no resulta la solución adecuada para una serie de escenarios:

- **Datos orientados a la indización y a la consulta:** muchas soluciones basadas en la nube no precisan de las funciones avanzadas de una base de datos relacional, como uniones y transacciones complejas. Si la aplicación está más orientada a la indización y a la consulta de datos, puede observar que Amazon SimpleDB puede resultar una opción más adecuada para satisfacer sus necesidades.
- **Numerosos objetos binarios grandes (BLOB):** aunque MySQL (y también Amazon RDS) puede almacenar objetos binarios grandes, si su aplicación hace un uso elevado de ellos (archivos de audio, vídeos, imágenes, etc.), es posible que Amazon Simple Storage Service (Amazon S3) sea una solución más apropiada.
- **Elasticidad automática:** como se ha mencionado anteriormente, es responsabilidad de los administradores configurar su entorno de Amazon RDS para conseguir elasticidad. Si necesita más elasticidad (y las estructuras de sus datos son adecuadas), puede elegir otra opción de almacenamiento como Amazon SimpleDB o Amazon S3.
- **Otras plataformas de base de datos:** en este punto, Amazon RDS ofrece una base de datos MySQL. Si necesita otra plataforma de base de datos (como IBM DB2 o Informix, Microsoft SQLServer, Oracle, PostgreSQL o Sybase), necesita implementar una AMI de base de datos relacional especializada de Amazon EC2 o crear una por su cuenta.

Base de datos relacional en Amazon EC2/AMI de base de datos relacional

Amazon EC2, junto con los volúmenes de EBS, ofrece una plataforma ideal para que pueda utilizar su propia base de datos relacional en la nube. Hay muchas soluciones de base de datos importantes que se encuentran disponibles como AMI de Amazon EC2 listas para utilizarse y previamente compiladas, entre otras, IBM DB2 e Informix, Oracle Database, MySQL, Microsoft SQL Server, PostgreSQL, Sybase, EnterpriseDB y Vertica.

Escenario ideal de uso

La ejecución de una base de datos relacional en Amazon EC2 y en EBS constituye el escenario ideal para los usuarios cuya aplicación requiere una base de datos relacional tradicional y específica, o bien para otros usuarios que requieran un nivel máximo de control administrativo y capacidad de configuración.

Rendimiento

El rendimiento de una instancia de base de datos relacional en Amazon EC2 depende de muchos factores, incluido el tipo de instancia de Amazon EC2, el número y la configuración de volúmenes de EBS, el software de base de datos y su configuración, y la carga de trabajo de la aplicación. En general, puede esperar que el rendimiento de la base de datos de Amazon EC2 sea similar al rendimiento de la misma base de datos instalada en un equipo in situ configurado de forma similar. Le instamos a que compare su aplicación real con varios tipos de instancias de Amazon EC2 usando varias configuraciones de almacenamiento a fin de seleccionar la configuración que mejor se adapte a sus necesidades.

Para aumentar el rendimiento de la base de datos, puede aumentar la memoria y los recursos informáticos mediante la elección de un tamaño superior de instancia de Amazon EC2. Para aumentar el rendimiento de E/S, puede cambiar el

número de volúmenes de EBS o utilizar el software RAID 0 (fragmentación del disco) en varios volúmenes de EBS, lo que aumentará la IOPS total y el ancho de banda. En muchos casos, también puede escalar el rendimiento total del sistema de la base de datos mediante el escalado horizontal con clústeres de base de datos, replicación y varios elementos subordinados de lectura. En general, el entorno de Amazon EC2 ofrece las mismas opciones de ajuste del rendimiento de la base de datos que un entorno de servidor físico.

Durabilidad y disponibilidad

Las bases de datos relacionales de Amazon EC2 ofrecen almacenamiento persistente para datos estructurados usando volúmenes de EBS como el almacén de datos, por lo que todas las notas sobre la durabilidad y disponibilidad de los datos de EBS también se aplican en este caso. Una vez más, también se puede aumentar la durabilidad y la disponibilidad básicas de los datos relacionales almacenados en volúmenes de EBS mediante la utilización de instantáneas de EBS, o bien mediante utilidades de copia de seguridad de base de datos de terceros (como RMAN de Oracle) para almacenar las copias de seguridad de base de datos en Amazon S3.

Coste

Mediante la ejecución de una base de datos en Amazon EC2, solo ha de pagar por lo que use, sin necesidad de tener que pagar tarifas mínimas ni de asumir contratos a largo plazo. El coste que supone la ejecución de su propia base de datos en Amazon EC2 depende del tamaño y el número de instancias de Amazon EC2 utilizadas para ejecutar la base de datos, del tamaño de los volúmenes de EBS utilizados para el almacenamiento de la base de datos, del volumen de datos transferidos dentro y fuera de Amazon EC2 y, en muchos casos, del coste de la licencia del software de base de datos de terceros. Muchos paquetes de base de datos de código abierto utilizan modelos de licencia gratuitos; algunos proveedores de software comercial utilizan el modelo Amazon DevPay; muchos otros ofrecen un modelo de licencia "bring-your-own-license". Póngase en contacto con su proveedor de software de base de datos o con Amazon Web Services para conocer el modelo aplicable para los precios de licencia.

Puede encontrar información detallada acerca de los precios establecidos para Amazon EC2, EBS y la transferencia de datos en el sitio web <http://aws.amazon.com/ec2/pricing/>.

Elasticidad y escalabilidad

En muchos casos, los usuarios de las soluciones de base de datos relacionales tradicionales de Amazon EC2 pueden beneficiarse de la elasticidad y la escalabilidad de la plataforma subyacente de AWS. Por ejemplo, después de configurar una instancia de Amazon EC2 con su solución de base de datos, puede agrupar la instancia en una AMI personalizada, utilizando los comandos Bundle para las AMI de almacén de instancia, o bien utilizando el comando Create Image para las AMI de EBS. Posteriormente, puede crear varias instancias nuevas de la configuración de la base de datos en unos instantes.

Base de datos relacional de antipatrones de Amazon EC2

La ejecución de su propia base de datos relacional en Amazon EC2 es una solución ideal para muchos usuarios, pero existen una serie de escenarios en los que otras soluciones pueden resultar más apropiadas:

- **Datos orientados a la indización y a la consulta:** muchas soluciones basadas en la nube no precisan de las funciones avanzadas de una base de datos relacional, como uniones o transacciones complejas. Si la aplicación está más orientada a la indización y a la consulta de datos, es posible que observe que Amazon SimpleDB puede resultar una opción más apropiada para satisfacer sus necesidades y, además, su gestión puede resultar más sencilla.

- **Numerosos objetos binarios grandes (BLOB):** muchas bases de datos relacionales admiten BLOB (archivos de audio, vídeo, imágenes, etc.). Si su aplicación hace un uso elevado de ellos, es posible que Amazon S3 sea una opción más apropiada, mediante la utilización de una base de datos relacional o de SimpleDB para la gestión de metadatos.
- **Elasticidad automática:** como se ha mencionado anteriormente, los usuarios de bases de datos relacionales en AWS, en muchos casos, se pueden beneficiar de la elasticidad y la escalabilidad de la plataforma subyacente de AWS, pero para ello es necesario que los administradores del sistema o los administradores de la base de datos realicen una tarea manual o programada. Suponiendo que precisa de una elasticidad totalmente automatizada (y que las estructuras de sus datos son adecuadas), puede elegir otra opción de almacenamiento como Amazon SimpleDB o Amazon S3.

Casos prácticos de almacenamiento en la nube

Para consultar las ilustraciones sobre el uso real de las opciones de almacenamiento de AWS, consulte el documento técnico adjunto *Storage Options in the AWS Cloud: Use Cases* en

http://media.amazonwebservices.com/AWS_Storage_Use_Cases.pdf.

Referencias y más documentación

Servicios de almacenamiento de AWS

Amazon Elastic Block Store (EBS), disponible en <http://aws.amazon.com/ebs>

Volúmenes de almacenamiento de instancias de Amazon EC2, disponible en

http://docs.amazonwebservices.com/AWSAmazon_EC2/latest/UserGuide

(consulte las secciones sobre tipos de instancias, almacenamiento de instancias y asignación de dispositivos de bloques)

Amazon Simple Storage Service (Amazon S3), disponible en <http://aws.amazon.com/s3>

Amazon Simple Queue Service (Amazon SQS), disponible en <http://aws.amazon.com/sqs>

Amazon SimpleDB, disponible en <http://aws.amazon.com/simpliedb>

Ejecución de bases de datos en AWS, disponible en http://aws.amazon.com/running_databases

Amazon Relational Database Service (Amazon RDS), disponible en <http://aws.amazon.com/rds>

Artículos sobre el almacenamiento de AWS

Artículo de Werner Vogel, titulado “Choosing Consistency”, disponible en

http://www.allthingsdistributed.com/2010/02/strong_consistency_simpledb.html

Artículo del blog Amazon Web Services Blog, “Amazon SimpleDB Consistency Enhancements”, disponible en

<http://aws.typepad.com/aws/2010/02/amazon-simpledb-consistency-enhancements.html>

Artículo del centro de desarrolladores de Amazon Web Services, “Amazon SimpleDB Consistency Enhancements”,

disponible en <http://developer.amazonwebservices.com/connect/entry.jspa?externalID=3572&categoryID=152>

Amazon Web Services Simple Monthly Calculator, disponible en <http://aws.amazon.com/calculator>

Oracle in the Cloud FAQ, disponible en <http://www.oracle.com/technetwork/topics/cloud/faq-098970.html>

Artículo del centro de desarrolladores de Amazon Web Services, titulado "Ejecución de bases de datos en AWS", disponible en http://aws.amazon.com/running_databases/

Oracle Case Study: Oracle Database in the Cloud, disponible en

<http://www.oracle.com/technetwork/database/features/availability/311356-129177.pdf>

Información adicional acerca de AWS

Capa de uso gratuito de AWS, disponible en <http://aws.amazon.com/free>

Conjuntos de datos públicos en AWS, disponible en <http://aws.amazon.com/publicdatasets>

AWS Import/Export, disponible en <http://aws.amazon.com/importexport>

Amazon CloudFront, disponible en <http://aws.amazon.com/cloudfront>

Casos prácticos de AWS, disponible en <http://aws.amazon.com/solutions/case-studies>

Anexo: Referencia rápida sobre el almacenamiento de AWS

	Datos no estructurados			Datos estructurados		
	Almacenamiento de instancias de Amazon EC2	Volúmenes de Amazon EBS	Amazon S3	Amazon SimpleDB	Otras base de datos relacionales (en EC2 y EBS)	Amazon RDS
Rendimiento	Alto	Alto	De moderado (un subproceso) a muy alto (varios subprocesos)	De moderado a alto (operaciones Put/Get por lotes)	Alto	Alto
Resistencia	Bajo	Moderado	Alto	Alto	Alto	Moderado
Coste	Incluido en el coste de EC2	Aprovisionado por GB/mes	Almacenado por GB/mes	Aprovisionado Primer GB gratuito; posteriormente por GB/mes	Aprovisionado (igual que en EBS)	Aprovisionado por GB/mes (5 GB mínimo)
Disponibilidad	Bajo	De moderado a alto (usando instantáneas de EBS)	Alto	Alto	De moderado a alto	Alto
Elasticidad/Escalabilidad	No	Manual (añadiendo más volúmenes)	Automático	Automático	Manual	Manual (un comando para modificar instancias de base de datos)
Límites de tamaño	De 160 GB a 1,6 TB (las instancias más grandes tienen volúmenes más grandes y más volúmenes)	De 1 GB a 1 TB por volumen(puede utilizar varios volúmenes o fragmentar para capacidades más grandes)	Realmente ilimitado (5 TB por objeto, objetos ilimitados por depósito)	10 GB/dominio 100 dominios (más dominios disponibles bajo demanda)	(igual que en EBS)	De 5 GB a 1 TB por instancia de base de datos
Persistencia por creación de instancias	No	Sí	Sí	Sí	Sí	Sí
Interfaces	Dispositivo de bloques, acceso a través del SO/sistema de archivos de EC2	N/D, acceso a través del SO/sistema de archivos de EC2	HTTP, REST o SOAP	REST o SOAP	Bibliotecas MySQL o JDBC	Bibliotecas MySQL o JDBC
Seguridad (cifrado en reposo)	Ejecución de sistema de archivos cifrados	Ejecución de sistema de archivos cifrados	Cifrar usando AES de 256 bits	Cifrar usando AES de 256 bits		
Seguridad (cifrado en tránsito)	N/D	N/D	SSL (HTTPS)	SSL (HTTPS)	SSL (HTTPS)	SSL (HTTPS)
Plataformas RDBMS admitidas	MySQL, SQL Server, Oracle, DB2, etc.	MySQL, SQL Server, Oracle, DB2, etc.	N/D	N/D	MySQL, SQL Server, Oracle, DB2, etc.	MySQL 5.1
Modelo (relacional o de otro tipo)	Bloque	Bloque	Objeto	Almacenamiento de entidades, esquema flexible, no relacional	Relacional	Relacional

	Datos no estructurados			Datos estructurados		
	Almacenamiento de instancias de Amazon EC2	Volúmenes de Amazon EBS	Amazon S3	Amazon SimpleDB	Otras base de datos relacionales (en EC2 y EBS)	Amazon RDS
Grado de automatización	Ninguno	Reflejo automático	Replicación automática, control de versiones	Indización, replicación, aprovisionamiento, revisiones	Depende de la base de datos	Copias de seguridad automáticas, revisiones de software
Grado de redundancia	No redundante	Redundante dentro de una zona de disponibilidad	Muy redundante en varios centros de datos	Mantener varias copias distintas geográficamente de todos los datos de usuario	Ninguno (replicación asíncrona disponible)	Ofrece las opciones de una única instancia de base de datos (una AZ) y de zonas de disponibilidad múltiples (Multi-AZ)
Acceso entre instancias (por ejemplo, capacidad de compartir)	No	No	Sí	Sí	Sí	Sí
Gestión y administración	Manual	Manual	Auto	Auto	Manual	Auto