



## **Opções de armazenamento em nuvem da AWS**

*Joseph Baron, Amazon Web Services*

*Robert Schneider, Think88*

*Dezembro de 2010*

## Introdução

A Amazon Web Services (AWS) é uma plataforma de computação em nuvem flexível, econômica e fácil de usar que inclui uma variedade de opções de armazenamento de dados baseados em nuvem. Essas alternativas fornecem uma ampla gama de opções para arquitetos e desenvolvedores.

Este whitepaper ajuda a entender as opções de armazenamento de dados primários disponíveis com a plataforma de computação em nuvem da AWS. Fornecemos uma visão geral de cada opção de armazenamento, descrevemos cenários de uso ideal e examinamos outras importantes características específicas de nuvem como elasticidade e escalabilidade. Dedicamos atenção especial para identificar o nível de durabilidade fornecida por cada opção de armazenamento. Este whitepaper encerra com uma tabela de referência rápida que compara as opções de armazenamento apresentadas aqui. Certifique-se de verificar a seção “Referências e Leitura complementar” para recursos adicionais.

Finalmente, apresentamos um [whitepaper complementar com vários casos de utilização de armazenamento](#) que mostram como usar várias opções de armazenamento em nuvem da AWS ao mesmo tempo. Você pode empregar esses casos de uso como um guia ao projetar sua própria arquitetura de armazenamento.

## Alternativa de armazenamento tradicional vs. armazenamento baseado em nuvem

Arquitetos de aplicativos e infraestrutura de TI tradicionais e no local tem inúmeras opções de armazenamento de dados, incluindo o seguinte:

- **Memória** – Em memória de armazenamento, como caches de arquivo, caches de objeto, bancos de dados na memória e discos de RAM fornecem acesso muito rápido aos dados.
- **Fila de mensagem** – Armazenamento temporário de dados enviados de forma assíncrona entre sistemas de computador ou componentes do aplicativo.
- **Rede de área de armazenamento (SAN)** – Bloquear dispositivos (LUNs de disco virtual) em SANs dedicados fornece muitas vezes o mais alto nível de desempenho de disco e durabilidade tanto para o armazenamento de dados quanto para o banco de dados de arquivos essenciais para os negócios, mas eles são certamente um dos sistemas mais caros.
- **Armazenamento de conexão direta (DAS)** – Unidades de disco rígido locais ou arrays que residem em cada servidor fornecem maior desempenho do que uma SAN, mas a durabilidade para arquivos temporários e persistentes, o armazenamento de banco de dados e o armazenamento de inicialização do sistema operacional (SO) são mais baixos do que em uma SAN.
- **Armazenamento conectado à rede (NAS)** – O armazenamento NAS fornece uma interface de nível de arquivo para armazenamento que pode ser compartilhado entre vários sistemas. NAS tende a ser mais lento do que SAN ou DAS.
- **Banco de dados** – Os dados estruturados geralmente são mantidos e acessados usando um banco de dados relacional como MySQL, PostgreSQL, Oracle, Microsoft SQL Server e DB2, ou em repositórios de banco de dados não relacionais. Os volumes de armazenamento de banco de dados geralmente residem em dispositivos SAN ou DAS.

- **Off-line** – Dados armazenados para fins de backup e arquivamento normalmente são colocados em mídia de disco tais como fitas, CDs, DVDs e assim por diante, que muitas vezes são armazenados em locais remotos e seguros para recuperação de desastres.

Cada uma dessas opções de armazenamento tradicional difere em desempenho, durabilidade e custo, bem como em suas interfaces. Arquitetos consideram todos estes fatores quando selecionam a solução de armazenamento adequada para a tarefa em mãos. Notavelmente, a maioria das infraestruturas de TI e arquiteturas de aplicativos utilizam várias tecnologias de armazenamento em conjunto, cada uma delas foi selecionada para satisfazer as necessidades de uma subclasse específica de armazenamento de dados. Estas combinações formam uma hierarquia de níveis de armazenamento de dados.

Como veremos ao longo deste whitepaper, a AWS oferece várias opções de armazenamento baseado em nuvem. Cada uma tem uma combinação única de desempenho, durabilidade, custo e interface e é reforçada por fatores adicionais tais como elasticidade, disponibilidade e escalabilidade. Estes fatores adicionais são críticos para as soluções de nuvem com base em escala da web. Como com aplicativos tradicionais no local, você pode usar várias opções de armazenamento em nuvem em conjunto para formar uma hierarquia de armazenamento de dados abrangente.

Este whitepaper analisa as seguintes opções de armazenamento em nuvem da AWS:

- **Volumes do Amazon EC2 Elastic Block Storage (EBS)**
- **Volumes do Amazon EC2 Local Instance Store (Ephemeral)**
- **Amazon Simple Storage Service (Amazon S3)**
- **Amazon Simple Queue Service (SQS)**
- **Amazon SimpleDB**
- **Amazon EC2 Relational Databases**
- **Amazon Relational Database Service (RDS)**

Vamos analisar detalhadamente cada uma das opções de armazenamento da AWS. Para cada opção, apresentaremos as seguintes informações:

- Nome e descrição
- Cenário de uso ideal
- Desempenho
- Durabilidade e disponibilidade
- Custos
- Elasticidade e escalabilidade
- Interfaces
- Antipadrões, situações onde outras opções de armazenamento seriam uma escolha melhor

Para as categorias adicionais de comparação entre a coleção de armazenamento da AWS, certifique-se de revisar a Referência rápida de armazenamento da AWS na página 36.

## Volumes do Amazon Elastic Block Store (EBS)

Os Volumes do Amazon Elastic Block Store (EBS) fornecem armazenamento durável de nível de bloco para uso com as instâncias do Amazon EC2 (máquinas virtuais). Os volumes do Amazon EBS são armazenamento fora de instância, conectados à rede que persiste independentemente da tempo de execução de uma única instância do Amazon EC2. Depois que um volume do EBS for anexado a uma instância do Amazon EC2, você pode interagir com ele como faria com uma unidade física de disco rígido, geralmente formatá-lo com um sistema de arquivo de sua escolha. Você pode usar um volume do EBS para iniciar uma instância do Amazon EC2 (AMIs EBS apenas) e anexar vários volumes do EBS a uma única instância do Amazon EC2. Observe, no entanto, que qualquer volume único do EBS pode ser ligado à apenas uma instância do Amazon EC2 a qualquer momento. Um volume do EBS não pode ser compartilhado com outros usuários, a menos que você crie um snapshot do EBS (consulte a seção “Durabilidade e disponibilidade” a seguir). Tamanhos para volumes do EBS variam de 1 GB a 1 TB e são alocados em incrementos de 1 GB.

### Cenário de uso ideal

O Amazon EBS destina-se a dados que mudam relativamente com frequência e requer persistência a longo prazo. O EBS fornece armazenamento em bloco virtual de modo persistente para servidores virtuais do Amazon EC2, assim você pode usá-lo como você usaria um disco rígido em um servidor físico. O Amazon EBS é particularmente bem adaptado para uso como o armazenamento principal para um sistema de arquivos, banco de dados ou para todos os aplicativos que necessitem de atualizações granulares finas e o acesso ao armazenamento de blocos de matérias-primas não formatados.

### Desempenho

Em geral, você pode esperar os volumes individuais do EBS para ter desempenho, tempo médio para falha (MTTF) e confiabilidade comparável a uma unidade USB alimentada externamente. Observe que enquanto volumes do EBS aparecem como unidades de disco locais, eles estão atualmente conectados à rede para uma instância do Amazon EC2. Portanto, outra rede E/S executada por instância, bem como a carga total na rede compartilhada, pode afetar o desempenho de volume de EBS individual.

Enquanto cada aplicativo (e seu desempenho associado) é exclusivo, você pode projetar e implantar muitas técnicas de otimização de taxa de transferência de disco tradicional com volumes do EBS. A combinação do Amazon EC2 e com o EBS permite usar várias das mesmas técnicas de otimização de desempenho que você usa com armazenamento e servidores locais. Por exemplo, você pode criar vários volumes e, em seguida, anexar todos a uma única instância do Amazon EC2. Com vários volumes do EBS anexados você pode particionar a carga de E/S total do aplicativo alocando um volume para os dados de registro, um volume para o banco de dados e ainda um outro volume para os dados de arquivo. Como alternativa, você pode classificar seus dados em múltiplos volumes do EBS usando um software RAID de driver de dispositivo 0, podendo assim agregar IOPs disponíveis, o rendimento volume total e o tamanho do volume total.

## Durabilidade e disponibilidade

Cada volume do Amazon EBS é automaticamente replicado na mesma zona de disponibilidade para evitar perda de dados devido a falha de qualquer componente de hardware único. O Amazon EBS também permite criar snapshots de volumes de um determinado momento, que persistem no Amazon S3. Esses Snapshots podem ser usados como ponto inicial para novos volumes Amazon EBS e para proteger dados para uma durabilidade a longo prazo.

A durabilidade do seu volume do EBS depende tanto do tamanho de seu volume quanto da porcentagem de dados que foi alterada desde seu último snapshot. Snapshots do EBS são backups incrementais e de determinado momento contendo apenas os blocos de dados alterados desde o último snapshot. Os volumes do EBS que operam com 20 GB ou menos de dados modificados desde que seu mais recente snapshot possa esperar uma taxa anual de falha (AFR) entre 0,1% - 0,5%. Para maximizar a disponibilidade e a durabilidade dos dados armazenados em volumes de EBS os usuários devem fazer snapshots de seus volumes EBS com frequência. Caso ocorra uma falha no seu volume do Amazon EBS, o que é uma situação improvável, todos os snapshots daquele volume permanecerão intactos e permitirão que você recrie seu volume a partir do momento do último snapshot.

Os volumes do Amazon EBS são projetados para serem altamente confiáveis e disponíveis. No entanto, em função de os volumes do EBS serem criados em uma determinada Zona de disponibilidade, eles estarão disponíveis se a própria Zona de disponibilidade não estiver disponível. Observe que, embora qualquer volume único do EBS seja restrito a uma única zona de disponibilidade, um snapshot de um volume do EBS está disponível em todas as Zonas de disponibilidade dentro de uma região e você pode usar um snapshot do EBS para criar um ou mais novos volumes do EBS em qualquer zona de disponibilidade. Os snapshots do EBS também podem ser compartilhados com outras contas de usuário. Isso facilita uso do backup de “clonagem de disco”, backup de “imagem de disco” e mecanismo de compartilhamento.

Para maximizar a disponibilidade e a durabilidade dos dados de EBS os usuários devem fazer snapshots de seus volumes EBS com frequência.

## Custos

Como em todos os serviços da Amazon Web Services, com o Amazon Elastic Block Store você paga somente pelo que usar, sem taxas mínimas ou contratos de longo prazo. O Amazon EBS é cobrado por GB ao mês de armazenamento mantido e por milhões de solicitações de E/S. O volume de armazenamento é cobrado pela quantidade que você aloca nele, até que você o libere. As snapshots do Amazon EBS são cobradas por GB ao mês de dados armazenados, bem como por 1.000 solicitações PUT e por 10.000 solicitações GET, quando salvar e carregar as snapshots. Pelas snapshots do EBS, você é cobrado somente pelo armazenamento que for realmente *utilizado (consumido)*. Observe que as snapshots do EBS são incrementais e comprimidas, de modo que o armazenamento utilizado em qualquer snapshot é geralmente muito menor do que o armazenamento utilizado em um volume EBS. Detalhes do preço total estão disponíveis em <http://aws.amazon.com/ec2/pricing/>.

Observe que não há cobrança por transferência de informação entre os vários armazenamentos oferecidos pela AWS (ou seja, uma instância do Amazon EC2 com EBS, Amazon S3, Amazon RDS, e assim por diante) contanto que estejam na mesma região da AWS.

## Elasticidade e escalabilidade

A AWS facilita a expansão de seu espaço de armazenamento disponível na sua instância do Amazon EC2. Você pode criar um novo volume do EBS e anexá-lo à instância e começar a utilizá-lo com as suas instâncias já existentes. Caso você não queira criar e manter novos volumes, veja como expandir o tamanho de um único volume:

- Confirme o aplicativo ou sistema de arquivos.
- Faça uma snapshot de seus dados do volume do EBS para o Amazon S3 (utilizando “Criar Snapshot do volume”).
- Crie um novo volume EBS a partir da snapshot, mas especifique um tamanho maior do que o do volume original.
- Anexe o volume novo e maior à sua instância do Amazon EC2.
- Separe e delete o volume do EBS original.

## Interfaces

Para criar, deletar, descrever, anexar e separar volumes do EBS da sua instância do Amazon EC2, o Amazon oferece APIs de controle em ambos formatos: SOAP e REST. Você pode utilizar as APIs para criar, deletar e descrever snapshots do EBS para o Amazon S3, bem como seus atributos associados. Se você prefere trabalhar com ferramentas gráficas, o AWS Management Console e a extensão do Firefox ElasticFox lhe proporcionará capacidade total da API em uma interface de fácil navegação. Independentemente de como você criar o seu volume do EBS, observe que todo armazenamento é distribuído no momento da criação do volume e que você é cobrado por este armazenamento, mesmo que você não faça uso do mesmo.

O EBS apresenta uma interface de dispositivo de bloco para a instância do Amazon EC2. Ou seja, para a instância do Amazon EC2, um volume do EBS apresenta-se como uma unidade de disco local. Para gravar e ler os dados dos volumes do EBS, você, portanto, utilizará as interfaces do sistema de arquivo de origem do seu sistema operacional escolhido.

## Antipadrões do Amazon EBS

Como descrito anteriormente, o EBS é ideal para informações que necessitam ser mantidas além da vida de uma única instância do Amazon EC2. No entanto, em certas situações outras opções de armazenamento da AWS podem ser mais apropriadas:

- **Armazenamento temporário** – Se você não está preocupado com fato de que seus dados permanecerão acessíveis após a sua instância do Amazon EC2 tenha sido encerrada (arquivos temporários, disco de trabalho, buffers, e assim por diante), considere aproveitar-se do volume de armazenamento que é fornecido automaticamente com a maioria das instâncias do Amazon EC2. Esses volumes efêmeros são fornecidos sem custo adicional além do preço padrão pela instância do Amazon EC2. Para mais informações, consulte a próxima sessão.
- **Armazenamento altamente durável** – Se você precisa armazenamento com alta durabilidade, utilize o Amazon S3. O armazenamento padrão do Amazon S3 é desenvolvido para um durabilidade anual de 99.999999999% por objeto. Por outro lado, os volumes do EBS com menos de 20 GB de dados modificados desde o último snapshot são projetados para um durabilidade anual entre 99,5% e 99,9%; pode-se esperar que volumes com mais dados modificados tenham durabilidade proporcionalmente menor

- **Dados estáticos ou conteúdo da web** – Se seus dados não são alterados com frequência, o Amazon S3 pode representar uma solução mais escalável e econômica para armazenamento dessas informações fixas. Além disso, o conteúdo da web oferecido por meio do EBS exige que um servidor da web seja executado no Amazon EC2, enquanto você pode disponibilizar conteúdo da web diretamente pelo Amazon S3.
- **Informações de valor essencial** – Como veremos na seção “Amazon SimpleDB” na página 14, o Amazon SimpleDB oferece um meio escalável e elástico para armazenagem de dados “schema-less”. Se você está tentando utilizar tecnologia relacional (incluindo volumes do EBS associados) para manter informações de valor essencial, vale a pena explorar o SimpleDB como uma alternativa.

## Volumes de armazenamento de instância local no Amazon Elastic Compute Cloud (EC2)

---

Os volumes de armazenamento de instância local do Amazon EC2 (também chamados drives efêmeros) oferecem armazenamento temporário de nível de bloco para instâncias do Amazon EC2. Ao criar uma instância do Amazon EC2 a partir de uma Amazon Machine Image (AMI), na maioria das vezes, ela é acompanhada de um bloco pré-configurado de armazenamento de disco pré-anexo.<sup>1</sup> Ao contrário dos volumes do EBS, os dados nos volumes de armazenamento de instância persistem apenas durante a vida da instância associada ao Amazon EC2. A quantidade de armazenamento deste disco varia de 160 GB até a 1,7 TB e varia conforme o tipo de instância do Amazon EC2. As maiores instâncias do Amazon EC2 tem mais e maiores volumes de armazenamento. No entanto, embora esses valores pareçam muito generosos e muitas vezes possam ser úteis, este armazenamento é temporário e melhor usado como um volume de trabalho ou disco RAM.

### Cenário de uso ideal

Volumes de armazenamento de instância local são ideais para o armazenamento temporário de informações que estão mudando continuamente, tais como buffers, caches, dados temporários e outros conteúdos temporários, ou para dados que são replicados em toda a frota de instâncias, tais como um pool de balanceamento de carga de servidores web. Armazenamento de instância do Amazon EC2 é desenvolvido para esta finalidade. Trata-se de dispositivo de inicialização (apenas para AMIs de armazenamento de instância) da máquina virtual, além de um ou mais volumes adicionais que são dedicados à instância do Amazon EC2 (para EBS AMIs e instância armazenam AMIs). Esse armazenamento é utilizável apenas para uma única instância do Amazon EC2 durante sua vida útil. Ao contrário dos volumes do EBS, os volumes de armazenamento de instância não podem ser desanexados ou anexados a outra instância.

### Desempenho

Pelo fato de a instância de virtual machine do Amazon EC2 e os volumes de armazenamento de instância local estarem localizados no mesmo servidor físico, a interação com esse armazenamento é muito rápida, particularmente para o acesso sequencial. Para aumentar as operações de E/S por segundo (IOPS) ou para melhorar a produtividade do disco, o armazenamento de várias instância de volumes pode ser agrupado junto usando software RAID 0 (distribuição de disco).

---

<sup>1</sup> A maioria dos tipos de instância do Amazon EC2 fornece volumes de armazenamento de instância local; no entanto, microinstâncias, tais como tipo t1.micro, fornecem somente o armazenamento do EBS. Além disso, as instâncias que usam o Amazon EBS para o dispositivo raiz (“inicialização através do EBS”), por padrão, não expõem os volumes de armazenamento de instância efêmera. Se desejar, você pode expor os volumes de armazenamento de instância no momento da inicialização de instância, especificando um Mapeamento de dispositivo de bloco. Consulte o Guia do usuário do Amazon EC2 para obter detalhes.



## Durabilidade e disponibilidade

Volumes de armazenamento de instância local do Amazon EC2 não se destinam a serem usados como armazenamento duradouro em disco. Os dados armazenados em volumes de armazenamento de instância local persistem apenas para o tempo de vida da instância do Amazon EC2. Os dados nos volumes de armazenamento de instância local Amazon EC2 são persistentes entre reinicializações de instância ordenadas, mas não em situações onde a instância do Amazon EC2 termina ou passa por um ciclo de falha ou reinicialização.

Você não deve usar volumes de armazenamento de instância local para quaisquer dados que devem persistir ao longo do tempo, tais como armazenamento permanente de arquivos ou banco de dados. No entanto, apesar dos volumes de armazenamento de instância única não serem persistentes, você pode manter seus dados periodicamente copiando ou fazendo o backup em EBS ou Amazon S3.

## Custos

O custo da instância do Amazon EC2 inclui todos os volumes e instâncias únicas (efêmeras). Embora não haja nenhuma cobrança para armazenamento de dados na instância local a fim de armazenar volumes, observe que os dados transferidos para e/ou a partir dos volumes de armazenamento Amazon EC2 em instância de fora do Amazon EC2 acarretará em taxas de transferência de dados e taxas adicionais no uso de qualquer armazenamento persistente, tais como Amazon S3, volumes EBS e snapshots do EBS. Informações detalhadas sobre o Amazon EC2, o EBS e os preços de transferência de dados podem ser encontradas na web em <http://aws.amazon.com/ec2/pricing/>.

## Elasticidade e escalabilidade

Volumes de armazenamento de instância local são fixos em tamanho para um dado tipo de instância do Amazon EC2 e vinculados a uma instância específica, de modo que este tipo de armazenamento é relativamente inelástico. No entanto, você pode obter plena elasticidade de armazenamento, incluindo uma das outras opções de armazenamento adequadas tais como o Amazon S3 ou o Elastic Block Storage (EBS) em sua estratégia de armazenamento no Amazon EC2.

## Interfaces

A instância do Amazon EC2 trabalha com volumes de armazenamento de instância local como se eles fossem unidades de disco locais. Isso significa que você pode interagir com dados hospedados em seu armazenamento de instância local usando mecanismos de sistema de arquivos nativo, tais como Windows NTFS ou Linux XFS. Observe que, em alguns casos, um dispositivo de volume de armazenamento de instância local será conectado à instância do Amazon EC2 mediante a inicialização, porém deve ser formatado com um sistema de arquivo apropriado e montado antes do uso.

## Antipadrões de Volume de Armazenamento de Instância Local no Amazon EC2

Os volumes de armazenamento de instância local do Amazon EC2 são rápidos, gratuitos (ou seja, incluídos no preço da instância do Amazon EC2) “volumes-rascunho” mais adequados para armazenar dados temporários que podem ser facilmente regenerados. Porém, em muitas outras situações, outras opções de armazenamento da AWS podem ser mais apropriadas:

- **Armazenamento persistente** – se você necessita de armazenamento persistente em disco virtual semelhante ao de uma unidade de disco física para arquivos ou outros dados que devam resistir mais do que o tempo de vida de uma única instância do Amazon EC2, os volumes EBS ou Amazon S3 são mais apropriados.



- **Armazenamento de banco de dados** – na maioria dos casos, bancos de dados requerem armazenamento que persistam mais do que o tempo de vida de uma única instância do Amazon EC2, tornando os volumes do EBS a escolha natural.
- **Armazenamento compartilhado** – volumes de armazenamento de instância são dedicados a uma única instância do Amazon EC2 e não podem ser compartilhados com outros sistemas ou usuários. Se você necessita de armazenamento que pode ser destacado de uma instância e anexado a uma instância diferente, ou se lhe for necessária a capacidade de compartilhar dados com facilidade, o Amazon S3 ou os volumes do EBS são a melhor escolha.
- **Snapshots** – se você precisar da conveniência, da durabilidade a longo prazo, da disponibilidade e também compartilhar snapshots de disco em determinado momento, os volumes do EBS são a melhor opção.

## Amazon Simple Storage Service (Amazon S3)

---

O Amazon S3 é um armazenamento de objetos distribuídos altamente flexível, durável e disponível projetado para armazenamento de dados primários e de missão crítica com uma interface de serviço web fácil de usar. Em aplicativos tradicionais locais, este tipo de dados normalmente seria armazenado em SAN ou NAS. No entanto, um mecanismo baseado em nuvem como o Amazon S3 é bem mais ágil, flexível e geo-redundante. O Amazon S3 fornece uma interface simples de serviço da web que pode ser usada para armazenar e recuperar qualquer quantidade de dados, a qualquer momento, de qualquer lugar na web. Você pode gravar, ler e deletar objetos contendo de 1 byte até 5 terabytes de dados cada, e o número de objetos que você pode armazenar em um bucket do Amazon S3 é ilimitado. O Amazon S3 também é altamente escalável, permitindo o acesso simultâneo à leitura e à gravação dos dados do Amazon S3 por diferentes clientes ou threads de aplicativo.

### Cenário de uso ideal

Um uso bastante comum para o Amazon S3 é o armazenamento de conteúdo estático da web. Este conteúdo pode ser enviado diretamente a partir do Amazon S3 por meio de um servidor web, visto que cada objeto do Amazon S3 possui seu próprio endereço URL HTTP, ou enviado por meio de uma rede de fornecimento de conteúdo (CDN), tal como o Amazon CloudFront. Graças à elasticidade do Amazon S3, ele funciona muito bem para hospedagem de conteúdo da web com exigências de largura de banda extremamente espinhosas. Além disso, como não há provisionamento de armazenamento, o Amazon S3 funciona bem para websites em rápido ritmo de crescimento de conteúdo de dados gerados pelo usuário, tais como compartilhamento de vídeos e fotos.

O Amazon S3 também é comumente usado como armazenamento de dados para computação em larga escala, tais como análise de transações financeiras ou dados de sequência de cliques e transcodificação de mídia. Por causa de escalabilidade horizontal do Amazon S3, você pode acessar seus dados a partir de múltiplos nós de computação simultaneamente sem ficar restrito por uma conexão única.

Por último, o Amazon S3 é muito usado como armazenamento de dados de missão crítica de origem, armazenamento altamente confiável para snapshot de volumes Amazon EBS, aplicações de armazenamento de backup e soluções “quentes” de recuperação de desastres para continuidade de negócios. Dado que o Amazon S3 armazena objetos de forma redundante em múltiplos dispositivos através de várias instalações, fornece a infraestrutura de armazenamento altamente durável necessária para essas situações.

## Desempenho

O acesso ao Amazon S3 de dentro do Elastic Compute Cloud da Amazon (Amazon EC2) na mesma região é rápido. Se você acessar o Amazon S3 usando várias linhas de execução, aplicativos ou clientes simultaneamente, a taxa de transferência agregada total do Amazon S3 será tipicamente escalada para taxas que ultrapassam largamente às que qualquer servidor único pode gerar ou consumir.

A fim de acelerar o acesso aos dados pertinentes, muitos desenvolvedores emparelham os dados do Amazon S3 com os do Amazon SimpleDB. O Amazon S3 armazena as informações reais e o SimpleDB serve como repositório dos metadados associados (por exemplo, nome do objeto, tamanho, palavras-chave e assim por diante). O SimpleDB fornece indexação automática, tornando bastante eficaz a localização da referência de um objeto através de uma pesquisa de metadados. Este resultado pode então ser usado para identificar e, em seguida, ser recuperado de dentro do próprio Amazon S3.

## Durabilidade e disponibilidade

Armazenando em sincronia e automaticamente seus dados através de vários dispositivos e várias instalações dentro de sua região geográfica selecionada, o armazenamento do Amazon S3 fornece o mais alto nível de durabilidade de dados e disponibilidade dentro da plataforma AWS. A correção de erros é embutida, e não há um ponto sequer de falha. O Amazon S3 foi projetado para sustentar a perda simultânea de dados em duas instalações, tornando-o muito bem adaptado para servir como o dispositivo de armazenamento de dados primários para dados de missão crítica. De fato, o Amazon S3 foi projetado para durabilidade de 99.999999999% (“11 noves”) por objeto e 99,99% de disponibilidade ao longo do período de um ano. Além de sua redundância embutida, os dados contidos no Amazon S3 também podem ser protegidos de falhas de aplicativo e exclusões não intencionais por meio do controle de versões do Amazon S3. Você também pode habilitar o controle de versões do Amazon S3 com o MFA Delete. Com esta opção habilitada em um bucket, duas formas de autenticação são requeridas para excluir uma versão de um objeto do Amazon S3: credenciais válidas de uma conta AWS mais um código de seis dígitos (uma senha de uso único, baseada em tempo) originados em dispositivo de token físico.

Para dados não críticos que podem ser reproduzidos facilmente se necessário, tais como mídia transcodificada ou miniaturas de imagens, você pode usar a opção de armazenamento de redundância reduzida (RRS) no Amazon S3, que fornece um nível inferior de durabilidade a um menor custo de armazenamento. Objetos armazenados usando a opção RRS têm menos redundância do que objetos armazenados em um armazenamento padrão do Amazon S3. Em ambos casos, seus dados estarão armazenados em múltiplos dispositivos e múltiplas localizações. O RRS é projetado para fornecer uma durabilidade de 99,99% de objetos durante um ano determinado. Enquanto que o RRS é menos durável que o Amazon S3 padrão, ainda assim é 400 vezes mais durável do que uma unidade de disco típica.

## Custos

Assim como em todos os serviços web da Amazon, com o Amazon S3, você paga apenas pelo que usar, sem taxas mínimas ou contratos de longo prazo. O Amazon S3 tem três componentes de fixação de preços: armazenamento (por GB por mês), transferência de dados download ou upload (por GB por mês) e solicitações (por n mil solicitações por mês). Para novos clientes, a AWS fornece um pacote de uso livre que inclui até 5 GB de armazenamento do Amazon S3. A tabela de preços completa do Amazon S3 pode ser encontrada em <http://aws.amazon.com/s3/pricing/>.

## Elasticidade e escalabilidade

O Amazon S3 foi projetado para oferecer um nível muito elevado de elasticidade e escalabilidade automaticamente. Ao contrário de um sistema de arquivos típico, que costuma encontrar problemas ao armazenar grande número de arquivos em um diretório, o Amazon S3 oferece suporte a um número ilimitado de arquivos em qualquer bucket. Além disso, ao contrário de uma unidade de disco, que impõe um limite na quantidade total de dados que podem ser armazenados antes de você particionar os dados em unidades e/ou servidores, um bucket do Amazon S3 pode armazenar um número ilimitado de bytes. Você pode armazenar qualquer número de objetos e o Amazon S3 gerenciará o dimensionamento e a distribuição de cópias redundantes de suas informações em outros servidores e outros locais na mesma região, todos usando a infraestrutura de alto desempenho da Amazon.

## Interfaces

O Amazon S3 fornece ambas APIs SOAP e RESTful em serviços web. Essas APIs permitem que os objetos (arquivos) do Amazon S3 sejam armazenados em buckets exclusivamente nomeados (pastas de nível superior). Cada objeto deve ter uma chave de objeto exclusivo (nome de arquivo) que servirá como um identificador para o objeto contido nesse bucket. Sendo o Amazon S3 um sistema de armazenamento de objetos baseado na web, em vez de um sistema de arquivos tradicional, você pode facilmente emular uma hierarquia de sistema de arquivos (pasta1/pasta2/arquivo) no Amazon S3, criando nomes-chaves de objeto que correspondam ao caminho completo de cada arquivo.

A maioria dos desenvolvedores cria aplicativos no Amazon S3 usando um kit de ferramentas de alto nível. A AWS e terceiros criaram kits de ferramentas de interface para o Amazon S3 e bibliotecas para várias linguagens de desenvolvimento de software popular e plataformas, tais como:

- **Java**, usando o SDK da AWS para Java (<http://aws.amazon.com/sdkforjava/>) e o kit de ferramentas AWS para Eclipse (<http://aws.amazon.com/eclipse/>).
- **C#**, usando o SDK da AWS para .NET (<http://aws.amazon.com/sdkfornet/>) e os Modelos do Visual Studio.
- **PHP**, usando a instalação de base do PHP5, ou o SDK da AWS para PHP (<http://aws.amazon.com/sdkforphp/>).
- **Perl**, usando o Digest::SHA1, Bundle::LWP, and XML::Simple modules, os quais podem ser baixados do Comprehensive Perl Archive Network em <http://www.cpan.org>.

Visto que o Amazon S3 foi projetado para ser acessado principalmente por meio de APIs, o AWS Management Console e ferramentas de terceiros também fornecem interfaces gráficas e de linha de comando para interagir com seus buckets e objetos do Amazon S3. O Amazon S3 também fornece suporte ao protocolo BitTorrent, o que permite que os consumidores recuperem simultaneamente as informações do S3, bem como de outros provedores.

Para baixar ou fazer upload de grandes volumes de dados você também pode usar o serviço AWS Import/Export (<http://aws.amazon.com/importexport/>). O AWS Import/Export acelera a movimentação de grandes volumes de dados para dentro e para fora da AWS usando dispositivos de armazenamento portáteis para transporte. A Amazon montará seu dispositivo e copiará os dados para ou a partir de um bucket Amazon S3 designado. Esta abordagem é frequentemente mais rápida e mais em conta do que a transferência de grandes quantidades de dados através da Internet.

Observe que o RRS usa as mesmas interfaces que as interfaces padrão do Amazon S3; a opção RRS é específica para um objeto ou bucket do Amazon S3 configurando a propriedade de classe de armazenamento durante uma operação PUT, ou selecionando a configuração de “Utilização de armazenamento de redundância reduzida” no AWS Management Console.

## Antipadrões do Amazon S3

O Amazon S3 é ideal para armazenar várias classes de informações relativamente estáticas e beneficiar-se de sua durabilidade, disponibilidade e recursos de elasticidade. No entanto, em algumas situações, o Amazon S3 não seria a melhor solução:

- **Sistema de arquivos** – O Amazon S3 usa um espaço de nomes plano e não se destina a servir como um sistema de arquivos autônomo e compatível com POSIX. No entanto, usando delimitadores (comumente os caracteres '/' ou '\') você é capaz de construir chaves para emular a estrutura de pastas hierárquica do sistema de arquivos dentro de um dado bucket.
- **Dados estruturados com consulta** – O Amazon S3 não oferece recursos de consulta: para recuperar um objeto específico, você precisará saber o nome do bucket e a chave. Logo, você não pode usar o Amazon S3 como um banco de dados por si só. Um padrão de utilização comum é colocar um objeto no Amazon S3 e então, utilizar o Amazon SimpleDB para manter os metadados do objeto. Você também pode usar outras tecnologias de banco de dados, tais como o Amazon RDS ou um banco de dados executando em uma instância do Amazon EC2. Você pode então pesquisar no banco de dados para localizar no bucket o nome do objeto e a respectiva chave e, em seguida, recuperar o próprio objeto do bucket do Amazon S3.
- **Dados que mudam rapidamente** – Dados que devem ser atualizados com muita frequência podem ser melhor servidos por uma solução de armazenamento com menor latência de leitura/gravação, tais como o EBS ou um banco de dados.

## Amazon Simple Queue Service

O Amazon Simple Queue Service (Amazon SQS) fornece um serviço de enfileiramento de mensagens confiável, altamente escalável, hospedado para armazenamento temporário e entrega de mensagens curtas de dados baseadas em texto (até 64 kB). Uma fila do Amazon SQS é um repositório de dados temporário para mensagens que estão aguardando processamento (normalmente uma mensagem produzida pelo componente de um aplicativo e esperando para ser consumido por outro). As mensagens do Amazon SQS podem ser enviadas e recebidas por servidores ou componentes de aplicativos distribuídos dentro do ambiente do Amazon EC2 ou em qualquer lugar da Internet. O Amazon SQS suporta um número ilimitado de filas e filas desordenadas, e no mínimo uma única entrega de mensagens.

Enquanto que o Amazon SQS e outros serviços de enfileiramento de mensagens são normalmente considerados como protocolos de comunicação assíncrona, o Amazon SQS também pode ser visto como uma loja que provém armazenamento de dados temporário porém durável para várias classes de aplicativos. A utilização do Amazon SQS como armazenamento temporário pode minimizar o uso de outros mecanismos de armazenamento, tais como arquivos temporários de disco.

### Cenário de uso ideal

O Amazon SQS é mais adequado para qualquer cenário onde múltiplos componentes do aplicativo devem comunicar e coordenar os seus trabalhos de maneira mais flexível. Isso ocorre particularmente em cenários de produtor-consumidor onde alguns componentes podem funcionar mais rápido ou mais devagar do que outros, ou quando o número de componentes interagentes muda com o passar do tempo ou de acordo com a carga. O Amazon SQS pode servir como a linguagem cola que habilite componentes a se comunicarem com fiabilidade sem estarem apertadamente acoplados ou altamente dependentes durante a operação síncrona, ou em um número fixo de componentes.

Um uso clássico do Amazon SQS é coordenar um pipeline de processamento de multi-etapas, onde cada mensagem está associada a uma tarefa que deve ser processada. Cada tarefa fica descrita por uma mensagem do Amazon SQS indicando a tarefa a ser realizada e um ponteiro mostrando os dados da tarefa no Amazon S3.

Para ilustrar, suponhamos que você tenha um número x de arquivos de imagem para codificar. Em uma fila de trabalho do SNS, você cria uma mensagem de SNS para cada arquivo especificando o comando (codificação jpeg) e a localização do arquivo no Amazon S3. Um pool de instâncias do Amazon EC2 executando a imagem necessária de processamento do software em questão faz o seguinte:

1. Retira as mensagens da tarefa em questão da fila, assincronamente
2. Recupera o arquivo em questão
3. Processa a sua conversão
4. Grava a imagem de volta para o Amazon S3
5. Escreve uma mensagem “tarefa concluída” para outra fila
6. Exclui a mensagem original da tarefa
7. Verifica se há mais mensagens na fila de trabalho

O uso da fila do Amazon SQS permite expandir ou reduzir o número de instâncias de trabalho e também permite expandir ou reduzir o poder de processamento de cada instância de trabalho única, de acordo com a carga de trabalho total, sem alterar o aplicativo.

## Desempenho

O Amazon SQS é um sistema distribuído de enfileiramento de mensagens otimizado para escalabilidade horizontal e não para velocidades de envio ou recebimento de segmento único. Um único cliente pode enviar ou receber mensagens do Amazon SQS a uma taxa de cerca de 5 a 50 mensagens por segundo. Maior desempenho de recebimento pode ser atingido solicitando várias mensagens (até 10) em uma única chamada. Pode levar alguns segundos para que uma mensagem que tenha sido colocada na fila fique disponível para ser recebida.

## Durabilidade e disponibilidade

Por design, as mensagens do Amazon SQS são altamente duráveis porém temporárias. Para evitar que mensagens sejam perdidas ou fiquem indisponíveis, todas as mensagens são armazenadas redundantemente entre vários servidores e Datacenters. O tempo de retenção da mensagem é configurável tomando por base a fila, com um mínimo de uma hora para um máximo de 14 dias. As mensagens são mantidas em fila até que sejam explicitamente excluídas, ou até que sejam excluídas automaticamente após a expiração do tempo de retenção.

## Custos

Assim como todos os serviços web da Amazon, com o Amazon SQS, você paga apenas pelo que usar, sem taxas mínimas ou contratos de longo prazo. A fim de começar o trabalho e fornecer suporte a aplicativos simples, o Amazon SQS fornece um pacote de serviços que fornece 100.000 solicitações por mês sem nenhum custo adicional. Além do pacote gratuito, os preços do Amazon SQS baseiam-se no número de solicitações (taxados a cada 10.000 solicitações) e na quantidade de dados transferidos para fora e para dentro (taxados por GB por mês). A tabela de preços completa do Amazon SQS pode ser encontrada em <http://aws.amazon.com/s/pricing/>.

## Elasticidade e escalabilidade

O Amazon SQS é altamente elástico e maciçamente escalável. O Amazon SQS foi projetado para permitir que um número ilimitado de computadores leiam e gravem um número ilimitado de mensagens a qualquer momento. Oferece suporte a um número ilimitado de filas e um número ilimitado de mensagens por fila para qualquer usuário.

## Interfaces

O Amazon SQS pode ser acessado através de interfaces de consulta (HTTP) e SOAP. Cinco APIs facilitam a familiarização dos desenvolvedores com o SQS: CreateQueue, SendMessage, ReceiveMessage, ChangeMessageVisibility, e DeleteMessage. As APIs adicionais estão disponíveis para fornecer funcionalidade avançada. Em todos os casos, as APIs de SOAP e de consulta podem ser usadas com Java, C#, Perl e PHP.

## Antipadrões do Amazon SQS

- **Dados binários ou extensos** – As mensagens do Amazon SQS devem ser obrigatoriamente texto e de no máximo 64 KB de tamanho. Se os dados que você precisa armazenar em uma fila excederem esse tamanho, ou forem binários, é melhor usar o Amazon S3 ou o RDS para armazenar os dados binários ou extensos e armazenar um ponteiro para os dados no Amazon SQS.
- **Armazenamento de longo prazo** – Se os dados da mensagem precisarem ser armazenados por mais de 14 dias, o Amazon S3 ou algum outro mecanismo de armazenamento será mais apropriado.

## Amazon SimpleDB

O Amazon SimpleDB representa uma nova abordagem para armazenar e gerenciar dados estruturados na nuvem, o que difere da técnica tradicional de implantação de servidores de banco de dados relacionais. O SimpleDB é um armazenamento de dados não relacionais altamente disponível, escalável e flexível que libera boa parte do trabalho da administração do banco de dados e gerenciamento de sistemas associado. Com o Amazon SimpleDB, você tem um modelo de dados menos esquemático, e você pode armazenar itens de dados compostos por um número flexível de pares nome/valor.

## Cenário de uso ideal

Esta alternativa de armazenamento é bem adequada para situações nas quais você tem dados estruturados e refinados os quais você necessita que persistam e, em seguida, sejam consultados (principalmente orientados para interações de leitura), com alta disponibilidade e durabilidade, mas para os quais você não queira ou não necessita que gerem a sobrecarga de administrar um banco de dados relacional full-blown.

Várias organizações usam o SimpleDB como parte de uma arquitetura de armazenamento baseada em uma nuvem maior. Um padrão comum é utilizar o SimpleDB para controlar metadados sobre informações armazenadas em outras opções da AWS. Por exemplo, suponha que você esteja construindo uma solução baseada em nuvem que armazenará milhões de imagens em multi-megabytes. Nesta situação, você pode usar o Amazon Elastic Block Storage (EBS) ou o Amazon Simple Storage Service (Amazon S3) para armazenar as imagens em si, e o SimpleDB servindo como repositório para guardar os detalhes sobre cada imagem. Você pode então consultar o SimpleDB a fim de localizar uma determinada imagem e usar essa informação para recuperar os resultados das outras opções de armazenamento da Amazon.



## Desempenho

Aproveitando a infraestrutura comprovada de alta velocidade da Amazon, a abordagem SimpleDB cria um banco de dados rápido e altamente disponível que é muito escalável. Para uma taxa de transferência mais alta e uma latência menor, o Amazon SimpleDB oferece, por padrão, leituras realmente consistentes. Para aplicações que necessitam de “ler minha última gravação”, também é fornecida uma leitura consistente.

## Durabilidade e disponibilidade

O SimpleDB oferece uma durabilidade bastante elevada de dados por meio de sua replicação totalmente automática e geo-redundante. Várias réplicas de cada item de dados do SimpleDB são armazenadas em locais diferentes da sua região geográfica selecionada.

## Custos

Assim como todos os serviços web da Amazon, com o Amazon SimpleDB, você paga apenas pelo que usar, sem taxas iniciais ou contratos de longo prazo. Os custos para o SimpleDB incluem o armazenamento de dados estruturados (por GB ao mês), transferência de dados (por GB ao mês) e horas de máquina (por mês) associados às operações PUT e GET. O primeiro GB ao mês de armazenamento de dados estruturados e as primeiras 25 horas de utilização da máquina em um mês (aproximadamente 2M de operações API GET ou SELECT) são gratuitos. Usando este nível gratuito, muitos aplicativos nunca acarretarão em quaisquer encargos adicionais do SimpleDB.

A tabela de preços completa do Amazon SimpleDB pode ser encontrada em <http://aws.amazon.com/simplydb/pricing/>.

## Elasticidade e escalabilidade

Quando você cria um domínio SimpleDB, a AWS aloca um máximo de 10 GB de armazenamento por domínio. Como você paga apenas pelos recursos de computação e armazenamento à medida que você os utiliza, você é capaz de reduzi-los de forma transparente para até zero e, da mesma forma, expandir. Os requisitos de armazenamento e indexação são abordados pelo Amazon SimpleDB, que armazena os seus dados indexados automaticamente de maneira redundante dentro da região que você selecionou quando criou seu domínio do SimpleDB. Isso elimina a administração do banco de dados e permite que você simplesmente opere PUT e GET com os dados. Para a escalabilidade horizontal aumentar a taxa de transferência ou o tamanho do conjunto de dados incrementar além das limitações de um único domínio, você pode criar vários domínios SimpleDB e executar várias operações em paralelo. Para conjuntos de dados maiores do que 10 GB, você pode considerar particionar seus dados em múltiplos domínios do SimpleDB.

## Interfaces

Você tem um número de diferentes técnicas à sua disposição para interação programática com o SimpleDB. A Amazon oferece uma API em serviço web, disponível nos estilos SOAP e RESTful. A API atualmente oferece nove operações, permitindo o gerenciamento de domínio (criar, listar, excluir e obter metadados) e trabalhando com atributos (obter, escrever, individualmente ou em lote e excluir). Apesar do SQL padrão não estar disponível para o SimpleDB, você pode usar a operação SimpleDB Select para criar consultas SQL a fim de recuperar um conjunto de atributos com base em critérios que você fornecer.



Se você preferir usar um kit de desenvolvimento de software, kits de ferramentas e bibliotecas para plataformas e linguagens de desenvolvimento de software populares estão disponíveis, tais como:

- **Java**, usando o SDK da AWS para Java (<http://aws.amazon.com/sdkforjava/>) e o kit de ferramentas AWS para Eclipse (<http://aws.amazon.com/eclipse/>).
- **C#**, usando o SDK da AWS para .NET (<http://aws.amazon.com/sdkfornet/>) e Modelos do Visual Studio
- **PHP**, usando a instalação de base do PHP5, ou o SDK da AWS para PHP (<http://aws.amazon.com/sdkforphp/>).
- **Perl**, usando o Digest::SHA1, Bundle::LWP, and XML::Simple modules, os quais podem ser baixados do Comprehensive Perl Archive Network em <http://www.cpan.org>.

## Antipadrões do Amazon SimpleDB

É fácil determinar quais cenários de uso não são adequados para o SimpleDB. Se qualquer um dos seguintes critérios for aplicável, então considere utilizar uma das outras opções de armazenamento AWS:

- **Aplicativos pré-gravados vinculados a um banco de dados relacional tradicional**—Se você estiver tentando portar um aplicativo existente para a nuvem da AWS e precisa continuar usando o banco de dados relacional, você pode optar por usar o Amazon RDS (se o seu banco de dados for do tipo MySQL), ou um dos vários AMIs de bancos de dados pré-configurados do Amazon EC2. Você também fica livre para criar sua própria instância do Amazon EC2 e nela instalar um mecanismo de banco de dados.
- **Junções e/ou transações complexas**—Ainda que muitas soluções sejam capazes de alavancar o SimpleDB para oferecer suporte a seus usuários, é possível que o seu aplicativo requeira junções, transações complexas e outras infra-estruturas relacionais fornecidas por plataformas de banco de dados tradicionais. Se for o caso, você pode querer explorar o Amazon RDS ou o Amazon EC2 com um banco de dados instalado.
- **Dados BLOB (Objetos binários grandes)**—Se você planeja armazenar dados binários (por exemplo, vídeo, fotos ou música), você vai querer considerar como opção o Amazon Elastic Block Storage (EBS) ou o Amazon Simple Storage Service (Amazon S3). No entanto, o SimpleDB ainda tem um papel a desempenhar neste cenário, a fim de manter o controle dos metadados (por exemplo, nome do item, tamanho, data da criação, proprietário, localização e assim por diante) sobre seus objetos binários.
- **Dados digitados (numéricos)**—O SimpleDB armazena todos os dados como sequências de texto portanto, se você precisar manipular dados digitados ou numéricos, o RDS ou um banco de dados tradicional no Amazon EC2 podem ser uma solução melhor.
- **Grandes quantidades de dados**—Como já foi dito, o SimpleDB oferece um modelo de armazenamento para muitos aplicativos a um preço atraente. No entanto, se você se depara com a manutenção de conjuntos de dados muito grandes, você também deve considerar as alternativas de armazenamento Amazon S3 ou EBS. Além disso, você também poderá empregar o SimpleDB como um mecanismo eficiente para manter o controle dos metadados sobre seus grandes conjuntos de dados.

## Amazon Relational Database Service (Amazon RDS)

---

O Amazon Relational Database Service (RDS) é uma base de dados do tipo MySQL totalmente funcional e relacional fornecida como um serviço gerenciado, baseado em nuvem. Se o aplicativo exigir o armazenamento relacional, mas você quiser reduzir o tempo gasto na gestão do banco de dados, o Amazon RDS automatizará as tarefas administrativas comuns para reduzir a complexidade e o custo total da propriedade. O Amazon RDS faz o backup automático do banco de dados e mantém o software do banco de dados, permitindo que você dedique mais tempo ao desenvolvimento do aplicativo. É ideal para dados estruturados de forma tradicional que requeiram consultas mais sofisticadas e mais recursos de juntas do que o oferecido pelo outro banco de dados da Amazon, o SimpleDB.

### Cenário de uso ideal

O Amazon RDS é uma ótima escolha para qualquer aplicativo que se baseia no MySQL como seu repositório de informações e se você quiser tirar proveito de um banco de dados altamente escalável, de baixa manutenção, econômico, baseado em nuvem sem a necessidade de fazer quaisquer alterações no código.

### Desempenho

O Amazon RDS fornece alto desempenho através de uma combinação de instâncias configuráveis em execução em infraestrutura de classe mundial com operações de backup e manutenção totalmente automatizadas comprovadas pela Amazon. As configurações de banco de dados disponíveis variam de pequena instância (plataforma de 64 bits com 1,7 GB de RAM e 1 unidade de computação elástica (ECU)) até uma instância quádrupla extra-grande (plataforma de 64 bits com 68 GB de RAM e 26 ECUs).

Para obter o desempenho ideal, os administradores e desenvolvedores de banco de dados são responsáveis por selecionar o perfil de instância adequada (incluindo RAM e armazenamento) para suas necessidades específicas de computação. Se o ajuste for necessário, eles estão livres para usar a API fornecida pela Amazon para ajustar suas configurações de banco de dados. O Amazon CloudWatch fornece métricas que podem ajudar a determinar quando o ajuste de desempenho e/ou elasticidade adicional é necessário. Finalmente, os administradores que optarem por mover-se para uma configuração de banco de dados maior devem procurar superar os recursos da sua instância atual.

### Elasticidade e escalabilidade

Os recursos do RDS podem ser dimensionados elasticamente em várias dimensões: tamanho de armazenamento de banco de dados, capacidade de computação de instância de banco de dados e o número de réplicas de leitura.

Para alcançar a elasticidade de computação adicional, você pode configurar instâncias adicionais do Amazon RDS e aproveitar o particionamento para distribuir a carga de trabalho.

Para dimensionar de forma elástica o armazenamento de banco de dados do RDS, você pode usar as ferramentas de linha de comando, API ou o AWS Management Console para solicitar armazenamento adicional. Dependendo das necessidades, esse armazenamento pode ser adicionado imediatamente ou durante a próxima janela de manutenção.

Expansão ou redução de recursos computacionais é facilmente realizado com um único comando de API ou do AWS Management Console. Por exemplo, talvez seja necessário poder adicional de computação para criar faturas no final de cada mês. É simples expandir temporariamente até uma instância quádrupla extra-grande, execute as cargas de trabalho computacionais intensivas e, em seguida, retorne para uma configuração menor, mais econômica para o restante do mês.

Para dimensionar recursos do banco de dados do RDS dentro ou fora (dimensionamento horizontal usando várias instâncias do RDS), os administradores podem criar uma ou mais réplicas de leitura do RDS. As réplicas de leitura usam o recurso de replicação assíncrona interna do MySQL para permitir que você dimensione elasticamente os recursos do RDS para oferecer suporte as leituras de cargas de trabalho pesadas.

Finalmente, os administradores podem configurar as instâncias adicionais do Amazon RDS e aproveitar o particionamento do banco de dados ou compartilhamento para distribuir a carga de trabalho, alcançar uma maior elasticidade e escalabilidade de banco de dados

## **Durabilidade e disponibilidade**

Para maior durabilidade, o RDS oferece dois tipos de backups de bancos de dados que são replicados em várias Zonas de disponibilidade: backups automatizados de instância de banco de dados e snapshots de banco de dados iniciada pelo usuário. Se você habilitar backups automatizados de instância de banco de dados, o RDS executará automaticamente um backup completo diário de seus dados durante a janela de backup especificada e também irá capturar os logs de transações do banco de dados. Esses backups automatizados são fornecidos sem custo adicional, podem ser retidos por até oito dias e podem ser usados para fazer uma restauração em determinado momento para qualquer ponto desde o início do período de retenção até os últimos cinco minutos da hora atual. Os snapshots de banco de dados iniciados pelo usuário, podem ser criados a qualquer momento e são mantidos até serem excluídos. Os snapshots de banco de dados permitem restaurar seu banco de dados para um estado conhecido.

O recurso de implantação do RDS Multi-AZ aumenta a durabilidade e a disponibilidade do seu banco de dados sincronicamente por replicar os dados entre uma instância de banco de dados RDS primária e uma instância em espera em outra zona de disponibilidade. No caso improvável de uma falha de componente de banco de dados ou falha em uma Zona de disponibilidade, o RDS fará automaticamente o failover para o modo de espera (que geralmente leva cerca de três minutos) e as transações de banco de dados podem ser retomadas logo que o modo de espera for promovido. A replicação sincronizada garante que não há nenhuma perda de dados.

Observe que a replicação sincronizada fornecida pelo recurso de implantação do RDS Multi-AZ é complementar para a replicação assíncrona incorporada fornecida pela réplica de leitura do RDS. Você pode usar o recurso sozinho ou ambos em combinação.

## **Custos**

Como em todos os serviços da Amazon Web Services, com o Amazon RDS você paga somente pelo que usar, sem taxas mínimas ou contratos de longo prazo. O Amazon RDS oferece uma estrutura hierárquica de preços, com base no tamanho da instância do banco de dados, o tipo de implantação (AZ única/Multi-AZ) e a região da AWS. A definição de preço para o Amazon RDS é baseada em vários fatores: as horas de instância de banco de dados (por hora), a quantidade de armazenamento de banco de dados configurado (por GB ao mês e por milhões de solicitações de E/S), armazenamento de backup adicional (por GB ao mês) e a transferência de dados de entrada/saída (por GB ao mês). Informações completas sobre a definição de preço do RDS podem ser encontradas em <http://aws.amazon.com/rds/pricing/>.

## **Interfaces**

Para ajudá-lo a começar, a Amazon oferece uma série de scripts de linha de comando fáceis de entender para criação, manutenção, monitoramento e travamento de sua instância de banco de dados. Como alternativa, você pode usar as operações baseadas em serviços web da Amazon para realizar a mesma série de tarefas.

Uma vez que sua instância foi criada, você pode configurar seu schema e dados usando qualquer ferramenta que trabalhe com um banco de dados relacional. Dependendo da quantidade e localização dos dados existentes no local, em vez disso você pode optar por extrair os dados locais usando o utilitário mysqldump e canalizar diretamente para o executável do MySQL para inserção no Amazon RDS. Para conjuntos de dados maiores, pode ser mais vantajoso construir primeiro o seu schema no Amazon RDS, extrair seus dados localmente em um arquivo simples e, em seguida, usar o utilitário mysqlimport para carregá-lo no Amazon RDS. O tópico de migração para o Amazon RDS será discutido com mais detalhes em um próximo whitepaper.

Depois que seu schema e dados estiverem no lugar, você interage com suas informações através do padrão MySQL SQL, bem como JDBC e outras APIs populares, bem como quaisquer ferramentas gráficas que possam trabalhar com dados relacionais. Não há nenhuma alteração de código para ser feita para permitir que o aplicativo interaja com RDS: você simplesmente substitui o endereço do seu servidor de banco de dados (por exemplo, dbserver.yourcompany.com) com a ponto de acesso público DNS (por exemplo, myinstance.c0cafggtpzd2.us-east-1.rds.amazonaws.com) fornecido pelo AWS quando você cria a instância. Esse ponto de acesso do DNS permanecerá o mesmo para o tempo de vida de sua instância. Além de configurar o ponto de acesso, tudo sobre seu aplicativo baseado em MySQL é inalterado.

### Antipadrões do Amazon RDS

O Amazon RDS é uma ótima solução para dados de MySQL com base em nuvem, mas em um número de cenários não é a escolha certa:

- **Índice e consulta de dados concentrados** – Muitas soluções baseadas em nuvem não requerem recursos avançados encontrados em um banco de dados relacional, como associações e transações complexas. Se seu aplicativo é mais orientado para indexação e consulta de dados, você pode encontrar no Amazon SimpleDB a solução mais adequada às suas necessidades.
- **Numerosos objetos binários grandes (BLOBs)** – Enquanto o MySQL (e o Amazon RDS) está equipado para armazenar BLOBs, se seu aplicativo usam muito eles (arquivos de áudio, vídeos, imagens e assim por diante), você pode encontrar no Amazon Simple Storage Service (Amazon S3) uma escolha melhor.
- **Elasticidade automática** – Como mencionado anteriormente, cabe aos administradores configurar seu ambiente do Amazon RDS para alcançar a elasticidade. Se você deseja elasticidade automatizada (e suas estruturas de dados possuem um bom ajuste), você pode optar por outra opção de armazenamento, como o Amazon SimpleDB ou o Amazon S3.
- **Outras plataformas de banco de dados** – Neste momento, o Amazon RDS fornece um banco de dados MySQL. Se você precisar de outra plataforma de banco de dados (como o IBM DB2 ou Informix, Microsoft SQL Server, Oracle, PostgreSQL ou Sybase) você precisará implantar um banco de dados relacional especializado em AMI do Amazon EC2 ou criar uma de sua preferência.

### Banco de dados relacional do Amazon EC2 / AMIs do banco de dados relacional

O Amazon EC2, juntamente com os volumes do EBS, fornece uma plataforma ideal para que você possa operar seu próprio banco de dados relacional em nuvem. Muitas soluções de banco de dados líderes estão disponíveis como pré-construídas, AMIs prontas para usar do Amazon EC2, incluindo IBM DB2 e Informix, Oracle Database, MySQL, Microsoft SQL Server, PostgreSQL, Sybase, EnterpriseDB e Vertica.

## Cenário de uso ideal

Executar um banco de dados relacional no Amazon EC2 e no EBS é o cenário ideal para utilizadores cuja aplicação requer um banco de dados relacional tradicional específico ou para aqueles usuários que exigem um nível máximo de controle administrativo e configurabilidade.

## Desempenho

O desempenho de uma instância de banco de dados relacional no Amazon EC2 depende de muitos fatores, incluindo o tipo de instância do Amazon EC2, o número e configuração dos volumes do EBS, o software de banco de dados, a sua configuração e a carga de trabalho do aplicativo. Em geral, você pode esperar que o desempenho do banco de dados no Amazon EC2 seja semelhante ao desempenho do mesmo banco de dados instalado em equipamentos com configuração semelhante no local. Recomendamos você avalie o desempenho de seu aplicativo em funcionamento em vários tipos de instância do Amazon EC2 usando várias configurações de armazenamento para selecionar a melhor configuração.

Para aumentar o desempenho do banco de dados você pode expandir a memória e os recursos de computação, escolhendo um tamanho maior de instância do Amazon EC2. Para expandir o desempenho E/S você pode alterar o número de volumes do EBS ou usar o software RAID 0 (distribuição de disco) em vários volumes do EBS, que irão aumentar o IOPS e a largura de banda total. Em muitos casos, você também pode dimensionar o desempenho de sistema de banco de dados total dimensionando horizontalmente com banco de dados de cluster, replicação e várias leituras secundárias. Em geral, você tem as mesmas opções de ajuste de desempenho do banco de dados no ambiente do Amazon EC2 que você tem em um ambiente de servidor físico.

## Durabilidade e disponibilidade

Bancos de dados relacionais no Amazon EC2 fornecem armazenamento persistente para dados estruturados usando os volumes do EBS como o armazenamento de dados, então todas as notas sobre a durabilidade e a disponibilidade de dados do EBS se aplicam aqui também. E, novamente, a durabilidade e a disponibilidade básica e de dados relacionais armazenados em volumes do EBS podem ser reforçadas usando os snapshots do EBS ou usando os utilitários de backup de banco de dados de terceirizados (tais como o RMAN da Oracle ) para armazenar backups do banco de dados no Amazon S3.

## Custos

Executando um banco de dados no Amazon EC2, você paga apenas pelo que usar, sem taxas mínimas ou contratos de longo prazo. O custo da execução do seu próprio banco de dados no Amazon EC2 depende do tamanho e do número de instâncias do Amazon EC2 usados para executar o seu banco de dados, o tamanho dos volumes do EBS usado para o armazenamento de banco de dados, a quantidade de dados transferidos dentro e fora do Amazon EC2, e, em muitos casos, o custo da licença do software de banco de dados de empresas terceirizadas. Muitos pacotes de banco de dados de código aberto usam um modelo de licença sem custo; alguns fornecedores de software comercial usam o modelo Amazon DevPay; muitos outros fornecem um modelo que usa a licença que o cliente já possui. Entre em contato com seu fornecedor de software de banco de dados ou a Amazon Web Services para entender os modelos de definição de preços de licença utilizado.

Informações detalhadas sobre o Amazon EC2, o EBS e os preços de transferência de dados podem ser encontradas na web em <http://aws.amazon.com/ec2/pricing/>.

## Elasticidade e escalabilidade

Em muitos casos, os usuários de soluções tradicionais de banco de dados relacional no Amazon EC2 podem aproveitar a elasticidade e escalabilidade da plataforma subjacente da AWS. Por exemplo, depois de configurar uma instância do Amazon EC2 com a sua solução de banco de dados, você pode empacotar a instância em uma AMI personalizada, usando qualquer conjunto de comandos para AMIs de armazenamento de instância ou usando o comando Criar imagens de comando para AMIs do EBS. Você pode criar várias instâncias novas de sua configuração de banco de dados rapidamente.

## Banco de dados relacional no antipadrões do Amazon EC2

Executar o seu próprio banco de dados relacional no Amazon EC2 é uma ótima solução para muitos usuários, mas existe um número de cenários onde outras soluções podem ser a melhor opção:

- **Índice e consulta de dados concentrados** – Muitas soluções baseadas em nuvem não requerem recursos avançados encontrados em um banco de dados relacional, como associações ou transações complexas. Se seu aplicativo é mais orientado para indexação e consulta de dados, você pode encontrar no Amazon SimpleDB a solução mais adequada às suas necessidades e ele é significativamente mais fácil de gerenciar.
- **Numerosos objetos binários grandes (BLObs)** – Muitos bancos de dados relacionais oferecem suporte à BLObs (arquivos de áudio, vídeos, imagens e assim por diante). Se seu aplicativo usa muito eles, você pode encontrar no Amazon S3 uma boa escolha, usando um banco de dados relacional ou o SimpleDB para o gerenciamento de metadata.
- **Elasticidade automática** – Como mencionado anteriormente, os usuários de bancos de dados relacionais na AWS podem em muitos casos aproveitar a elasticidade e a escalabilidade da plataforma subjacente da AWS, mas isso exige que os administradores de sistema ou DBAs executem uma tarefa manual ou com scripts. Supondo que você precise ou quer elasticidade totalmente automatizada (e que suas estruturas de dados possuem um bom ajuste), você pode optar por outra opção de armazenamento, como o Amazon SimpleDB ou o Amazon S3.

## Armazenamento em nuvem Casos de uso

Para exemplos de usos reais das opções de armazenamento da AWS, consulte o whitepaper complementar:

Opções de armazenamento em nuvem da AWS: Casos de uso em:

[http://media.amazonwebservices.com/AWS\\_Storage\\_Use\\_Cases.pdf](http://media.amazonwebservices.com/AWS_Storage_Use_Cases.pdf).

## Referências e leitura complementar

### Serviços de armazenamento da AWS

---

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

Volumes de armazenamento de instância do Amazon EC2 –

[http://docs.amazonwebservices.com/AWSAmazon\\_EC2/latest/UserGuide](http://docs.amazonwebservices.com/AWSAmazon_EC2/latest/UserGuide)

(consulte as seções Tipos de instância, Armazenamento de instância e Mapeamento de dispositivo de bloco)

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

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

Amazon SimpleDB – <http://aws.amazon.com/simpledb>

Banco de dados em execução na AWS – [http://aws.amazon.com/running\\_databases](http://aws.amazon.com/running_databases)

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





## Artigos de armazenamento AWS

---

Artigo de Werner Vogels, “Choosing Consistency” –

[http://www.allthingsdistributed.com/2010/02/strong\\_consistency\\_simpledb.html](http://www.allthingsdistributed.com/2010/02/strong_consistency_simpledb.html)

Artigo do blog da Amazon Web Services, “Amazon SimpleDB Consistency Enhancements” –

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

Artigo do Centro do desenvolvedor da Amazon Web Services, “Amazon SimpleDB Consistency Enhancements” –

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

Calculadora mensal simples da Amazon Web Services – <http://aws.amazon.com/calculator>

Perguntas frequentes sobre a Oracle em nuvem – <http://www.oracle.com/technetwork/topics/cloud/faq-098970.html>

Artigo do Centro do desenvolvedor da Amazon Web Services, “Running Databases on AWS” –

[http://aws.amazon.com/running\\_databases/](http://aws.amazon.com/running_databases/)

Estudo de caso da Oracle: Banco de dados Oracle em nuvem –

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

## Outras informações da AWS

---

Nível de uso gratuito da AWS – <http://aws.amazon.com/pt/free>

Conjuntos de dados públicos na AWS – <http://aws.amazon.com/publicdatasets>

AWS Import/Export – <http://aws.amazon.com/importexport>

Amazon CloudFront – <http://aws.amazon.com/cloudfront>

Estudos de caso da AWS – <http://aws.amazon.com/solutions/case-studies>



## Apêndice: Referência rápida de armazenamento

	Dados desestruturados			Dados estruturados		
	Armazenamento de instância do Amazon EC2	Volumes do Amazon EBS	Amazon S3	Amazon SimpleDB	Outros bancos de dados relacionais (no EC2 e no EBS)	Amazon RDS
<b>Desempenho</b>	Alto	Alta	Moderado (thread único) a muito alto (vários threads)	Moderado a alto (lote Puts / Gets)	Alta	Alta
<b>Durabilidade</b>	Baixa	Moderada	Alta	Alta	Alta	Moderada
<b>Custos</b>	Custo incluídos no EC2	Provisionado por GB/mês	Armazenado por GB/mês	Provisionado Primeiro GB grátis, então por GB/mês	Provisionado (mesmo que o EBS)	Provisionado por GB/mês (mínimo de 5 GB)
<b>Disponibilidade</b>	Baixa	Moderada a alta (usando snapshots do EBS)	Alta	Alta	Moderada a alta	Alta
<b>Elasticidade e escalabilidade</b>	Não	Manual (adicionando mais volumes)	Automático	Automático	Manual	Manual (um comando para modificar a instância de banco de dados)
<b>Limites de tamanhos</b>	160 GB a 1,6 TB (instâncias maiores têm mais e maiores volumes)	1 GB a 1 TB por volume (podem ser usados vários volumes ou distribuição para capacidades maiores)	Efetivamente ilimitado (5 TB por objeto, objetos ilimitados por bucket)	10 GB/domínio 100 domínios (mais domínios disponíveis mediante solicitação)	(mesmo que o EBS)	5 GB a 1 TB por instância de banco de dados
<b>Persistência em instâncias</b>	Não	Sim	Sim	Sim	Sim	Sim
<b>Interfaces</b>	Dispositivo de bloco, acesso via sistema SO / sistema de arquivos no EC2	N/D, acesso através do SO do EC2 / sistema de arquivos	HTTP, REST ou SOAP	REST ou SOAP	Bibliotecas MySQL ou JDBC	Bibliotecas MySQL ou JDBC
<b>Segurança (criptografia em repouso)</b>	Executar FS criptografado	Executar FS criptografado	Criptografar usando AES de 256 bits.	Criptografar usando AES de 256 bits.		
<b>Segurança (criptografia em trânsito)</b>	N/D	N/D	SSL (HTTPS)	SSL (HTTPS)	SSL (HTTPS)	SSL (HTTPS)
<b>Plataformas suportadas pelo RDBMS</b>	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
<b>Modelo (relacional ou não)</b>	Bloco	Bloco	Objeto	Esquema não relacional, flexível, loja de entidade	Relacional	Relacional

	Dados desestruturados			Dados estruturados		
	Armazenamento de instância do Amazon EC2	Volumes do Amazon EBS	Amazon S3	Amazon SimpleDB	Outros bancos de dados relacionais (no EC2 e no EBS)	Amazon RDS
<b>Grau de automação</b>	Nenhum	Auto espelhamento	Auto-replicação, Versão	Indexação, replicação, provisionamento, aplicação de patches	Depende do banco de dados	Backups e patch de software automatizados
<b>Grau de redundância</b>	Não redundante	Redundante dentro de uma Zona de disponibilidade	Altamente redundante em vários datacenters	Manter várias cópias de todos os dados do usuário, diversas geograficamente	Nenhum (replicação assíncrona disponível)	Oferecer ambas as opções de instância de banco de dados única (um AZ) e Multi-AZ
<b>Acesso à instância cruzada (por exemplo, compartilhável)</b>	Não	Não	Sim	Sim	Sim	Sim
<b>Gerenciamento e administração</b>	Manual	Manual	Auto	Auto	Manual	Auto