Abordagens de backup e recuperação que usam a AWS

Junho de 2016

Avisos

Este documento é fornecido apenas para fins informativos. Ele relaciona as atuais ofertas de produtos e práticas da AWS a contar da data de emissão deste documento, que estão sujeitas a alterações sem aviso prévio. Os clientes são responsáveis por fazer sua própria avaliação independente das informações deste documento e de qualquer uso dos produtos ou serviços da AWS, cada um dos quais é fornecido “no estado em que se encontra”, sem garantia de qualquer tipo, expressa ou implícita. Este documento não cria quaisquer garantias, representações, compromissos contratuais ou condições da AWS, suas afiliadas, seus fornecedores ou licenciadores. As responsabilidades e as obrigações da AWS em relação aos seus clientes são controladas por acordos da AWS, e este documento não integra nem modifica qualquer acordo entre a AWS e seus clientes.

Índice

Resumo

Este documento se destina a arquitetos de soluções empresariais, arquitetos de backup e administradores de TI responsáveis pela proteção de dados em seus ambientes de TI corporativos. Ele apresenta cargas de trabalho de produção e arquiteturas que podem ser implementadas usando a AWS para aumentar ou substituir uma solução de backup e recuperação. Essas abordagens oferecem custos mais baixos, maior escalabilidade e mais durabilidade para atender aos requisitos de Recovery Time Objective (RTO – Objetivo do tempo de recuperação) e Recovery Point Objective (RPO – Objetivo do ponto de recuperação) e de conformidade.

Introdução

Com a aceleração do aumento dos dados empresariais, a tarefa de protegê-los se torna mais desafiadora. Dúvidas sobre a durabilidade e a escalabilidade de métodos de backup são comuns, incluindo esta: Como a nuvem ajuda a atender às minhas necessidades de backup e arquivamento?

Este documento abrange várias arquiteturas de backup (aplicativos nativos de nuvem, ambientes híbridos e no local) e serviços da AWS associados que podem ser usados para desenvolver soluções de proteção de dados escaláveis e confiáveis.

Por que usar a AWS como uma plataforma de proteção de dados?

A Amazon Web Services (AWS) é uma plataforma de computação em nuvem segura, de alto desempenho, flexível, econômica e fácil de usar. A AWS cuida do trabalho pesado não diferenciado e oferece ferramentas e recursos que você pode usar para desenvolver soluções de backup e recuperação escaláveis.

Há muitas vantagens no uso da AWS como parte da sua estratégia de proteção de dados:

- **Durabilidade:** O Amazon Simple Storage Service (Amazon S3) e o Amazon Glacier foram desenvolvidos para proporcionar 99,999999999% (11 noves) de durabilidade para os objetos armazenados neles. Ambas as plataformas oferecem locais confiáveis para dados de backup.
- **Segurança:** A AWS oferece várias opções para controle de acesso e criptografia de dados em trânsito e em repouso.
- **Infraestrutura global:** Os serviços da AWS estão disponíveis em todo o mundo. Portanto, você pode fazer backup e armazenar dados na região que atende aos seus requisitos de conformidade.
- **Conformidade:** A infraestrutura da AWS é certificada para conformidade com padrões, como Service Organization Controls (SOC - Relatório de controles de empresa de serviços), Statement on Standards for Attestation Engagements (SSAE - Relatório de declaração sobre normas para comprovação de contratos) 16, International Organization for Standardization (ISO - Organização Internacional de Normalização) 27001, Payment Card Industry Data Security Standard (PCI DSS - Padrão de segurança de dados da Indústria de cartões de pagamento), Health Insurance Portability and Accountability Act (HIPPA - Lei de portabilidade e responsabilidade de provedores de saúde), SEC[1] e Federal Risk and Authorization Management Program (FedRAMP - Programa federal de gerenciamento de risco e autorização) para que você possa facilmente adaptar a solução de backup ao seu regime de conformidade.
- **Escalabilidade:** Com a AWS, você não precisa se preocupar com a capacidade. É possível aumentar ou reduzir a escala do consumo conforme a variação das suas necessidades sem custos indiretos administrativos.
- **TCO mais baixo:** A escala de operações da AWS reduz os custos de serviço e ajuda a reduzir o custo total de propriedade (TCO) do armazenamento. A AWS repassa essas economias de custo aos clientes na forma de queda de preços.
- **Definição de preço com pagamento conforme o uso:** Adquira serviços da AWS conforme sua necessidade e somente pelo período que planeja usá-los. A definição de preço da AWS não tem taxas iniciais, multas por rescisão nem contratos de longo prazo.

[1] https://aws.amazon.com/about-aws/whats-new/2015/09/amazon-glacier-receives-third-party-compliance-assessment-for-sec-rule-17a-4f-from-cohasset-associates-inc/

Serviços de armazenamento da AWS para proteção de dados

O Amazon S3 e o Amazon Glacier são serviços ideais para backup e arquivamento. Ambos são plataformas de armazenamento duráveis e de baixo custo. Ambos oferecem capacidade ilimitada e não exigem gerenciamento de volume ou mídia à medida que os conjuntos de dados de backup crescem. O modelo de pagamento conforme o uso e o baixo custo por GB/mês tornam esses serviços uma boa opção para casos de uso de proteção de dados.

Amazon S3

O Amazon S3 oferece armazenamento de objetos altamente seguro e escalável.

Você pode usar o Amazon S3 para armazenar e recuperar qualquer volume de dados, a qualquer momento, de qualquer lugar na Web. O Amazon S3 armazena dados como objetos nos recursos chamados *buckets*. O AWS Storage Gateway e muitas soluções de backup de terceiros podem gerenciar objetos do Amazon S3 por você. O bucket permite armazenar quantos objetos você quiser, além de gravar, ler e excluir objetos. Um único objeto pode ter um tamanho de até 5 TB.

O Amazon S3 oferece diversas classes de armazenamento criadas para diferentes casos de uso. Entre elas estão:

- **Amazon S3 Standard** para armazenamento de finalidade geral de dados acessados com frequência.
- **Amazon S3 Standard - Acesso pouco frequente** para dados duradouros, mas acessados com menos frequência.
- **Amazon Glacier** para arquivamento de longo prazo.

O Amazon S3 também oferece políticas de ciclo de vida que você pode configurar para gerenciar seus dados durante todo o ciclo de vida deles. Após a definição de uma política, seus dados serão migrados para a classe de armazenamento apropriada sem alterações no aplicativo. Para obter mais informações, consulte Classes de armazenamento do S3.

Amazon Glacier

O Amazon Glacier é um serviço de armazenamento de arquivos em nuvem de custo extremamente baixo que fornece armazenamento seguro e durável para backup on-line e arquivamento de dados. Para manter os custos baixos, o Amazon Glacier é otimizado para dados que são acessados com pouca frequência e para os quais são aceitáveis tempos de recuperação de várias horas. Com o Amazon Glacier, você pode armazenar com segurança grandes ou pequenas quantidades de dados por apenas 0,007 USD por gigabyte por mês, o que representa uma economia significativa em comparação com soluções no local. O Amazon Glacier é ideal para o armazenamento de dados de backup com requisitos de retenção longa ou indefinida e para arquivamento de dados em longo prazo. Para obter mais informações, consulte Amazon Glacier.

AWS Storage Gateway

O AWS Storage Gateway conecta um dispositivo de software no local ao armazenamento baseado em nuvem para proporcionar uma integração perfeita e altamente segura entre o ambiente de TI no local e a infraestrutura de armazenamento da AWS. Para obter mais informações, consulte AWS Storage Gateway.

Serviços de transferência da AWS

Além de gateways e conectores de terceiros, é possível usar opções da AWS, como AWS Direct Connect, AWS Snowball, AWS Storage Gateway e Amazon S3 Transfer Acceleration para transferir dados rapidamente. Para obter mais informações, consulte Migração de dados para a nuvem.

Criação de uma solução de backup e recuperação

Ao desenvolver uma estratégia abrangente para fazer backup e restaurar dados, primeiro é preciso identificar a falha ou situações de desastre que possam ocorrer e seu possível impacto nos negócios. Em alguns setores, é preciso considerar os requisitos regulatórios para segurança de dados, privacidade e retenção de registros.

Você deve implementar processos de backup que oferecerão o nível apropriado de granularidade para atender ao RTO e ao RPO da empresa, incluindo:

- Recuperação no nível do arquivo
- Recuperação no nível do volume
- Recuperação no nível do aplicativo (por exemplo, bancos de dados)
- Recuperação no nível da imagem

As seções a seguir descrevem as abordagens de backup, recuperação e arquivamento baseadas na organização da sua infraestrutura. A infraestrutura de TI pode ser amplamente categorizada como nativa de nuvem, no local e híbrida.

Infraestrutura nativa de nuvem

Este cenário descreve um ambiente de carga de trabalho que reside totalmente na AWS. Como mostra a figura a seguir, ele inclui servidores da Web, servidores de aplicativos, servidores de monitoramento, bancos de dados e Active Directory.

Se você estiver executando todos os seus serviços da AWS, poderá aproveitar muitos recursos integrados para atender às suas necessidades de recuperação e proteção de dados.

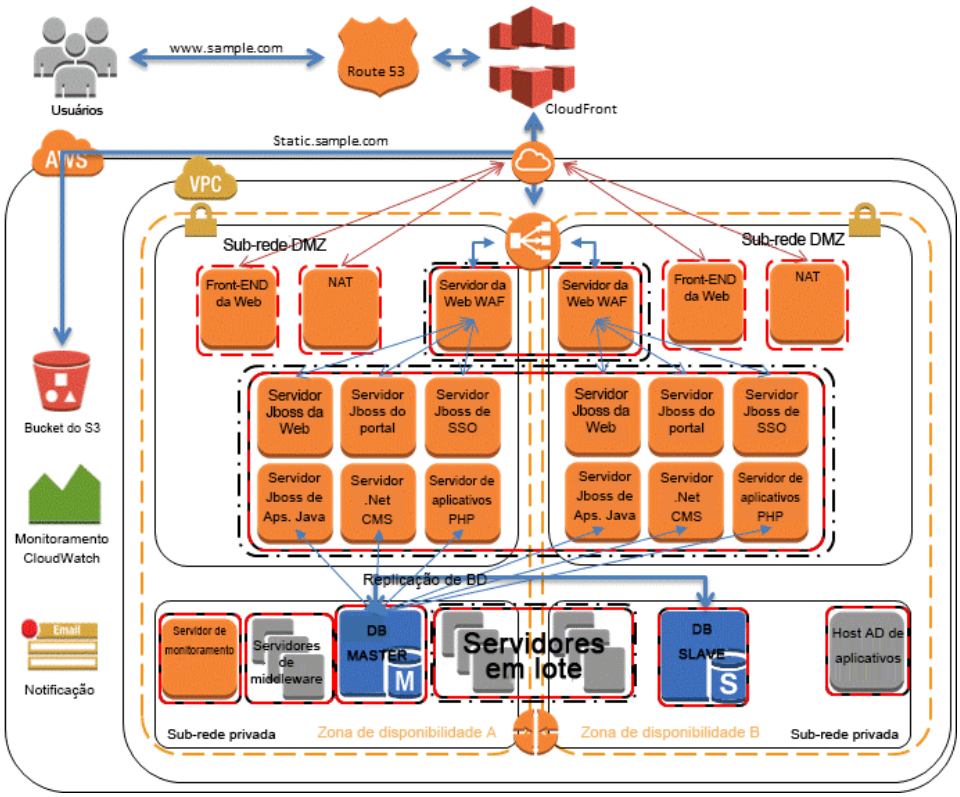

Figura 1: Cenário nativo da AWS

Proteção baseada em snapshot do EBS

Quando os serviços estão em execução no Amazon Elastic Compute Cloud[2] (Amazon EC2), as instâncias de computação podem usar os volumes do Amazon Elastic Block Store (Amazon EBS) para armazenar e acessar dados primários. É possível usar esse armazenamento em blocos para dados estruturados, como bancos de dados, ou dados não estruturados, como arquivos em um sistema de arquivos no volume.

O Amazon EBS permite criar snapshots (backups) de qualquer volume do Amazon EBS. Ele coloca uma cópia do volume no Amazon S3, onde ela é armazenada de maneira redundante em várias zonas de disponibilidade. O primeiro snapshot é uma cópia completa do volume. Snapshots contínuos armazenam somente alterações incrementais no nível do bloco.

Esse é um modo rápido e confiável de restaurar dados do volume inteiro. Se você precisa apenas de uma restauração parcial, pode anexar o volume à instância em execução com um nome diferente, montá-lo e, em seguida, usar os comandos de cópia do sistema operacional para copiar os dados do volume de backup para o volume de produção.

Também é possível copiar os snapshots do Amazon EBS entre as regiões da AWS usando o recurso de cópia de snapshot do Amazon EBS disponível no console ou na linha de comando, conforme descrito no Guia do usuário do Amazon Elastic Cloud Compute.[3] Você pode usar esse recurso para armazenar seu backup em outra região sem precisar gerenciar a tecnologia de replicação subjacente.

---

[2] http://aws.amazon.com/ec2/

[3] http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-copy-snapshot.html

Criação de snapshots do EBS

Ao criar um snapshot, você protege seus dados diretamente em um armazenamento durável baseado em disco. É possível usar o Console de Gerenciamento da AWS, a interface de linha de comando (CLI) ou as APIs para criar o snapshot do Amazon EBS.

No console do Amazon EC2, na página **Elastic Block Store Volumes**, escolha **Criar snapshot** no menu **Ações**. Na caixa de diálogo **Criar snapshot**, escolha **Criar** para criar um snapshot que será armazenado no Amazon S3.

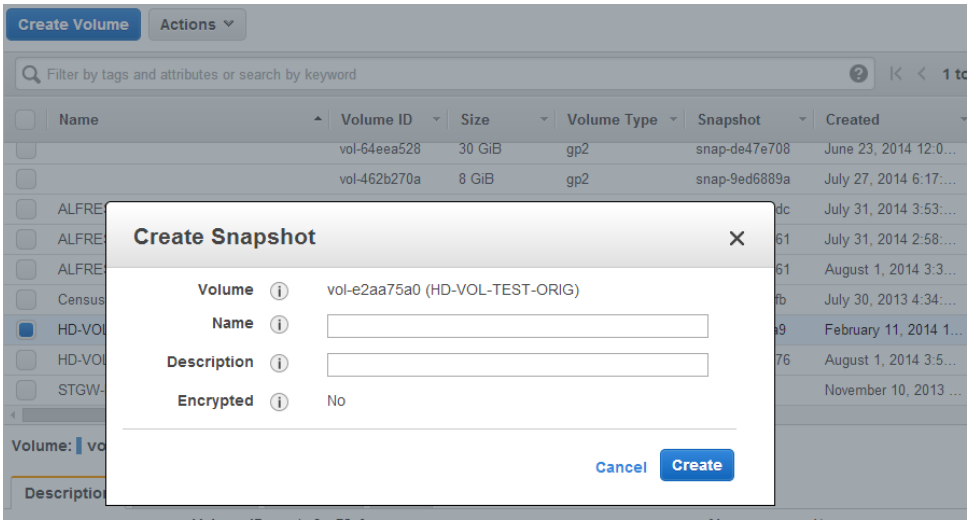

Figura 2: Uso do console do EC2 para criar um snapshot

Para usar o comando CLI com a finalidade de criar o snapshot, execute o comando a seguir:

```
➢  aws ec2 create-snapshot
```

É possível programar e executar o comando `aws ec2 create-snapshot` regularmente para fazer backup dos dados do EBS. A definição de preço econômico do Amazon S3 permite reter muitas gerações de dados. E como os snapshots são baseados em blocos, você só consome espaço para dados alterados após a criação do snapshot inicial.

Restauração a partir de um snapshot do EBS

Para restaurar dados a partir de um snapshot, você pode usar o Console de Gerenciamento da AWS, a CLI ou as APIs para criar um volume a partir de um snapshot existente.

Por exemplo, siga estas etapas para restaurar um volume para um backup feito anteriormente:

1. Use o comando a seguir para criar um volume a partir do snapshot de backup:

```
➢ aws ec2 create-volume –-region us-west-1b –-snapshot-
  id mysnapshot-id
```

2. Na instância do Amazon EC2, desmonte o volume existente.

   No Linux, use `umount`. No Windows, use o Logical Volume Manager (LVM).

3. Use o comando a seguir para separar o volume existente da instância:

```
➢ aws ec2 detach-volume --volume-id oldvolume-id –
  instance-id myec2instance-id
```

4. Use o comando a seguir para anexar o volume que foi criado a partir do snapshot:

```
➢ aws ec2 attach-volume --volume-id newvolume-id
  --instance-id myec2instance-id --device /dev/sdf
```

5. Monte novamente o volume na instância em execução.

Criação de backups consistentes ou a quente

Quando você faz um backup, é melhor que o sistema não esteja executando nenhuma E/S. No caso ideal, a máquina não está aceitando tráfego, mas isso fica cada vez mais raro à medida que as operações de TI 24/7 se estabelecem como norma.

Por esse motivo, é preciso aquietar o sistema de arquivos ou o banco de dados para fazer um backup sem erros. O modo como você faz isso depende do banco de dados ou do sistema de arquivos.

O processo de um banco de dados é assim:

- Se possível, passe o banco de dados para o modo de backup a quente.
- Execute os comandos de snapshot do Amazon EBS.
- Tire o banco de dados do modo de backup a quente ou, se estiver usando uma réplica de leitura, encerre a instância dela.

O processo para um sistema de arquivos é semelhante, mas depende dos recursos do sistema operacional ou do sistema de arquivos. Por exemplo, XFS é um sistema de arquivos capaz de liberar os dados para um backup consistente. Para obter mais informações, consulte xfs_freeze.[4]

Se o sistema de arquivos não oferecer suporte à capacidade de congelar, você deverá desmontá-lo, emitir o comando de snapshot e montar novamente o sistema de arquivos. Alternativamente, é possível facilitar esse processo usando um gerenciador de volumes lógicos que seja compatível com o congelamento de E/S.

Como o processo de snapshot continua em segundo plano e a criação do snapshot é de rápida execução e captura um momento, os volumes dos quais você está fazendo backup só precisam ser desmontados por alguns segundos. Como a janela de backup é a menor possível, o tempo de interrupção é previsível e pode ser programado.

---

[4] https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Storage_Administration_Guide/xfsfreeze.html

Como fazer backups de vários volumes

Em alguns casos, é possível segmentar os dados entre vários volumes do Amazon EBS usando um gerenciador de volumes lógicos para aumentar a taxa de transferência em potencial. Quando você usa um gerenciador de volumes lógicos (por exemplo, mdadm ou LVM), é importante realizar o backup a partir da camada do gerenciador de volumes, não dos volumes de EBS subjacentes. Isso garante a consistência entre todos os metadados e a coerência dos volumes de subcomponentes.

Há várias maneiras de fazer isso. Por exemplo, é possível usar o script criado por alestic.com[5]. Os buffers de memória devem ser descarregados no disco; a E/S do sistema de arquivos para o disco deve ser interrompida, e um snapshot deve ser iniciado simultaneamente para todos os volumes que constituem o conjunto de RAID. Depois de iniciar o snapshot dos volumes (geralmente um segundo ou dois), o sistema de arquivos pode continuar com as operações. Os snapshots devem ser marcados para que você possa gerenciá-los coletivamente durante uma restauração.

Você também pode realizar esses backups a partir do gerenciador de volumes lógicos ou do nível do sistema de arquivos. Nesses casos, o uso de um agente de backup tradicional permite que o backup dos dados seja feito pela rede. Várias soluções de backup baseadas em agente estão disponíveis na Internet e no AWS Marketplace.[6] Lembre-se de que o software de backup baseado em agente conta com um nome de servidor e endereço IP consistentes. Como resultado, o uso dessas ferramentas com instâncias implantadas em uma Amazon Virtual Private Cloud (VPC)[7] é a melhor maneira de garantir a confiabilidade.

Uma abordagem alternativa é criar uma réplica dos volumes existentes do sistema primário em um único volume grande. Isso simplifica o processo de backup porque é necessário fazer backup apenas de um volume grande, e tal backup não ocorre no sistema primário. Entretanto, primeiro você deve determinar se o volume único é capaz de operar de maneira suficiente durante o backup e se o tamanho máximo do volume é adequado ao aplicativo.

---

[5] https://github.com/alestic/ec2-consistent-snapshot

[6] https://aws.amazon.com/marketplace/

[7] http://aws.amazon.com/vpc/

Abordagens de backup de banco de dados

A AWS possui muitas opções para bancos de dados. Você pode operar seu próprio banco de dados em uma instância do EC2 ou usar uma das opções de banco de dados de serviços gerenciados fornecidas pelo Amazon Relational Database Service [8](Amazon RDS). Se estiver operando seu próprio banco de dados em uma instância do EC2, você poderá fazer backup dos dados em arquivos usando ferramentas nativas (por exemplo, MySQL[9], Oracle[10], MSSQL[11], PostgreSQL[12]) ou criar um snapshot dos volumes que contêm os dados usando um dos métodos descritos em "Proteção baseada em snapshot do EBS."

Uso de backups de réplica de banco de dados

Para bancos de dados criados em conjuntos de RAID de volumes do Amazon EBS, é possível remover a carga de backups no banco de dados principal criando uma réplica de leitura do banco de dados. Esta é uma cópia atualizada do banco de dados executada em uma instância separada do Amazon EC2. A instância da réplica do banco de dados pode ser criada com o uso de vários discos semelhantes à origem ou os dados podem ser consolidados em um único volume de EBS. Você pode seguir um dos procedimentos descritos em "Proteção baseada em snapshot do EBS" para criar um snapshot dos volumes de EBS. Essa abordagem é usada com frequência para grandes bancos de dados que precisam ser executados 24/7. Quando for esse o caso, a janela de backup exigida será muito longa, e o banco de dados de produção não poderá ser paralisado por períodos tão longos.

Uso do Amazon RDS para backups

O Amazon RDS inclui recursos para automatizar backups de bancos de dados. O Amazon RDS cria um snapshot do volume de armazenamento de sua instância de banco de dados, fazendo o backup de toda a instância do DB, não apenas dos bancos de dados individuais.

---

[8] https://aws.amazon.com/rds/

[9] http://dev.mysql.com/doc/refman/5.7/en/backup-and-recovery.html

[10] http://docs.oracle.com/cd/E11882_01/backup.112/e10642/rcmbckba.htm#BRADV8003

[11] http://msdn.microsoft.com/en-us/library/ms187510.aspx

[12] http://www.postgresql.org/docs/9.3/static/backup.html

O Amazon RDS oferece dois diferentes métodos para backup e restauração de suas instâncias de banco de dados:

- Os **backups automáticos** permitem a recuperação de um momento de sua instância de banco de dados. Os backups automatizados são ativados por padrão quando você cria uma nova instância de banco de dados. O Amazon RDS executa um backup diário completo de seus dados durante uma janela que você define quando cria a instância de banco de dados. Você pode configurar um período de retenção de até 35 dias para o backup automático. O Amazon RDS utiliza esses backups de dados periódicos em conjunto com seus logs de transação para permitir que você restaure sua instância de banco de dados para qualquer segundo durante seu período de retenção, até o `LatestRestorableTime` (geralmente até os últimos cinco minutos). Para localizar o último momento restaurável das suas instâncias de banco de dados, é possível usar a chamada de API `DescribeDBInstances` ou verificar a guia **Descrição** do banco de dados no Console de Gerenciamento da AWS.

  Quando você inicia a recuperação de um momento, os logs de transação são aplicados ao backup diário mais apropriado a fim de restaurar sua instância de banco de dados para o momento solicitado.

- **Snapshots de banco de dados** são backups iniciados pelo usuário que permitem fazer backup da sua instância de banco de dados em um estado conhecido e com a frequência desejada, para depois restaurá-la àquele estado quando quiser. É possível usar o Console de Gerenciamento da AWS ou a chamada de API `CreateDBSnapshot` para criar snapshots de banco de dados. Esses snapshots têm retenção ilimitada. Eles são mantidos até que você use o console ou a chamada de API `DeleteDBSnapshot` para excluí-los explicitamente.

Quando você restaura um banco de dados para um momento ou a partir de um DB snapshot, uma nova instância de banco de dados é criada com um novo endpoint. Dessa maneira, é possível criar várias instâncias de banco de dados a partir de um DB snapshot ou um momento específico.

Você pode usar o Console de Gerenciamento da AWS ou uma chamada `DeleteDBInstance` para excluir a antiga instância de banco de dados.

Uso da AMI para fazer backup das instâncias do EC2

A AWS armazena imagens do sistema nas chamadas Imagens de máquina da Amazon (AMIs). Essas imagens consistem no modelo do volume raiz necessário para lançar uma instância. É possível usar o Console de Gerenciamento da AWS ou o comando CLI `aws ec2 create-image` para fazer backup do volume raiz como uma AMI.

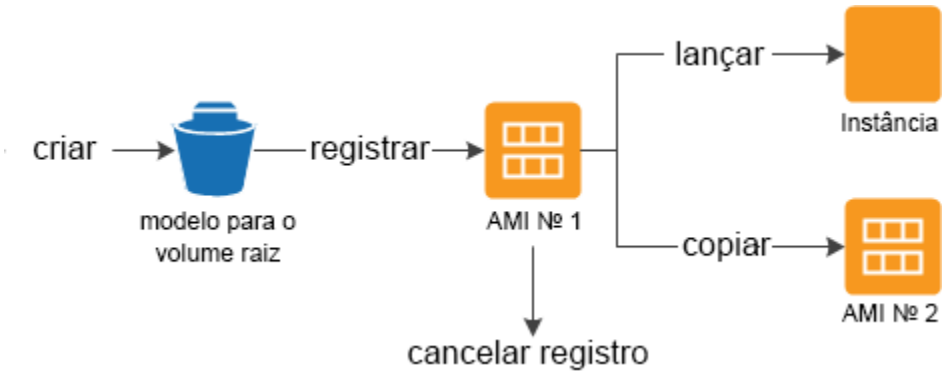

Figura 3: Uso de uma AMI para fazer backup e lançar uma instância

Quando você registra uma AMI, ela é armazenada automaticamente na sua conta usando snapshots do Amazon EBS. Esses snapshots residem no Amazon S3 e são resilientes.

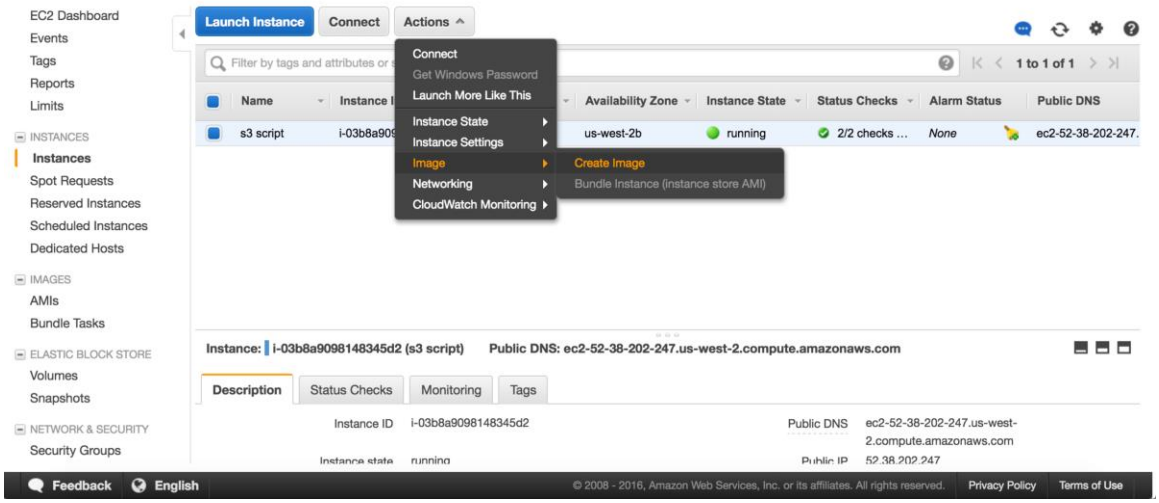

Figura 4: Uso do console EC2 para criar uma imagem de máquina

Depois de criar uma AMI de sua instância do Amazon EC2, é possível usar a AMI para recriar a instância ou executar mais cópias dela. Também é possível copiar AMIs de uma região para outra para migração de aplicativo ou recuperação de desastres.

Infraestrutura no local para a AWS

Este cenário descreve um ambiente de carga de trabalho sem componentes na nuvem. Todos os recursos, incluindo servidores da Web, servidores de aplicativos, servidores de monitoramento, bancos de dados, Active Directory etc. ficam hospedados no datacenter do cliente ou por meio de co-locação.

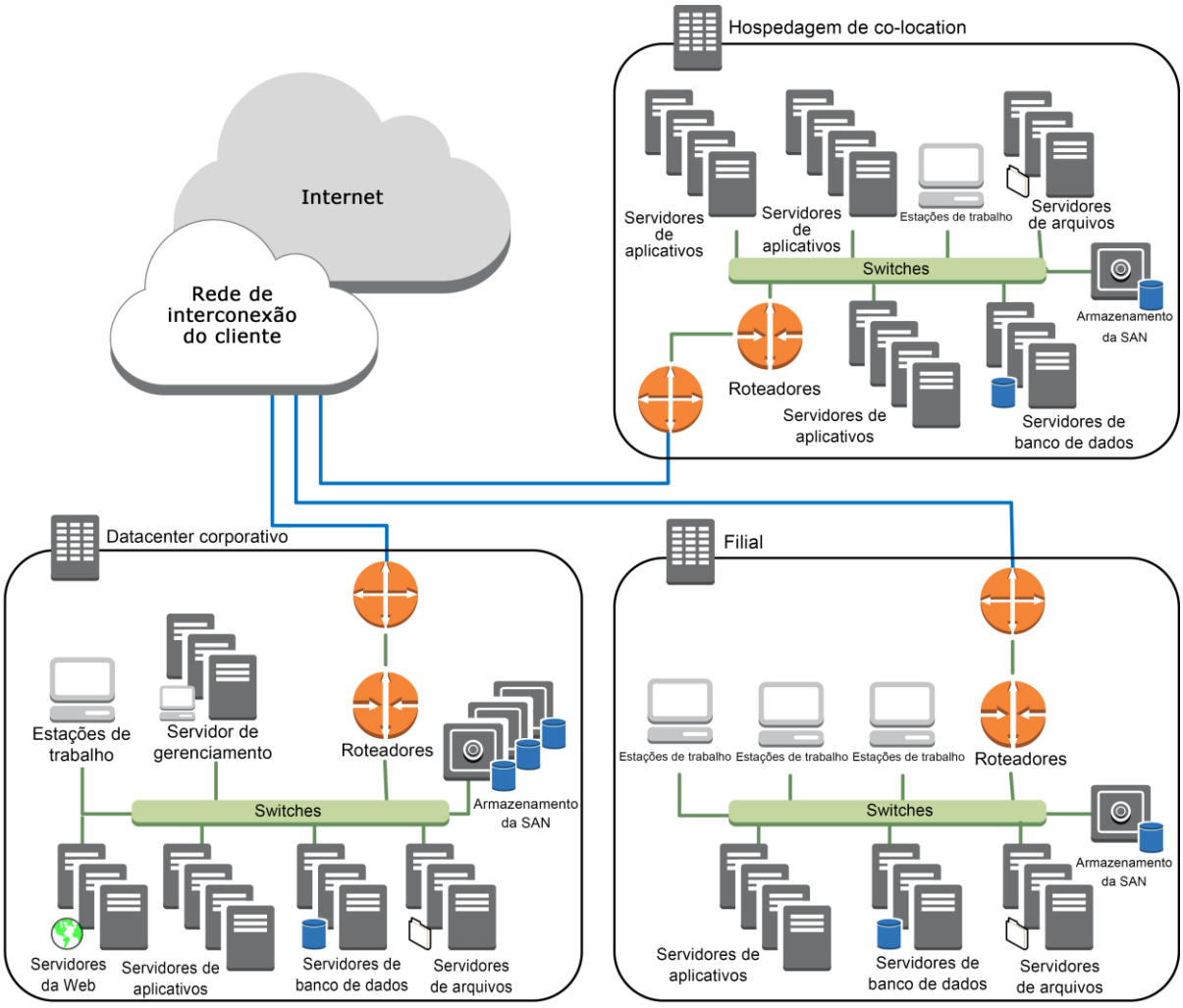

Figura 5: Ambiente no local

Ao usar os serviços de armazenamento da AWS neste cenário, você pode se concentrar nas tarefas de backup e arquivamento. Você não precisa se preocupar com a escalabilidade do armazenamento ou a capacidade da infraestrutura para realizar a tarefa de backup.

Amazon S3 e Amazon Glacier são nativamente baseados em API e estão disponíveis pela Internet. Isso permite que os fornecedores de softwares de backup integrem diretamente seus aplicativos às soluções de armazenamento da AWS, como mostrado na figura a seguir.

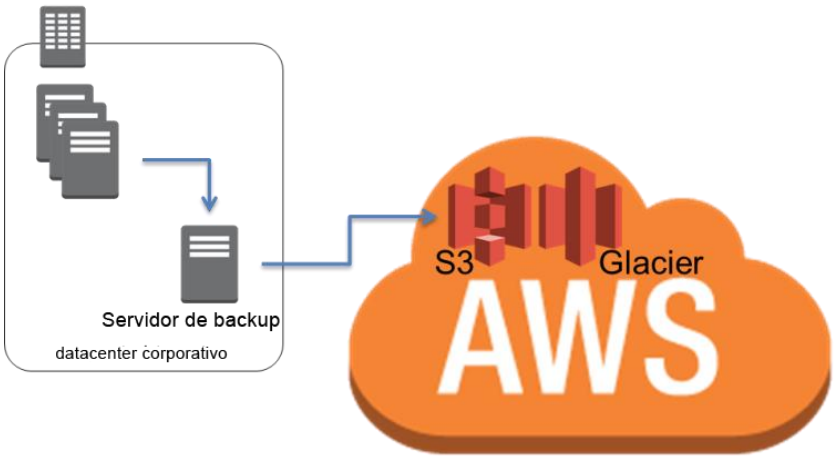

Figura 6: Conector de backup para Amazon S3 ou Amazon Glacier

Neste cenário, o software de backup e arquivamento que faz interface direta com a AWS por meio das APIs. Como o software de backup considera a AWS, ele fará backup dos dados dos servidores no local diretamente no Amazon S3 ou no Amazon Glacier.

Se seu software de backup atual não for nativamente compatível com a Nuvem AWS, você poderá usar os produtos AWS Storage Gateway. AWS Storage Gateway[13] é um dispositivo virtual que fornece integração uniforme e segura entre seu datacenter e a infraestrutura de armazenamento da AWS. O serviço permite armazenar dados com segurança na Nuvem AWS para um armazenamento escalável e econômico. O Storage Gateway é compatível com os protocolos de armazenamento padrão do setor que funcionam com seus aplicativos existentes enquanto armazenam com segurança todos os seus dados criptografados no Amazon S3 ou no Amazon Glacier.

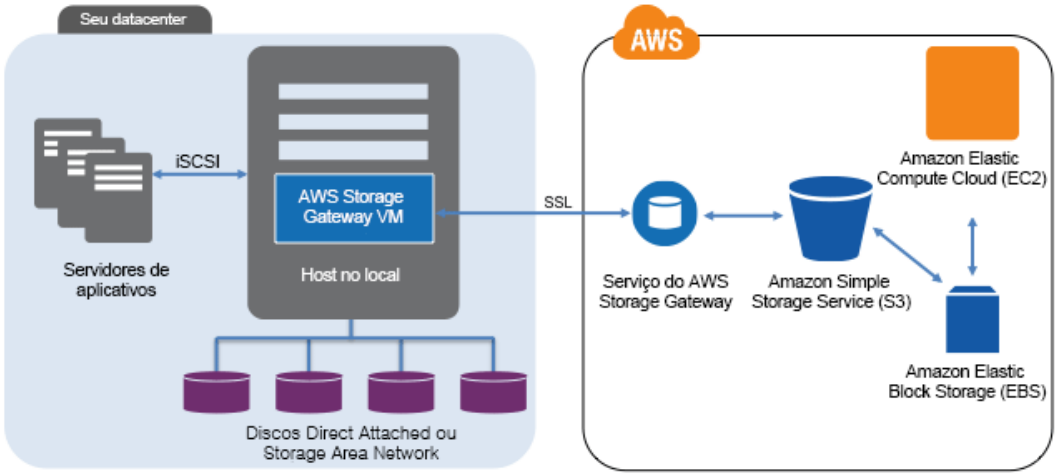

Figura 7: Conexão no local com o armazenamento da AWS

O AWS Storage Gateway oferece suporte às seguintes configurações:

- **Gateways de volume:** Os gateways de volume fornecem volumes de armazenamento de backup em nuvem que você pode montar como dispositivos Internet Small Computer System Interface (iSCSI) a partir de seus servidores de aplicativos no local. O gateway oferece suporte às seguintes configurações de volume:
  - **Volumes em cache no gateway:** Você pode armazenar os dados principais no Amazon S3 e reter localmente os dados acessados com frequência. Os volumes em cache no gateway proporcionam economia substancial no armazenamento principal, minimizam a necessidade de escalar seu armazenamento no local e retêm acesso de baixa latência aos seus dados acessados com frequência.

[13] http://aws.amazon.com/storagegateway/

  - **Volumes armazenados no gateway:** Se você precisar de acesso de baixa latência ao seu conjunto de dados inteiro, poderá configurar seu gateway de dados no local para armazenar seus dados principais localmente e fazer backup, de maneira assíncrona, de snapshots de momento desses dados no Amazon S3. Os volumes armazenados no gateway fornecem backups externos duráveis e baratos que você pode recuperar localmente ou do Amazon EC2.
- **Virtual Tape Library (VTL, Biblioteca de fitas virtuais) do gateway:** Com o Gateway-VTL, você pode ter uma coleção ilimitada de fitas virtuais. Cada fita virtual pode ser armazenada em uma biblioteca de fitas virtuais com backup feito pelo Amazon S3 ou uma prateleira de fitas virtuais com backup feito pelo Amazon Glacier. A biblioteca de fitas virtuais apresenta uma interface iSCSI padrão do setor que proporciona ao seu aplicativo de backup acesso online às fitas virtuais. Quando você não precisar mais de acesso imediato ou frequente aos dados contidos em uma fita virtual, poderá usar o seu aplicativo de backup para movê-los da biblioteca de fitas virtuais para a prateleira de fitas virtuais a fim de reduzir ainda mais os seus custos de armazenamento.

Esses gateways atuam como dispositivos plug-and-play que fornecem dispositivos iSCSI padrão capazes de se integrar à sua estrutura de backup ou arquivamento. Você pode usar os dispositivos de disco iSCSI como pools de armazenamento para o seu software de backup ou o gateway-VTL para descarregar o backup ou arquivamento baseado em fita diretamente no Amazon S3 ou no Amazon Glacier.

Com esse método, seus backups e arquivamentos são feitos fora do local automaticamente (para fins de conformidade) e armazenados em mídia durável, eliminando a complexidade e os riscos de segurança do gerenciamento de fitas externo.

Ambientes híbridos

As duas implantações de infraestrutura discutidas até agora, nativa de nuvem e no local, podem ser combinadas em um cenário híbrido em que o ambiente de carga de trabalho possui componentes de infraestrutura na AWS e no local. Os recursos, incluindo servidores da Web, servidores de aplicativos, servidores de monitoramento, bancos de dados, Active Directory etc. ficam hospedados no datacenter do cliente ou na AWS. Os aplicativos executados na Nuvem AWS são conectados aos aplicativos executados no local.

Esse cenário está se tornando comum para cargas de trabalho empresariais. Muitas empresas possuem datacenters próprios e usam a AWS para aumentar a capacidade. Esses datacenters de cliente frequentemente são conectados à rede da AWS por links de rede de alta capacidade. Por exemplo, com o [AWS Direct Connect][14], é possível estabelecer conectividade privada dedicada entre seu local e a AWS. Isso oferece largura de banda e latência consistente para carregar dados na nuvem para fins de proteção de dados e desempenho consistente, além de latência para cargas de trabalho híbridas.

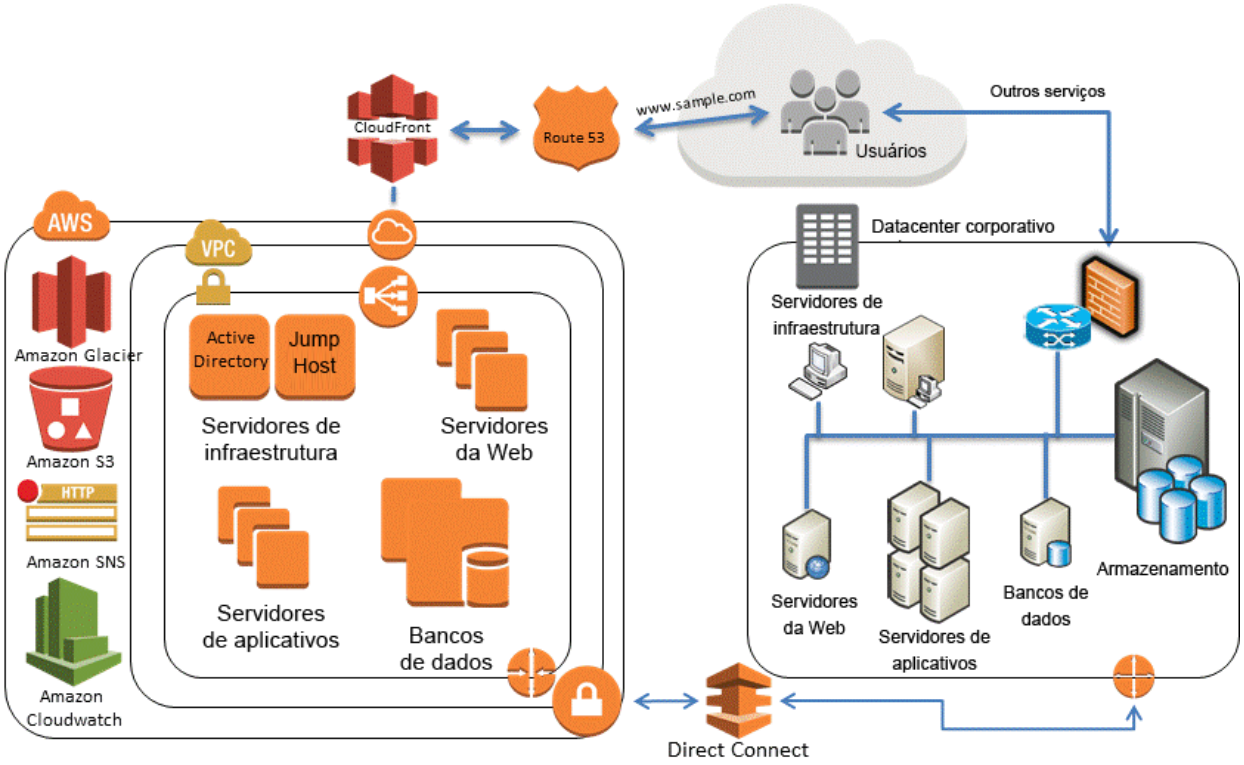

Figura 8: Um cenário de infraestrutura híbrida

As soluções de proteção de dados bem projetadas usam uma combinação dos métodos descritos nas soluções nativas de nuvem e no local.

[14] http://aws.amazon.com/directconnect/

Backup de aplicativos baseados na AWS para o datacenter

Se você já tem uma estrutura que faz backup de dados dos seus servidores no local, será fácil estendê-la aos recursos da AWS por uma conexão VPN ou por meio do AWS Direct Connect. É possível instalar o agente de backup nas instâncias do Amazon EC2 e fazer backup delas de acordo com suas políticas de proteção de dados.

Migração do gerenciamento de backup para a nuvem para disponibilidade

Dependendo da arquitetura de backup, você pode ter um servidor mestre de backup e um ou mais servidores de mídia ou armazenamento localizados no local com os serviços protegidos. Nesse caso, convém mover o servidor mestre de backup para uma instância do Amazon EC2 para protegê-lo contra desastres no local e viabilizar um infraestrutura de backup de alta disponibilidade.

Para gerenciar os fluxos de dados de backup, convém também criar um ou mais servidores de mídia nas instâncias do Amazon EC2. Os servidores de mídia próximos às instâncias do Amazon EC2 proporcionarão economia na transferência pela Internet e, ao fazer backup no S3 ou no Amazon Glacier, aumentarão o desempenho geral de backup e recuperação.

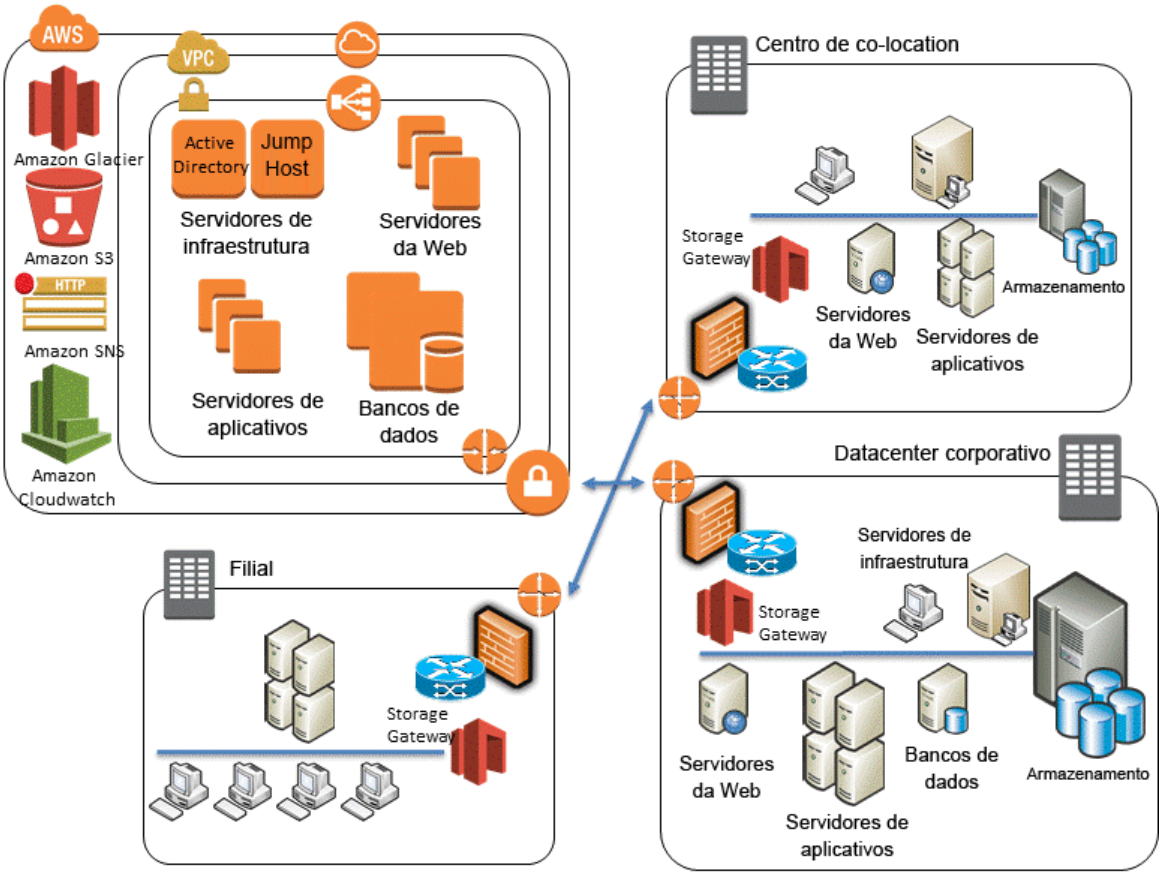

Figura 9: Uso de gateways no cenário híbrido

Exemplo de cenário híbrido

Imagine que você está gerenciando um ambiente em que faz backup de instâncias do Amazon EC2, servidores independentes, máquinas virtuais e bancos de dados. Esse ambiente possui 1.000 servidores, e você faz backup do sistema operacional, dados de arquivos, imagens de máquinas virtuais e bancos de dados. Há 20 bancos de dados (uma combinação de MySQL, Microsoft SQL Server e Oracle) para backup.

Seu software de backup tem agentes que fazem backup de sistemas operacionais, imagens de máquinas virtuais, volumes de dados, bancos de dados do SQL Server e bancos de dados Oracle (usando RMAN). No caso de aplicativos como o MySQL, para os quais seu software de backup não tem um agente, você pode usar o utilitário cliente mysqldump para criar um arquivo de despejo de banco de dados no disco em que os agentes de backup padrão podem proteger os dados.

Para proteger esse ambiente, muito provavelmente seu software de backup de terceiros tem um servidor de catálogo global ou um servidor mestre que controla as atividades de backup, arquivamento e restauração, bem como vários servidores de mídia conectados ao armazenamento baseado em disco, unidades de fita Linear Tape-Open (LTO) e serviços de armazenamento da AWS.

A maneira mais simples de ampliar sua solução de backup com os serviços de armazenamento da AWS é aproveitar seu suporte de fornecedor de backup para o Amazon S3 ou o Amazon Glacier. Sugerimos entrar em contato com seu fornecedor para entender suas opções de integração e de conector. Para obter uma lista de fornecedores de software de backup que trabalham com a AWS, consulte nosso diretório de parceiros[15].

Se seu software de backup atual não oferece suporte nativo ao armazenamento em nuvem para backup ou arquivamento, você pode usar um dispositivo de Storage Gateway como uma ponte entre o software de backup e o Amazon S3 ou o Amazon Glacier.

Há muitas soluções de gateway de terceiros. Também é possível usar dispositivos virtuais do AWS Storage Gateway para preencher essa lacuna porque eles usam técnicas genéricas, como volumes baseados em iSCSI e bibliotecas de fitas virtuais (VTLs). Essa configuração requer um hipervisor compatível (VMware ou Microsoft Hyper-V) e armazenamento local para hospedar o dispositivo.

Arquivamento de dados com a AWS

Quando for necessário preservar dados por motivos de conformidade ou empresariais, arquive-os. Diferentemente de backups, que normalmente são executados para manter uma cópia dos dados de produção por um breve período a fim de recuperar dados corrompidos ou perdidos, o arquivamento mantém todas as cópias dos dados até que a política de retenção expire.

Um bom arquivamento atende aos seguintes critérios:

- Durabilidade de dados para integridade em longo prazo

[15] http://www.aws-partner-directory.com/PartnerDirectory/PartnerSearch?type=ISV

- Segurança dos dados
- Facilidade de recuperação
- Baixo custo

Os armazenamentos de dados imutáveis podem ser outro requisito regulatório ou de conformidade.

O Amazon Glacier oferece arquivamentos baratos, criptografia nativa de dados em repouso, 99,999999999% de durabilidade e capacidade ilimitada.

Amazon S3 Standard - O acesso pouco frequente é uma boa opção para casos de uso que requerem uma recuperação rápida dos dados. O Amazon Glacier é uma boa opção para casos de uso em que os dados raramente são acessados e para os quais é aceitável ter tempos de recuperação de várias horas.

Os objetos podem ser escalonados no Amazon Glacier por meio de regras de ciclo de vida na API do S3 ou do Amazon Glacier. O recurso Amazon Glacier Vault Lock permite implantar e aplicar facilmente os controles de conformidade para cofres individuais do Amazon Glacier com uma política de vault lock. É possível especificar controles, como “uma gravação e muitas leituras” (WORM, write once, read many) em uma política de vault lock e bloquear a política contra edições futuras. Para obter mais informações, consulte Amazon Glacier.

Proteção de dados de backup na AWS

A segurança de dados é uma preocupação comum. A AWS trata a segurança com muita seriedade. Ela é a base de todos os serviços que lançamos. Os serviços de armazenamento, como o Amazon S3, oferecem recursos poderosos de controle de acesso e criptografia para dados em repouso e em trânsito. Todos os endpoints de API do Amazon S3 e do Amazon Glacier oferecem suporte à criptografia SSL para dados em trânsito. O Amazon Glacier criptografa todos os dados em repouso por padrão. Com o Amazon S3, os clientes podem optar pela criptografia no lado do servidor para objetos em repouso permitindo que a AWS gerencie as chaves de criptografia, fornecendo suas próprias chaves quando eles carregam um objeto ou usando a integração do AWS Key Management Service (AWS KMS) [16]para as chaves de criptografia. Como alternativa, os clientes podem sempre criptografar os dados antes de carregá-los na AWS. Para obter mais informações, consulte Amazon Web Services: Visão geral dos processos de segurança.

[16] http://docs.aws.amazon.com/AmazonS3/latest/dev/UsingKMSEncryption.html

Conclusão

A Gartner reconheceu a AWS como líder em serviços de armazenamento em nuvem pública[17]. A AWS tem tudo para ajudar as organizações a mover suas cargas de trabalho para plataformas baseadas na nuvem, o backup de última geração. A AWS fornece soluções econômicas e escaláveis para ajudar as organizações a equilibrar seus requisitos de backup e arquivamento. Esses serviços têm uma boa integração com as tecnologias que você usa atualmente.

Colaboradores

As seguintes pessoas contribuíram na elaboração deste documento:

- Pawan Agnihotri, arquiteto de soluções, Amazon Web Services
- Lee Kear, arquiteto de soluções, Amazon Web Services
- Peter Levett, arquiteto de soluções, Amazon Web Services

Revisões do documento

Atualizado em maio de 2016

[17] http://www.gartner.com/technology/reprints.do?id=1-1WWKTQ3&ct=140709&st=sb