O caminho para quem quer começar em Computação em Nuvem - Foco em AWS

Começe agora

COMPUTAÇÃO EM NUVEM COM FOCO EM AWS

Pensado para quem quer começar do Zero

GUILHERME TELES

Você vai ver como não é tão complicado quanto parece

Ano 2020

O caminho para quem quer começar em Computação em Nuvem - Foco em AWS

GUILHERME TELES

Você vai ver como não é tão complicado quanto parece

Ano 2020

ISBN: 978-65-00-01323-8

Contents

Primeiros Passos

O que é a Computação em Nuvem

A computação em nuvem é a entrega sob demanda de recursos e aplicativos de TI via Internet com preços pagos conforme o uso.

Quer você execute aplicativos que compartilhem fotos com milhões de usuários móveis ou forneça serviços que suportam as operações críticas de seus negócios, a nuvem fornece acesso rápido a recursos de TI flexíveis e de baixo custo.

Com a computação em nuvem, você não precisa fazer grandes investimentos iniciais em hardware e gastar muito tempo em gerenciamento.

Em vez disso, você pode fornecer exatamente o tipo e tamanho certos de recursos de computação necessários para alimentar sua mais nova ideia ou operar seu departamento de tecnologia.

Com a computação em nuvem, você pode acessar quantos recursos precisar, quase instantaneamente, e pagar apenas pelo que usar.

Em sua forma mais simples, a computação em nuvem fornece uma maneira fácil de acessar servidores, armazenamento, bancos de dados e um amplo conjunto de serviços de aplicativos pela Internet.

Os provedores de computação em nuvem, como a AWS, possuem e mantêm o hardware conectado à rede necessário para esses serviços de aplicativos, enquanto você provisiona e usa o que precisa para suas cargas de trabalho.

Vantagens da computação em nuvem

A computação em nuvem introduz uma mudança revolucionária na maneira como a tecnologia é obtida, usada e gerenciada, e na forma como as organizações fazem o orçamento e pagam pelos serviços de tecnologia.

Com a capacidade de reconfigurar o ambiente de computação rapidamente para se adaptar às mudanças nos requisitos de negócios, as organizações podem otimizar os seus respectivos gastos.

A capacidade pode ser aumentada ou diminuída automaticamente para atender a padrões de uso flutuantes. Os serviços podem ser temporariamente desligados ou desligados permanentemente, conforme as demandas da empresa.

Além disso, com o faturamento de pagamento por uso, os serviços da nuvem tornam-se uma despesa operacional em vez de uma despesa de capital.

Despesa variável vs. despesa de capital

Vamos começar com a capacidade de negociar despesa de capital por despesa operacional variável.

Em vez de ter que investir pesadamente em data centers e servidores antes de saber como usá-los, você pode pagar apenas quando consumir recursos de computação e pagar apenas por quanto consome.

Economias de escala

Outra vantagem da computação em nuvem é que as organizações se beneficiam de enormes economias de escala. Ao usar a computação em nuvem, você pode obter um custo variável mais baixo do que conseguiria por conta própria.

Como o uso de centenas de milhares de clientes é agregado na nuvem, fornecedores como a AWS podem obter maiores economias de escala, o que se traduz em preços mais baixos.

Pare de adivinhar a capacidade

Quando você toma uma decisão de tentar prever a capacidade antes de implantar um aplicativo, geralmente acaba sentado em recursos ociosos caros ou lidando com capacidade limitada.

Com a computação em nuvem, as organizações podem parar de adivinhar os requisitos de capacidade da infraestrutura necessária para atender às suas necessidades de negócios.

Eles podem acessar o quanto precisam e aumentar ou diminuir conforme necessário com apenas alguns minutos de antecedência.

Aumente a velocidade e a agilidade

Em um ambiente de computação em nuvem, os novos recursos de TI estão a um clique de distância, o que permite que as organizações reduzam o tempo necessário para disponibilizar esses recursos aos desenvolvedores de semanas para apenas alguns minutos.

Isso resulta em um aumento dramático na velocidade e agilidade para a organização, porque o custo e o tempo necessários para experimentar e desenvolver são significativamente menores.

Foco nos diferenciadores de negócios

A computação em nuvem permite que as organizações se concentrem em suas prioridades de negócios, em vez de no trabalho pesado de servidores de rack, empilhamento e alimentação. Ao adotar essa mudança de paradigma, as organizações podem parar de gastar dinheiro na execução e manutenção de data centers.

Isso permite que as organizações se concentrem em projetos que diferenciam seus negócios, como focar apenas em inovação.

Global em minutos

As organizações podem implantar facilmente seus aplicativos em vários locais ao redor do mundo com apenas alguns cliques. Isso permite que as organizações forneçam redundância em todo o mundo e proporcionem menor latência e melhores experiências aos seus clientes a um custo mínimo.

A globalização costumava ser algo que apenas as maiores empresas podiam se dar ao luxo de fazer, mas a computação em nuvem democratiza essa capacidade, tornando possível para qualquer organização.

Embora questões específicas sobre essas vantagens da computação em nuvem não sejam passíveis em exames de certificação, a exposição a esses benefícios pode ajudar a racionalizar as respostas apropriadas.

Modelos de implantação de computação em nuvem

Os dois principais modelos de implantação de computação em nuvem atualmente são implantações "all-in" baseadas em nuvem e implantações híbridas. É importante entender como cada estratégia se aplica às opções e decisões de arquitetura.

Um aplicativo completo baseado na nuvem é totalmente implantado na nuvem, com todos os componentes em execução na nuvem. Os aplicativos na nuvem foram criados na nuvem ou migrados de uma infraestrutura existente para aproveitar os benefícios de computação em nuvem.

Os aplicativos baseados em nuvem podem ser construídos em partes de infraestrutura de baixo nível ou podem usar serviços de nível superior que fornecem abstração dos requisitos de gerenciamento, arquitetura e dimensionamento da infraestrutura principal.

Uma implantação híbrida é uma abordagem comum adotada por muitas empresas que conecta infraestrutura e aplicativos entre recursos baseados na nuvem e recursos existentes, geralmente em um data center existente.

O método mais comum de implantação híbrida é entre a nuvem e a infraestrutura local existente para estender e aumentar a infraestrutura da organização enquanto conecta recursos da nuvem a sistemas internos.

A escolha entre um investimento existente em infraestrutura e a migração para a nuvem não precisa ser uma decisão binária.

A alavancagem da conectividade dedicada, federação de identidades e ferramentas integradas permite que as organizações executem aplicativos híbridos nos serviços locais e na nuvem.

Fundamentos da AWS

Na sua essência, a AWS fornece entrega sob demanda de recursos de tecnologia via Internet em uma plataforma segura de serviços em nuvem, oferecendo energia computacional, armazenamento, bancos de dados, entrega de conteúdo e outras funcionalidades para ajudar as empresas a escalar e crescer.

Usar os recursos da AWS em vez dos seus é como comprar eletricidade de uma empresa de energia em vez de operar seu próprio gerador e fornece as principais vantagens da computação em nuvem.

A capacidade corresponde exatamente à sua necessidade, você paga apenas pelo que usa, resultando em economias de escala em custos mais baixos, e o serviço é fornecido por um fornecedor com experiência na execução de redes de grande escala.

Infraestrutura global

A AWS atende a mais de um milhão de clientes ativos em mais de 190 países e continua a expandir sua infraestrutura global de maneira constante para ajudar as organizações a obter menor latência e maior produtividade para suas necessidades de negócios.

A AWS fornece uma plataforma de infraestrutura de tecnologia altamente disponível com vários locais em todo o mundo. Esses locais são compostos de regiões e zonas de disponibilidade. Cada região é uma área geográfica separada. Cada região possui vários locais isolados, conhecidos como zonas de disponibilidade. A AWS permite a colocação de recursos e dados em vários locais.

Os recursos não são replicados nas regiões, a menos que as organizações optem por fazê-lo.

Cada região é completamente independente e foi projetada para ser completamente isolada das outras regiões. Isso alcança a maior tolerância a falhas e estabilidade possíveis.

Cada zona de disponibilidade também é isolada, mas as zonas de disponibilidade em uma região são conectadas por meio de links de baixa latência. As zonas de disponibilidade são fisicamente separadas dentro de uma região metropolitana típica e localizadas em planícies de risco de menor risco (a categorização específica da zona de inundação varia de acordo com a região).

Além de usar uma fonte de alimentação ininterrupta (UPS) e geradores de backup no local, eles são alimentados por diferentes grades de utilitários independentes (quando disponíveis) para reduzir ainda mais os pontos únicos de falha.

As zonas de disponibilidade são todas redundantemente conectadas a vários provedores de transporte. Colocando recursos em zonas de disponibilidade separadas, você pode proteger seu site ou aplicativo contra uma interrupção no serviço que afeta um único local.

Você pode obter alta disponibilidade implantando seu aplicativo em várias zonas de disponibilidade. Instâncias redundantes para cada camada (por exemplo, web, aplicativo e banco de dados) de um aplicativo devem ser colocadas em zonas de disponibilidade distintas, criando assim uma solução multi-site.

No mínimo, o objetivo é ter uma cópia independente do cada aplicativo é empilhado em duas ou mais zonas de disponibilidade.

Segurança e conformidade

Seja no seu data center ou na nuvem, a segurança das informações é de suma importância para as organizações que executam cargas de trabalho críticas.

A segurança é um requisito funcional essencial que protege as informações críticas da missão contra roubo acidental ou deliberado, vazamento, comprometimento da integridade e exclusão.

Ajudar a proteger a confidencialidade, integridade e disponibilidade de sistemas e dados é da maior importância para a AWS, assim como mantém sua confiança em clientes e investidores.

Segurança

Se você já teve algum contato com a AWS, já deve ter ouvido, caso contrário, já guarde essa frase que é sempre reforçada:

A segurança da nuvem na AWS é a prioridade número um.

Todos os clientes da AWS se beneficiam das arquiteturas de data center e de rede criadas para atender aos requisitos das organizações mais sensíveis à segurança.

A AWS e seus parceiros oferecem centenas de ferramentas e recursos para ajudar as organizações a atingir seus objetivos de segurança em termos de visibilidade, auditabilidade, controlabilidade e agilidade.

Isso significa que as organizações podem ter a segurança de que precisam, mas sem os gastos de capital e com sobrecarga operacional muito menor do que em um ambiente local.

As organizações que utilizam a AWS herdam todas as melhores práticas de políticas, arquitetura e processos operacionais da AWS criados para atender aos requisitos dos clientes mais sensíveis à segurança. A infraestrutura da AWS foi projetada para fornecer a mais alta disponibilidade e, ao mesmo tempo, proteger fortemente a privacidade e a segregação do cliente.

Ao implantar sistemas na plataforma de computação em nuvem da AWS, a AWS ajuda compartilhando as responsabilidades de segurança com a organização. A AWS gerencia a infraestrutura subjacente e a organização pode proteger qualquer coisa que implantar na AWS. Isso proporciona a cada organização a flexibilidade e agilidade que eles precisam nos controles de segurança.

Essa infraestrutura é construída e gerenciada não apenas de acordo com as melhores práticas e padrões de segurança, mas também com as necessidades exclusivas da nuvem. A AWS usa controles redundantes e em camadas, com validações e testes contínuos e uma quantidade substancial de automação para garantir que a infraestrutura subjacente seja monitorada e protegida 24/7.

A AWS garante que esses controles sejam aplicados de maneira consistente em todos os novos data centers ou serviços.

Conformidade

Quando os clientes transferem suas cargas de trabalho de produção para a Nuvem AWS, ambas as partes se tornam responsáveis pelo gerenciamento do ambiente de TI.

Os clientes são responsáveis por configurar seu ambiente de maneira segura e controlada. Também precisam manter uma governança adequada em todo o ambiente de controle de TI. Ao unir o foco em governança, recursos de serviço compatíveis com a auditoria, com conformidade ou padrões de auditoria aplicáveis, a AWS permite que os clientes desenvolvam programas de conformidade tradicionais.

Isso ajuda as organizações a estabelecer e operar em um ambiente de controle de segurança da AWS.

As organizações mantêm controle e propriedade completos sobre a região em que seus dados estão localizados fisicamente, permitindo que eles atendam aos requisitos regionais de conformidade e residência de dados.

A infraestrutura de tecnologia que a AWS fornece às organizações é projetada e gerenciada em alinhamento com as melhores práticas de segurança e uma variedade de padrões de segurança de TI. A seguir, é apresentada uma lista parcial das muitas certificações e padrões com os quais a AWS está em conformidade:

- Controles da organização de serviços (SOC) 1 / Norma internacional sobre garantia de Compromissos (ISAE) 3402, SOC 2 e SOC 3
- Lei Federal de Gerenciamento de Segurança da Informação (FISMA)
- Processo de Certificação e Acreditação de Garantia da Informação do Departamento de Defesa (DIACAP)
- Programa Federal de Gerenciamento de Riscos e Autorizações (FedRAMP)
- Padrão de segurança de dados da indústria de cartões (PCI DSS) Nível 1
- Organização Internacional de Normalização (ISO) 9001, ISO 27001 e ISO 27018

A AWS fornece uma ampla gama de informações sobre seu ambiente de controle de TI para ajudar as organizações a cumprir compromissos regulatórios na forma de relatórios, certificações, acreditações e outros atestados de terceiros.

Plataforma de Computação em Nuvem da AWS

A AWS fornece muitos serviços de nuvem que você pode combinar para atender às necessidades de negócios. Embora o conhecimento de todos os serviços da plataforma permita que você seja um arquiteto de soluções completo, entender os serviços e conceitos fundamentais descritos aqui ajudará você bom um excelente entendimento para o CORE dos exames Associates.

Esta seção apresenta os principais serviços da AWS por categoria.

Acessando a plataforma

Para acessar os serviços da AWS, você pode usar o AWS Management Console, o AWS Command Line Interface (CLI) ou os AWS Software Development Kits (SDKs).

O AWS Management Console é um aplicativo da web para gerenciar serviços da AWS. O console fornece uma interface de usuário intuitiva para executar muitas tarefas. Cada serviço possui seu próprio console, que pode ser acessado no AWS Management Console.

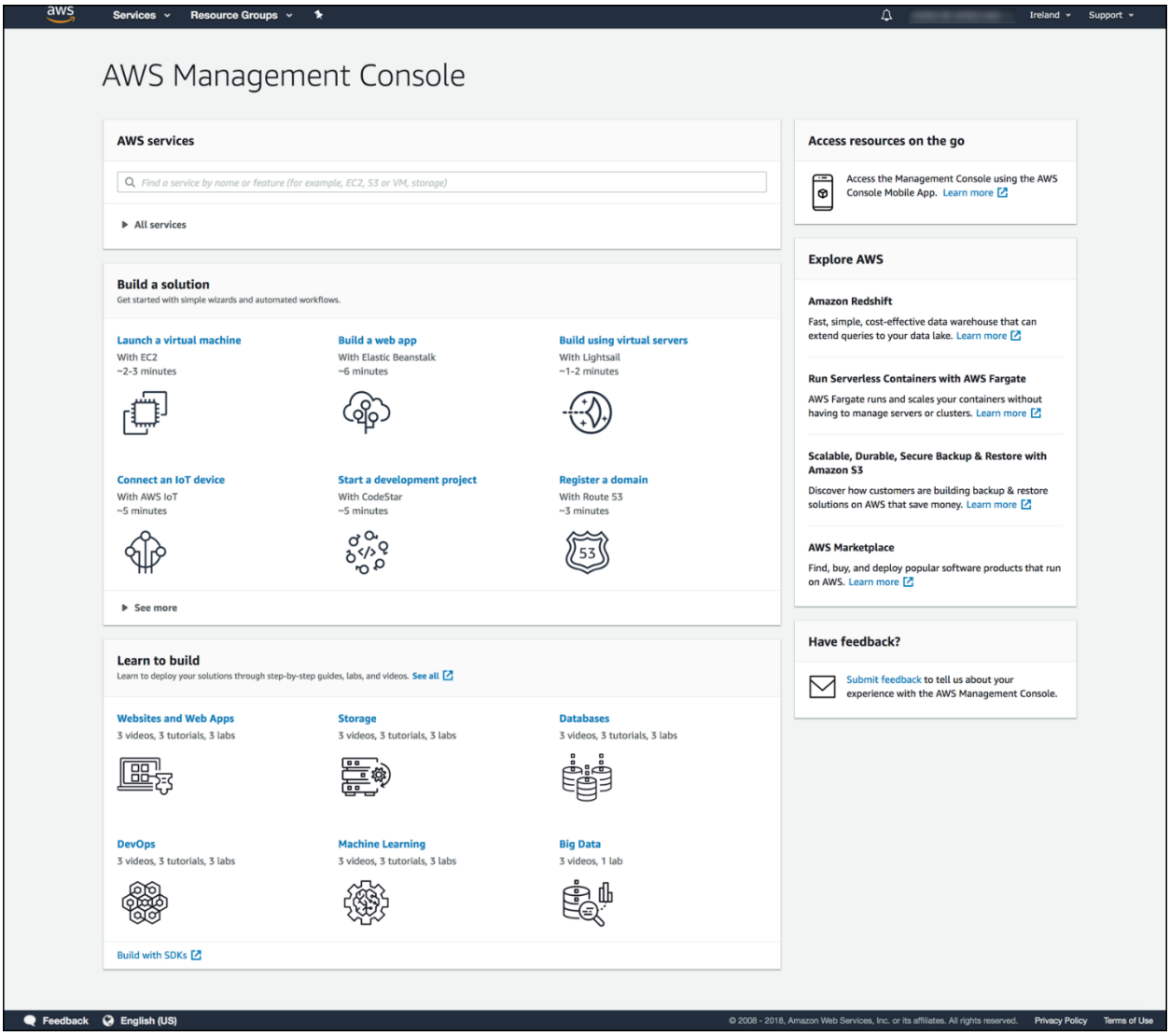

Figura 1 - AWS Management Console

A interface de linha de comando (CLI) da AWS é uma ferramenta unificada usada para gerenciar os serviços em nuvem da AWS.

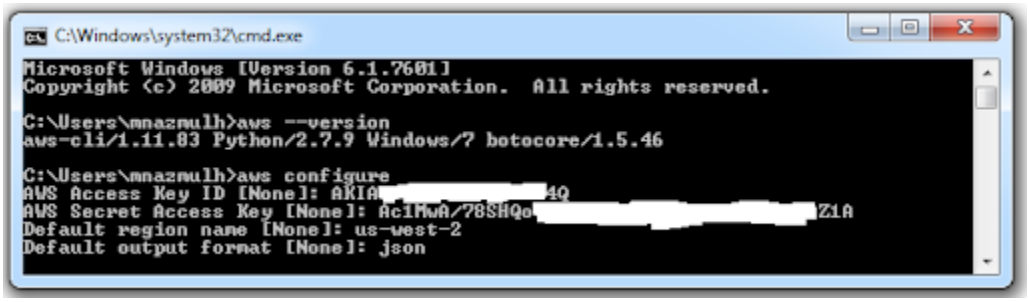

Figura 2 - AWS CLI

Com apenas uma ferramenta para baixar e configurar, você pode controlar vários serviços da linha de comando e automatizá-los por meio de scripts.

Os AWS Software Development Kits (SDKs) fornecem uma interface de programação de aplicativos (API) que interage com os serviços da Web que constituem fundamentalmente a plataforma da AWS.

Para maiores informações de AWS SDK´s: https://aws.amazon.com/pt/tools/

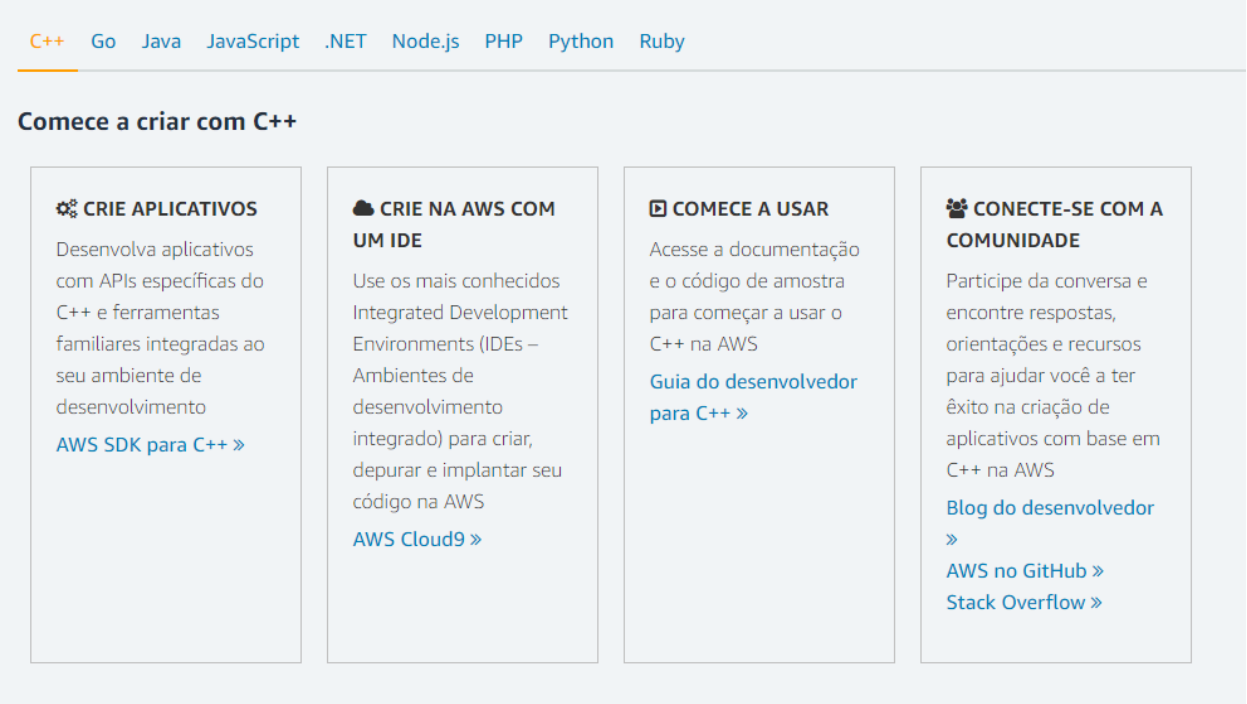

Figura 3 - AWS SDK

Os SDKs fornecem suporte para muitas linguagens e plataformas de programação diferentes para permitir que você trabalhe com seu idioma preferido. Embora você possa certamente fazer chamadas HTTP diretamente para os terminais de serviço da Web, o uso dos SDKs pode reduzir a complexidade da codificação, fornecendo acesso programático a muitos dos serviços.

Serviços de Computação e Rede

A AWS fornece uma variedade de serviços de computação e rede para fornecer funcionalidade essencial para as empresas desenvolverem e executarem suas cargas de trabalho.

Esses serviços de computação e de rede podem ser aproveitados com os serviços de armazenamento, banco de dados e aplicativos para fornecer uma solução completa.

Esta seção oferece uma descrição de alto nível dos principais serviços de computação e rede.

Amazon Elastic Compute Cloud (Amazon EC2)

O Amazon Elastic Compute Cloud (Amazon EC2) é um serviço da web que fornece capacidade de computação redimensionável na nuvem. Ele permite que as organizações obtenham e configurem servidores virtuais nos datacenters da Amazon e aproveitem esses recursos para criar e hospedar sistemas de software.

As organizações podem selecionar entre uma variedade de sistemas operacionais e configurações de recursos (memória, CPU, armazenamento etc.) ideais para o perfil do aplicativo de cada carga de trabalho.

O Amazon EC2 apresenta um verdadeiro ambiente de computação virtual, permitindo que as organizações iniciem recursos de computação com uma variedade de sistemas operacionais, carreguem-nos com aplicativos personalizados e gerenciem permissões de acesso à rede, mantendo o controle completo.

AWS Lambda

O AWS Lambda é uma plataforma de computação de administração zero para desenvolvedores da Web de back-end que executa seu código na Nuvem da AWS e fornece uma estrutura de preços refinada.

O AWS Lambda executa seu código de back-end em sua própria frota de computação da AWS de instâncias do Amazon EC2 em várias zonas de

disponibilidade em uma região, o que fornece alta disponibilidade, segurança, desempenho e escalabilidade da infraestrutura da AWS.

Escala Automática

A Escala Automática (Auto Scaling) permite que as organizações aumentem ou diminuam a capacidade do Amazon EC2 automaticamente, de acordo com as condições definidas para a carga de trabalho específica.

Não só pode ser usado para ajudar a manter a disponibilidade do aplicativo e garantir que o número desejado de instâncias do Amazon EC2 esteja em execução, mas também permite que os recursos aumentem e diminuam de acordo com as demandas de cargas de trabalho dinâmicas. Em vez de provisionar para o pico de carga, as organizações podem otimizar custos e usar apenas a capacidade realmente necessária.

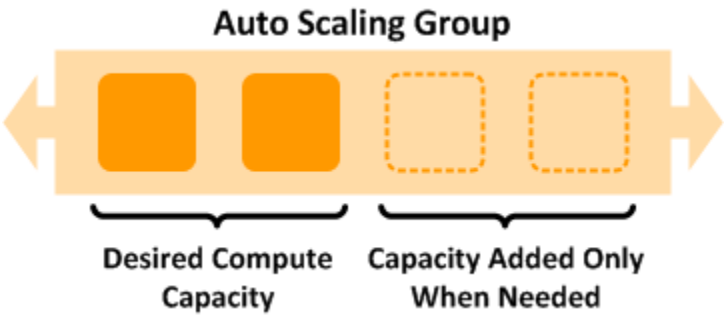

Figura 4 - Auto Scaling

O Auto Scaling é adequado tanto para aplicativos que possuem padrões de demanda estáveis quanto para aplicativos que apresentam variabilidade horária, diária ou semanal de uso.

Balanceamento de carga elástico

O Elastic Load Balancing distribui automaticamente o tráfego de aplicativos recebidos por várias instâncias do Amazon EC2 na nuvem. Ele permite que as organizações atinjam níveis mais altos de tolerância a falhas em seus aplicativos, fornecendo perfeitamente a quantidade necessária de capacidade de balanceamento de carga necessária para distribuir o tráfego de aplicativos.

AWS Elastic Beanstalk

O AWS Elastic Beanstalk é a maneira mais rápida e simples de colocar um aplicativo Web em funcionamento na AWS. Os desenvolvedores podem simplesmente fazer upload do código do aplicativo e o serviço lida automaticamente com todos os detalhes, como provisionamento de recursos, balanceamento de carga, Auto Scaling e monitoramento.

Ele fornece suporte para uma variedade de plataformas, incluindo PHP, Java, Python, Ruby, Node.js, .NET e Go. Com o AWS Elastic Beanstalk, as organizações mantêm controle total sobre os recursos da AWS que alimentam o aplicativo e podem acessar os recursos subjacentes, e outros recursos a qualquer momento.

Nuvem Virtual Privada da Amazon (Amazon VPC)

O Amazon Virtual Private Cloud (Amazon VPC) permite que as organizações provisionem uma seção isolada da AW, na qual podem iniciar os recursos da AWS em uma rede virtual que definem.

As organizações têm controle total sobre o ambiente virtual, incluindo seleção do intervalo de endereços IP, criação de sub-redes e configuração de tabelas de rotas e gateways de rede. Além disso, as organizações podem estender suas redes corporativas de data center à AWS usando conexões de rede virtual privada (VPN) de hardware ou software ou circuitos dedicados usando o AWS Direct Connect.

AWS Direct Connect

O AWS Direct Connect permite que as organizações estabeleçam uma conexão de rede dedicada de seus datacenters à AWS. Usando o AWS Direct Connect, as organizações podem estabelecer conectividade privada entre a AWS e seus ambientes de data center, escritório ou colocation, que em muitos casos podem reduzir os custos de rede, aumentar o rendimento da largura de banda e fornecer uma experiência de rede mais consistente do que as conexões VPN baseadas na Internet.

Amazon Route 53

O Amazon Route 53 é um serviço da Web de DNS (Domain Name System) altamente disponível e escalonável. Ele foi projetado para oferecer aos desenvolvedores e empresas uma maneira extremamente confiável e econômica de direcionar os usuários finais para aplicativos da Internet, traduzindo nomes legíveis por humanos, como www.examplo.com.br, nos endereços IP numéricos, como 192.168.2.1, que computadores usam para conectar um ao outro.

O Amazon Route 53 também serve como registrador de domínio, permitindo que você compre e gerencie domínios diretamente da AWS.

Armazenamento e Entrega de Conteúdo

A AWS fornece uma variedade de serviços para atender às suas necessidades de armazenamento, como Amazon Simple Storage Service, Amazon CloudFront e Amazon Elastic Block Store.

Esta seção fornece uma visão geral dos serviços de armazenamento e entrega de conteúdo.

Serviço de Armazenamento Simples da Amazon (Amazon S3)

O Amazon Simple Storage Service (Amazon S3) fornece aos desenvolvedores e equipes de tecnologia armazenamento de objetos altamente durável e escalável que lida com quantidades praticamente ilimitadas de dados e grande número de usuários simultâneos.

As organizações podem armazenar qualquer número de objetos de qualquer tipo, como páginas HTML, arquivos de código-fonte, arquivos de imagem e dados criptografados e acessá-los usando protocolos baseados em HTTP.

O Amazon S3 fornece armazenamento econômico de objetos para uma ampla variedade de casos de uso, incluindo backup e recuperação, análise de big data, recuperação de desastres, aplicativos em nuvem e distribuição de conteúdo.

Amazon Glacier – A Geleira da Amazon

O Amazon Glacier é um serviço de armazenamento seguro, durável e de custo extremamente baixo para arquivamento de dados e backup a longo prazo. As organizações podem armazenar grandes ou pequenas quantidades de dados de maneira confiável por um custo muito baixo por gigabyte por mês.

Para manter os custos baixos para os clientes, o Amazon Glacier é otimizado para dados acessados com pouca frequência, onde é necessário um tempo de recuperação de várias horas. O Amazon S3 se integra diretamente ao Amazon Glacier para permitir que as organizações escolham a camada de armazenamento correta para suas cargas de trabalho.

Amazon Elastic Block Store (Amazon EBS)

O Amazon Elastic Block Store (Amazon EBS) fornece volumes persistentes de armazenamento em nível de bloco para uso com instâncias do Amazon EC2.

Cada volume do Amazon EBS é replicado automaticamente em sua Zona de disponibilidade para proteger as organizações contra falhas de componentes, oferecendo alta disponibilidade e durabilidade.

Ao oferecer desempenho consistente e de baixa latência, a Amazon o EBS fornece o armazenamento em disco necessário para executar uma ampla variedade de cargas de trabalho.

Gateway de Armazenamento da AWS

O AWS Storage Gateway é um serviço que conecta um dispositivo de software local com armazenamento baseado em nuvem para fornecer integração perfeita e segura entre o ambiente de TI de uma organização e a infraestrutura de armazenamento da AWS.

O serviço suporta protocolos de armazenamento padrão do setor que funcionam com aplicativos existentes. Ele fornece desempenho de baixa latência, mantendo

um cache de dados acessados com frequência no local, enquanto armazena com segurança todos os dados criptografados no Amazon S3 ou Amazon Glacier.

Amazon CloudFront

O Amazon CloudFront é um serviço da Web de entrega de conteúdo. Ele se integra a outros serviços da AWS para oferecer aos desenvolvedores e empresas uma maneira fácil de distribuir conteúdo para usuários em todo o mundo com baixa latência, altas velocidades de transferência de dados e sem compromissos mínimos de uso.

O Amazon CloudFront pode ser usado para fornecer todo o conteúdo de um site, incluindo conteúdo dinâmico, estático, de streaming e interativo, usando uma rede global de pontos de presença.

As solicitações de conteúdo são roteadas automaticamente para o local da borda mais próximo, para que o conteúdo seja entregue com o melhor desempenho possível aos usuários finais em todo o mundo.

Serviços de Banco de Dados

A AWS fornece serviços de banco de dados relacional e NoSQL totalmente gerenciados e armazenamento em cache na memória como um serviço e uma solução de armazém de dados em escala de petabytes.

Serviço de banco de dados relacional da Amazon (Amazon RDS)

O Serviço de banco de dados relacional da Amazon (Amazon RDS) fornece um banco de dados relacional totalmente gerenciado, com suporte para muitos mecanismos populares de código-fonte aberto e comercial. É um serviço econômico que permite que as organizações iniciem bancos de dados seguros, altamente disponíveis, tolerantes a falhas e prontos para produção em minutos.

Como o Amazon RDS gerencia tarefas de administração que consomem tempo, incluindo backups, aplicação de patches, monitoramento, dimensionamento e replicação de software, os recursos organizacionais podem se concentrar em aplicativos e negócios geradores de receita, em vez de tarefas operacionais.

Amazon DynamoDB

O Amazon DynamoDB é um serviço de banco de dados NoSQL rápido e flexível para todos os aplicativos que precisam de latência consistente de um dígito de milissegundo em qualquer escala. É um banco de dados totalmente gerenciado e suporta modelos de dados de documento e chave / valor.

Seu modelo de dados flexível e desempenho confiável o tornam ideal para dispositivos móveis, web, jogos, tecnologia de anúncios, Internet das Coisas e muitos outros aplicativos.

Amazon Redshift

O Amazon Redshift é um serviço de data warehouse rápido, totalmente gerenciado e em escala de petabytes, que simplifica a análise de dados estruturados.

O Amazon Redshift fornece uma interface SQL padrão que permite que as organizações usem as ferramentas de inteligência de negócios existentes. Ao alavancar tecnologia de armazenamento colunar que melhora a eficiência de E / S e consultas paralelas em vários nós, o Amazon Redshift é capaz de oferecer um desempenho rápido das consultas.

A arquitetura Amazon Redshift permite que as organizações automatizem a maioria das tarefas administrativas comuns associadas ao provisionamento, configuração e monitoramento de um data warehouse na nuvem.

Amazon ElastiCache

O Amazon ElastiCache é um serviço da Web que simplifica a implantação, operação e dimensionamento de um cache na memória na nuvem.

O serviço melhora o desempenho dos aplicativos da Web, permitindo que as organizações recuperem informações de caches rápidos, gerenciados e na memória, em vez de confiar inteiramente em bancos de dados mais lentos e baseados em disco.

No momento da redação deste artigo, o Amazon ElastiCache suporta os mecanismos de cache Memcached e Redis.

Ferramentas de Gerenciamento

A AWS fornece uma variedade de ferramentas que ajudam as organizações a gerenciar seus recursos da AWS.

Amazon CloudWatch

O Amazon CloudWatch é um serviço de monitoramento dos recursos da AWS e dos aplicativos em execução na mesma.

Ele permite às organizações coletar e rastrear métricas, coletar e monitorar arquivos de log e definir alarmes. Ao aproveitar o Amazon CloudWatch, as organizações podem obter visibilidade em todo o sistema sobre a utilização de recursos, desempenho de aplicativos e integridade operacional.

Usando essas informações, as organizações podem reagir, conforme necessário, para manter os aplicativos funcionando sem problemas.

AWS CloudFormation

O AWS CloudFormation oferece aos desenvolvedores e administradores de sistemas uma maneira eficaz de criar e gerenciar uma coleção de recursos relacionados da AWS, provisionando e atualizando-os de maneira ordenada e previsível.

O AWS CloudFormation define uma linguagem de modelagem baseada em JSON que pode ser usada para descrever todos os recursos da AWS necessários para uma carga de trabalho.

Os modelos podem ser enviados ao AWS CloudFormation e o serviço cuidará do provisionamento e da configuração desses recursos na ordem apropriada.

AWS CloudTrail

O AWS CloudTrail é um serviço da Web que registra as chamadas da API da AWS para uma conta e entrega arquivos de log para auditoria e revisão.

As informações registradas incluem a identidade do chamador da API, a hora da chamada da API, o endereço IP de origem do chamador da API, os parâmetros de solicitação e os elementos de resposta retornados pelo serviço.

AWS Config

O AWS Config é um serviço totalmente gerenciado que fornece às organizações um inventário de recursos da AWS, histórico de configuração e notificações de alterações na configuração para habilitar a segurança e a governança.

Com o AWS Config, as organizações podem descobrir os recursos existentes da AWS, exportar um inventário de seus recursos da AWS com todos os detalhes de configuração e determinar como um recurso foi configurado a qualquer momento.

Esses recursos permitem auditoria de conformidade, análise de segurança, rastreamento de alterações de recursos e solução de problemas.

Segurança e Identidade

A AWS fornece serviços de segurança e identidade que ajudam as organizações a proteger seus dados e sistemas na nuvem.

Gerenciamento de identidade e acesso da AWS (IAM)

O AWS Identity and Access Management (IAM) permite que as organizações controlem com segurança o acesso aos serviços e recursos da AWS para seus usuários. Usando o IAM, as organizações podem criar e gerenciar usuários e grupos da AWS e usar permissões para permitir e negar seu acesso para recursos da AWS.

Serviço de Gerenciamento de Chaves da AWS (KMS)

O AMS Key Management Service (KMS) é um serviço gerenciado que facilita as organizações criar e controlar as chaves de criptografia usadas para criptografar seus dados e usa os HSMs (Hardware Security Modules) para proteger a segurança de suas chaves.

O AWS KMS está integrado a vários outros serviços da AWS Cloud para ajudar a proteger os dados armazenados com esses serviços.

Serviço de Diretório da AWS

O AWS Directory Service permite que as organizações configurem e executem o Microsoft Active Directory na nuvem da AWS ou conectem seus recursos da AWS a um Microsoft Active Directory local existente.

As organizações podem usá-lo para gerenciar usuários e grupos, fornecer logon único para aplicativos e serviços, criar e aplicar Diretivas de Grupo, instâncias de ingresso em domínio do Amazon EC2 e simplificar a implantação e o gerenciamento de cargas de trabalho Linux e Microsoft Windows baseadas em nuvem.

AWS Certificate Manager

O AWS Certificate Manager é um serviço que permite que as organizações provisionem, gerenciem e implantem facilmente certificados SSL / TLS (Secure Sockets Layer / Transport Layer Security) para uso com os serviços da AWS.

Ele remove o processo manual demorado de compra, upload e renovação de certificados SSL / TLS.

Com o AWS Certificate Manager, as organizações podem solicitar rapidamente um certificado, implantá-lo nos recursos da AWS, como as distribuições Elastic Load Balancing ou Amazon CloudFront, e permitir que o AWS Certificate Manager lide com renovações de certificados.

AWS Web Application Firewall (WAF)

O AWS Web Application Firewall (WAF) ajuda a proteger aplicativos da Web contra ataques e explorações comuns que podem afetar a disponibilidade de aplicativos, comprometer a segurança ou consumir recursos excessivos. O AWS WAF oferece às organizações o controle sobre qual tráfego permitir ou bloquear seus aplicativos da web, definindo regras de segurança da web personalizáveis.

Serviços de Aplicação

A AWS fornece uma variedade de serviços gerenciados para usar com aplicativos.

Amazon API Gateway

O Amazon API Gateway é um serviço totalmente gerenciado que facilita para os desenvolvedores a criar, publicar, manter, monitorar e proteger APIs em qualquer escala.

As organizações podem criar uma API que atua como uma "porta de entrada" para aplicativos acessarem dados, lógica comercial ou funcionalidade de serviços de

back-end, como cargas de trabalho em execução no Amazon EC2, código em execução no AWS Lambda ou qualquer aplicativo da Web.

O Amazon API Gateway lida com todas as tarefas envolvidas na aceitação e processamento de até centenas de milhares de chamadas simultâneas à API, incluindo gerenciamento de tráfego, autorização e controle de acesso, monitoramento e gerenciamento de versão da API.

Transcodificador Elástico Amazon

O Amazon Elastic Transcoder é uma transcodificação de mídia na nuvem. Ele foi projetado para ser uma maneira altamente escalável e econômica para que desenvolvedores e empresas convertam (ou transcodifiquem) arquivos de mídia de seus formatos de origem em versões que serão reproduzidas em dispositivos como smartphones, tablets e PCs.

Serviço de Notificação Simples da Amazon (Amazon SNS)

O Serviço de Notificação Simples da Amazon (Amazon SNS) é um serviço da Web que coordena e gerencia a entrega ou o envio de mensagens aos destinatários.

No Amazon SNS, existem dois tipos de clientes - editores e assinantes - também chamados de produtores e consumidores.

Os editores se comunicam de forma assíncrona com os assinantes, produzindo e enviando uma mensagem para um tópico, que é um ponto de acesso lógico e um canal de comunicação.

Os assinantes consomem ou recebem a mensagem ou notificação em um dos protocolos suportados quando estão inscritos no tópico.

Serviço de Email Simples da Amazon (Amazon SES)

O Amazon Simple Email Service (Amazon SES) é um serviço de email econômico que as organizações podem usar para enviar email transacional, mensagens de marketing ou qualquer outro tipo de conteúdo para seus clientes.

O Amazon SES também pode ser usado para receber mensagens e entregar para um bucket do Amazon S3, chame o código personalizado por meio de uma função do AWS Lambda ou publique notificações no Amazon SNS.

Serviço de Fluxo de Trabalho Simples da Amazon (Amazon SWF)

O Amazon Simple Workflow Service (Amazon SWF) ajuda os desenvolvedores a criar, executar e dimensionar tarefas em segundo plano com etapas paralelas ou sequenciais.

O Amazon SWF pode ser considerado um rastreador de estado e coordenador de tarefas totalmente gerenciados na nuvem. Em padrões arquiteturais comuns, se as etapas de seu aplicativo levarem mais de 500 milissegundos para concluir, é de vital importância acompanhar o estado do processamento e fornecer a capacidade de recuperar ou tentar novamente se uma tarefa falhar. O Amazon SWF ajuda as organizações a alcançar essa confiabilidade.

Serviço de fila simples da Amazon (Amazon SQS)

O Amazon SQS (Amazon Simple Queue Service) é um serviço de enfileiramento de mensagens rápido, confiável, escalável e totalmente gerenciado.

O Amazon SQS torna mais simples e econômico desacoplar os componentes de um aplicativo em nuvem. Com o Amazon SQS, as organizações podem transmitir qualquer volume de dados, em qualquer nível de taxa de transferência, sem perder mensagens ou exigir que outros serviços estejam sempre disponíveis.

S3 & Glacier

Aqui pretendemos fornecer a você um entendimento básico dos principais serviços de armazenamento de objetos disponíveis na AWS: Amazon Simple Storage Service (Amazon S3) e Amazon Glacier.

O Amazon S3 fornece aos desenvolvedores e equipes de TI armazenamento em nuvem seguro, durável e altamente escalável.

O Amazon S3 é um armazenamento de objetos fácil de usar, com uma interface simples de serviço da Web que você pode usar para armazenar e recuperar qualquer quantidade de dados de qualquer lugar da Web.

O Amazon S3 também permite que você pague apenas pelo armazenamento que realmente usa, o que elimina o planejamento e as restrições de capacidade associadas ao armazenamento tradicional. O Amazon S3 é um dos primeiros serviços introduzidos pela AWS e serve como um dos serviços da Web fundamentais - quase qualquer aplicativo em execução na AWS usa o Amazon S3, direta ou indiretamente.

O Amazon S3 pode ser usado sozinho ou em conjunto com outros serviços e oferece um nível muito alto de integração com muitos outros serviços em nuvem da AWS.

Por exemplo, o Amazon S3 serve como armazenamento de destino durável para o Amazon Kinesis e o Amazon Elastic MapReduce (Amazon EMR), é usado como o armazenamento para os snapshots do Amazon Elastic Block Store (Amazon EBS) e do Amazon Relational Database Service (Amazon RDS) e é usado como um mecanismo de armazenamento temporário ou de armazenamento de dados para o Amazon Redshift e o Amazon DynamoDB, entre muitas outras funções.

Como o Amazon S3 é tão flexível, altamente integrado e tão comumente usado, é importante entender esse serviço em detalhes.

Casos de uso comuns para armazenamento do Amazon S3 incluem:

- Backup e arquivamento de dados no local ou na nuvem
- Armazenamento e distribuição de conteúdo, mídia e software
- Análise de big data
- Hospedagem estática de sites
- Hospedagem de aplicativos móveis e de Internet nativos da nuvem
- Recuperação de desastre

Para dar suporte a esses casos de uso e muito mais, o Amazon S3 oferece uma variedade de classes de armazenamento projetadas para vários casos de uso genéricos: uso geral, acesso pouco frequente e arquivo morto.

Para ajudar a gerenciar dados durante seu ciclo de vida, o Amazon S3 oferece políticas configuráveis de ciclo de vida. Usando políticas de ciclo de vida, você pode migrar seus dados automaticamente para a classe de armazenamento mais apropriada, sem modificar o código do seu aplicativo.

O Amazon Glacier é outro serviço de armazenamento em nuvem relacionado ao Amazon S3, mas otimizado para arquivamento de dados e backup de longo prazo a um custo extremamente baixo. O Amazon Glacier é adequado para "dados frios", dados raramente acessados e para os quais é aceitável um tempo de recuperação de três a cinco horas.

O Amazon Glacier pode ser usado como uma classe de armazenamento do Amazon S3 (consulte os tópicos Classes de armazenamento e Gerenciamento do ciclo de vida do objeto na seção Recursos avançados do Amazon S3) e como um serviço de armazenamento de arquivo independente (consulte a seção Amazon Glacier).

Armazenamento de Objetos versus Armazenamento Tradicional de Blocos e Armazenamento de Arquivos

Nos ambientes tradicionais, dois tipos de armazenamento predominam: Armazenamento em Bloco e Armazenamento de Arquivos.

O armazenamento em bloco opera em um nível inferior - o nível do dispositivo de armazenamento bruto - e gerencia os dados como um conjunto de blocos numerados e de tamanho fixo.

O armazenamento de arquivos opera em um nível mais alto - o nível do sistema operacional - e gerencia os dados como uma hierarquia nomeada de arquivos e pastas.

O armazenamento de blocos e arquivos geralmente é acessado através de uma rede na forma de uma SAN (Storage Area Network) para armazenamento em bloco, usando protocolos como iSCSI ou Fibre Channel ou como um servidor de arquivos NAS (Network Attached Storage) ou "arquivador" para armazenamento de arquivos, usando protocolos como o CIFS (Common Internet File System) ou o NFS (Network File System). Seja anexado diretamente ou anexado à rede, bloco ou arquivo, esse tipo de armazenamento está intimamente associado ao servidor e ao sistema operacional que está usando o armazenamento.

O armazenamento de objetos do Amazon S3 é algo bem diferente. O Amazon S3 é o armazenamento de objetos na nuvem. Em vez de estar intimamente associado a um servidor, o armazenamento do Amazon S3 é independente de um servidor e é acessado pela Internet. Em vez de gerenciar dados como blocos ou arquivos usando os protocolos SCSI, CIFS ou NFS, os dados são gerenciados como objetos usando um Programa de Aplicativo de Interface (API) criada em verbos HTTP padrão.

Cada objeto do Amazon S3 contém dados e metadados. Os objetos residem em contêineres chamados buckets e cada objeto é identificado por uma chave especificada pelo usuário (nome do arquivo).

Os buckets são uma pasta simples e plana, sem hierarquia de sistema de arquivos. Ou seja, você pode ter vários buckets, mas não pode ter um sub-bucket dentro de um bucket. Cada bucket pode conter um número ilimitado de objetos.

É fácil pensar em um objeto Amazon S3 (ou na parte de dados de um objeto) como um arquivo e a chave como o nome do arquivo. No entanto, lembre-se de que o Amazon S3 não é um sistema de arquivos tradicional e difere de maneiras significativas.

No Amazon S3, você obtém um objeto ou coloca um objeto, operando em objeto inteiro de uma só vez, em vez de atualizar incrementalmente partes do objeto, como faria com um arquivo.

Você não pode "montar" um bucket, "abrir" um objeto, instalar um sistema operacional no Amazon S3 ou executar um banco de dados nele. Em vez de um sistema de arquivos, o Amazon S3 é um armazenamento de objetos altamente durável e escalonável, otimizado para leituras e construído com um conjunto de recursos intencionalmente minimalista.

Ele fornece uma abstração simples e robusta para armazenamento de arquivos que o libera de muitos detalhes subjacentes com os quais você normalmente precisa lidar no armazenamento tradicional. Por exemplo, com o Amazon S3, você não precisa se preocupar com os limites de armazenamento do dispositivo ou sistema de arquivos e com o planejamento da capacidade - um único bucket pode armazenar um número ilimitado de arquivos. Você também não precisa se preocupar com a durabilidade ou replicação dos dados nas zonas de disponibilidade. Os objetos do Amazon S3 são replicados automaticamente em vários dispositivos em várias instalações de uma região.

O mesmo ocorre com a escalabilidade - se a taxa de solicitações aumentar constantemente, o Amazon S3 particiona automaticamente os buckets para oferecer suporte a taxas de solicitações muito altas e acesso simultâneo por muitos clientes.

Noções Básicas do Amazon Simple Storage Service (Amazon S3)

Agora que você entende algumas das principais diferenças entre o armazenamento tradicional de blocos, arquivos e o armazenamento de objetos na nuvem, podemos explorar os conceitos básicos do Amazon S3 com mais detalhes.

Buckets

Um bucket é um contêiner (pasta da web) para objetos (arquivos) armazenados no Amazon S3. Todo objeto do Amazon S3 está contido em um bucket. Os buckets formam o namespace de nível superior para o Amazon S3 e nomes de bucket são globais. Isso significa que os nomes dos seus buckets devem ser exclusivos em todas as contas da AWS, como os nomes de domínio do DNS (Sistema de Nomes de Domínio), não apenas dentro da sua própria conta.

Os nomes dos intervalos podem conter até 63 letras minúsculas, números, hífens e pontos. Você pode criar e usar vários buckets. Você pode ter até 100 por conta, por padrão. Tendo solicitação direta a AWS, este número pode ser aumentado.

É uma prática recomendada usar nomes de grupos que contenham seu nome de domínio e estejam em conformidade com as regras para nomes DNS. Isso garante que os nomes dos seus buckets sejam seus, que possam ser usados em todas as regiões e que hospedem sites estáticos.

Regiões da AWS

Embora o espaço para nome dos buckets do Amazon S3 seja global, cada bucket do Amazon S3 é criado em uma região específica que você escolher. Isso permite que você controle onde seus dados são armazenados.

Você pode criar e usar buckets localizados perto de um conjunto específico de usuários finais ou clientes, a fim de minimizar a latência, ou localizados em uma região específica para satisfazer as preocupações de localidade e soberania de dados ou localizados longe de suas instalações principais, a fim de satisfazer as necessidades de recuperação e conformidade de desastres.

Você controla a localização dos seus dados. Os dados em um bucket do Amazon S3 são armazenados nessa região, a menos que você os copie explicitamente para outro bucket localizado em uma região diferente.

Objetos

Objetos são as entidades ou arquivos armazenados nos buckets do Amazon S3. Um objeto pode armazenar praticamente qualquer tipo de dados em qualquer formato. Os objetos podem variar em tamanho de 0 bytes a 5 TB, e um único depósito pode armazenar um número ilimitado de objetos.

Isso significa que o Amazon S3 pode armazenar uma quantidade praticamente ilimitada de dados. Cada objeto consiste em dados (o próprio arquivo) e metadados (dados sobre o arquivo). Isso significa que os dados de um objeto são tratados simplesmente como um fluxo de bytes - o Amazon S3 não sabe nem se importa com o tipo de dados que você está armazenando e o serviço não age de maneira diferente para dados de texto e dados binários.

Os metadados associados a um objeto Amazon S3 são um conjunto de pares de nome / valor que descrevem o objeto. Existem dois tipos de metadados: metadados do sistema e metadados do usuário.

Os metadados do sistema são criados e usados pelo próprio Amazon S3 e incluem itens como a data da última modificação, tamanho do objeto, resumo MD5 e tipo de conteúdo HTTP.

Os metadados do usuário são opcionais e só podem ser especificados no momento em que um objeto é criado. Você pode usar metadados personalizados para marcar seus dados com atributos significativos para você.

Chaves

Todo objeto armazenado em um bucket S3 é identificado por um identificador exclusivo chamado chave. Você pode pensar na chave como um nome de arquivo. Uma chave pode ter até 1024 bytes de caracteres Unicode UTF-8, incluindo barras, barras invertidas, pontos e traços incorporados.

As chaves devem ser exclusivas em um único depósito, mas diferentes depósitos podem conter objetos com a mesma chave. A combinação de bucket, chave e ID da versão opcional identificam exclusivamente um objeto do Amazon S3.

URL do objeto

O Amazon S3 é armazenamento para a Internet, e todo objeto do Amazon S3 pode ser endereçado por uma URL exclusiva formada usando o terminal de serviços da Web, o nome do bloco e a chave do objeto.

Por exemplo, com a URL:

http://meubucket.s3.amazonaws.com/jose.doc

meubucket é o nome do bucket S3 e **jose.doc** é a chave ou o nome do arquivo.

Se outro objeto for criado, por exemplo:

http://meubucket.s3.amazonaws.com/do/re/mi/fa/jose.doc

o nome do bucket ainda é meubucket, mas agora a chave ou o nome do arquivo é a taxa da string **/do/re/mi/fa/jose.doc**

Uma chave pode conter caracteres delimitadores, como barras ou barras invertidas, para ajudá-lo a nomear e organizar logicamente seus objetos do Amazon S3, mas para o Amazon S3 é simplesmente um nome de chave longo em um espaço de nome simples. Não existe uma hierarquia real de arquivos e pastas.

Operações do Amazon S3

A API do Amazon S3 é intencionalmente simples, com apenas algumas operações comuns. Elas incluem:

- Criar / excluir um bucket
- Escrever um objeto
- Ler um objeto
- Excluir um objeto
- Listar chaves em um bucket
- Interface REST

A interface nativa do Amazon S3 é uma API REST (Representational State Transfer). Com a interface REST, você usa solicitações HTTP ou HTTPS padrão para criar e excluir buckets, listar chaves e ler e gravar objetos. O REST mapeia “verbos” HTTP padrão (métodos HTTP) para as operações familiares CRUD (Criar, Ler, Atualizar, Excluir).

Criar é HTTP PUT (e às vezes POST); read é HTTP GET; delete é HTTP DELETE; e update é HTTP POST (ou algumas vezes PUT). Sempre use solicitações de HTTPS para API do Amazon S3 para garantir que suas solicitações e dados estejam seguros.

Na maioria dos casos, os usuários não usam a interface REST diretamente, mas interagem com o Amazon S3 usando uma das interfaces de nível superior disponíveis. Isso inclui os SDKs (AWS Software Development Kits) (bibliotecas de wrapper) para iOS, Android, JavaScript, Java, .NET, Node.js, PHP, Python, Ruby, Go e C ++, a interface de linha de comando (CLI) da AWS, e o AWS Management Console.

O Amazon S3 originalmente suportava uma API SOAP (Simple Object Access Protocol) além da API REST, mas você deve usar a API REST. O ponto de extremidade HTTPS herdado ainda está disponível, mas novos recursos não são suportados.

Durabilidade e Disponibilidade

A durabilidade e a disponibilidade dos dados estão relacionadas, mas são conceitos ligeiramente diferentes.

A durabilidade aborda a pergunta: "Meus dados ainda estarão lá no futuro?". A disponibilidade aborda a pergunta: "Posso acessar meus dados agora?". O Amazon S3 foi projetado para fornecer recursos muito altos de durabilidade e disponibilidade muito alta para seus dados.

O armazenamento padrão do Amazon S3 foi projetado para 99,999999999% de durabilidade e 99,99% de disponibilidade de objetos em um determinado ano. Por exemplo, se você armazenar 10.000 objetos com o Amazon S3, poderá esperar, em média, a perda de um único objeto a cada 10.000.000 anos.

O Amazon S3 alcança alta durabilidade, armazenando automaticamente dados de forma redundante em vários dispositivos em várias instalações de uma região. Ele foi projetado para suportar a perda simultânea de dados em duas instalações sem perda de dados do usuário. O Amazon S3 fornece uma infraestrutura de armazenamento altamente durável, projetada para armazenamento de dados primários e de missão crítica.

Se você precisar armazenar dados derivados não críticos ou facilmente reproduzíveis (como miniaturas de imagens) que não exijam esse alto nível de durabilidade, você pode optar por usar o armazenamento de redundância reduzida (RRS) a um custo menor. O RRS oferece 99,99% de durabilidade com um custo menor de armazenamento do que o armazenamento tradicional Amazon S3.

Embora o armazenamento do Amazon S3 ofereça durabilidade muito alta no nível da infraestrutura, ainda é uma prática recomendada proteger contra exclusão ou substituição acidental de dados no nível do usuário, usando recursos adicionais como controle de versão, replicação entre regiões e exclusão do MFA.

Consistência dos Dados

O Amazon S3 é um sistema eventualmente consistente. Como seus dados são replicados automaticamente em vários servidores e locais em uma região, as alterações nos dados podem levar até algumas horas para propagar para todos os locais.

Como resultado, há algumas situações em que as informações que você lê imediatamente após uma atualização podem retornar dados obsoletos. Para PUTs de novos objetos, isso não é uma preocupação. Nesse caso, o Amazon S3 fornece consistência de leitura e pós-gravação. No entanto, para PUTs para objetos existentes (substituição de objeto em uma chave existente) e para DELETEs de objeto, o Amazon S3 fornece consistência eventual.

Consistência eventual significa que, se você colocar novos dados em uma chave existente, um GET subsequente poderá retornar os dados antigos. Da mesma forma, se você excluir um objeto, um GET subsequente para esse objeto ainda pode ler o objeto excluído.

Em todos os casos, as atualizações em uma única chave são atômicas - para leituras eventualmente consistentes, você obtém os novos dados ou os dados antigos, mas nunca uma mistura inconsistente de dados.

Controle de acesso

O Amazon S3 é seguro por padrão. Quando você cria um bucket ou objeto no Amazon S3, apenas você tem acesso.

Para permitir que você conceda acesso controlado a outras pessoas, o Amazon S3 fornece controles de acesso de granulação (ACLs) e controles de acesso de granulação fina (políticas de bucket do Amazon S3, políticas de gerenciamento de acesso e identidade da AWS [IAM] políticas e autenticação de cadeia de consulta).

As ACLs do Amazon S3 permitem conceder certas permissões de granularidade grossa: READ, WRITE ou FULL-CONTROL no nível do objeto ou do bucket. As ACLs são um mecanismo de controle de acesso herdado, criado antes da existência do IAM.

Hoje, as ACLs são mais usadas para um conjunto limitado de casos de uso, como habilitar o log do bucket ou tornar um bucket que hospeda um site estático legível pelo mundo.

As políticas de bucket do Amazon S3 são o mecanismo de controle de acesso recomendado para o Amazon S3 e fornecem um controle muito mais refinado. As políticas de bucket do Amazon S3 são muito semelhantes às políticas do IAM.

- Eles estão associados ao recurso de bucket em vez de um principal do IAM.
- Eles incluem uma referência explícita ao principal do IAM na política.

Esse princípio pode ser associado a uma conta AWS diferente, portanto, as políticas de bucket do Amazon S3 permitem atribuir acesso entre contas aos recursos do Amazon S3. Usando uma política de bucket do Amazon S3, você pode especificar quem pode acessá-lo, de onde (pelo bloco ou endereço IP do CIDR) ou durante o horário do dia.

Por fim, as políticas do IAM podem ser associadas diretamente às entidades do IAM que concedem acesso a um bucket do Amazon S3, assim como podem conceder acesso a qualquer serviço e recurso da AWS. Obviamente, você só pode atribuir políticas do IAM a entidades principais nas contas da AWS que você controla.

Hospedagem Estática de Sites

Um caso de uso muito comum para armazenamento do Amazon S3 é a hospedagem estática de sites. Muitos sites, principalmente os micros sites, não precisam dos serviços de um servidor da Web completo.

Um site estático significa que todas as páginas do site contêm apenas conteúdo estático e não requerem processamento no servidor, como PHP, ASP.NET ou JSP. (Observe que isso não significa que o site não pode ser interativo e dinâmico. Isso pode ser realizado com scripts do lado do cliente, como JavaScript incorporado em páginas da Web HTML estáticas.)

Os sites estáticos têm muitas vantagens: são muito rápidos, muito escaláveis e podem ser mais seguros do que um site dinâmico típico. Se você hospedar um site estático no Amazon S3, também poderá aproveitar a segurança, durabilidade, disponibilidade e escalabilidade do Amazon S3.

Como todo objeto do Amazon S3 tem um URL, é relativamente simples transformar um bucket em um site. Para hospedar um site estático, basta configurar um bucket para hospedagem de sites e, em seguida, carregar o conteúdo do site estático no bucket.

Para configurar um bucket do Amazon S3 para hospedagem de site estático:

1. Crie um bucket com o **mesmo nome que o nome do host** do site desejado.

2. Carregue os arquivos estáticos no bucket.

3. Torne todos os arquivos públicos (legíveis pelo mundo).

4. Habilite a hospedagem estática de sites para o bucket. Isso inclui especificar um documento de índice e um documento de erro.

5. O site estará agora disponível no URL do site S3:

<nome do buket> .s3-website- <AWS-region> .amazonaws.com.

6. Crie um nome DNS amigável em seu próprio domínio para o site usando um DNS CNAME ou um alias do Amazon Route 53 que seja resolvido para o URL do site do Amazon S3.

7. O site estará agora disponível no nome de domínio do seu site.

Recursos avançados do Amazon S3

Além do básico, existem alguns recursos avançados do Amazon S3 com os quais você também deve estar familiarizado.

Prefixos e delimitadores

Embora o Amazon S3 use uma estrutura plana em um bucket, ele suporta o uso de parâmetros de prefixo e delimitador ao listar os nomes das chaves.

Esse recurso permite organizar, navegar e recuperar os objetos em um bucket hierarquicamente. Normalmente, você usaria uma barra (/) ou barra invertida (\) como delimitador e, em seguida, nomes de chaves com delimitadores incorporados para emular uma hierarquia de arquivos e pastas no espaço de nomes de chaves de objetos simples de um bucket. Por exemplo, convém armazenar uma série de logs do servidor pelo nome do servidor (como servidor2020), mas organizados por ano e mês, da seguinte forma:

logs / 2020 / janeiro / servidor2020.log

logs / 2020 / fevereiro / servidor2020.log

logs / 2020 / março / servidor2020.log

A API REST, os SDKs do wrapper, a AWS CLI e o Amazon Management Console suportam o uso de delimitadores e prefixos. Esse recurso permite organizar logicamente novos dados e facilmente manter a estrutura hierárquica de pasta e arquivo de dados existentes carregados ou armazenados em backup dos sistemas de arquivos tradicionais.

Usado junto com políticas de bucket do IAM ou Amazon S3, prefixos e delimitadores também permitem criar o equivalente a "subdiretórios" departamentais ou "diretórios pessoais" do usuário em um único bloco, restringindo ou compartilhando o acesso a esses "subdiretórios" (definidos por prefixos), conforme necessário.

Use delimitadores e prefixos de objetos para organizar hierarquicamente os objetos em seus buckets do Amazon S3, mas lembre-se sempre de que o Amazon S3 não é realmente um sistema de arquivos.

Classes de Armazenamento

O Amazon S3 oferece uma variedade de classes de armazenamento adequadas para vários casos de uso.

O Amazon S3 Standard oferece armazenamento de objetos de alta durabilidade, alta disponibilidade, baixa latência e alto desempenho para uso geral. Como fornece baixa latência de primeiro byte e alta taxa de transferência, o Standard é adequado para armazenamento de curto ou longo prazo de dados acessados com freqüência.

Na maioria dos casos de uso de uso geral, o Amazon S3 Standard é o por onde começar.

Amazon S3 Standard - Acesso infrequente (Standard-IA) oferece a mesma durabilidade, baixa latência e alta taxa de transferência que o Amazon S3 Standard, mas foi projetado para dados de longa duração e acesso com menos frequência. O Standard-IA tem um custo de armazenamento por GB-mês mais baixo que o Standard, mas o modelo de preço também inclui um tamanho mínimo de objeto (128 KB), duração mínima (30 dias) e custos de recuperação por GB, portanto, é mais adequado para infrequentemente dados acessados armazenados por mais de 30 dias.

O armazenamento de redundância reduzida (RRS) do Amazon S3 oferece durabilidade um pouco menor (4 noves) do que a Standard ou Standard-IA a um custo reduzido. É o mais apropriado para dados derivados que podem ser facilmente reproduzidos, como miniaturas de imagens.

Por fim, a classe de armazenamento Amazon Glacier oferece armazenamento em nuvem seguro, durável e de custo extremamente baixo para dados que não requerem acesso em tempo real, como arquivos e backups de longo prazo. Para manter os custos baixos, o Amazon Glacier é otimizado para dados acessados com pouca frequência, onde é adequado um tempo de recuperação de várias horas.

Para recuperar um objeto Amazon Glacier, você emite um comando de restauração usando uma das APIs do Amazon S3; três a cinco horas depois, o objeto Amazon Glacier é copiado para o Amazon S3 RRS. Observe que a restauração simplesmente cria uma cópia no Amazon S3 RRS. O objeto de dados original permanece no Amazon Glacier até excluído explicitamente.

Lembre-se também de que o Amazon Glacier permite recuperar até 5% dos dados do Amazon S3 armazenados no Amazon Glacier gratuitamente a cada mês. Restaurações além do subsídio diário de restauração incorrem em uma taxa de restauração. Consulte a página de preços do Amazon Glacier no site da AWS para obter detalhes completos. Além de atuar como uma camada de armazenamento no Amazon S3, o Amazon Glacier também é um serviço de armazenamento autônomo com uma API separada e algumas características exclusivas.

No entanto, quando você usa o Amazon Glacier como uma classe de armazenamento do Amazon S3, sempre interage com os dados por meio das APIs do Amazon S3. Consulte a seção Amazon Glacier para obter mais detalhes.

Defina uma política de recuperação de dados para limitar as restaurações na camada gratuita ou no limite máximo de GB por-hora para evitar ou minimizar as taxas de restauração do Amazon Glacier.

Gerenciamento do Ciclo de Vida do Objeto

O Amazon S3 Object Lifecycle Management é aproximadamente equivalente às camadas de armazenamento automatizadas nas infraestruturas tradicionais de armazenamento de TI.

Em muitos casos, os dados têm um ciclo de vida natural, começando como dados "quentes" (acessados com freqüência), movendo-se para dados "mornos" (acessados com menos frequência) à medida que envelhece e terminando sua vida como "fria" (backup a longo prazo ou arquivar) dados antes de uma eventual exclusão.

Por exemplo, muitos documentos comerciais são acessados com frequência quando são criados, e se tornam muito menos frequentemente acessados ao longo do tempo. Em muitos casos, no entanto, as regras de conformidade exigem que os documentos comerciais sejam arquivados e mantidos acessíveis por anos.

Da mesma forma, estudos mostram que os backups de arquivos, sistemas operacionais e bancos de dados são acessados com mais frequência nos primeiros dias após a criação, geralmente para restaurar após um erro inadvertido.

Após uma semana ou duas, esses backups permanecem um ativo crítico, mas são muito menos prováveis de serem acessados para uma restauração. Em muitos casos, as regras de conformidade exigem que um certo número de backups seja mantido por muitos anos.

Usando as regras de configuração do ciclo de vida do Amazon S3, você pode reduzir significativamente os custos de armazenamento ao fazer a transição automática de dados de uma classe de armazenamento para outra ou até mesmo excluir dados automaticamente após um período de tempo. Por exemplo, as regras do ciclo de vida para dados de backup podem ser:

1. Armazene dados de backup inicialmente no Amazon S3 Standard.
2. Após 30 dias, faça a transição para o Amazon Standard-IA.
3. Após 90 dias, faça a transição para o Amazon Glacier.
4. Após 3 anos, exclua.

As configurações do ciclo de vida são anexadas ao bucket e podem ser aplicadas a todos os objetos ou apenas aos objetos especificados por um prefixo.

Criptografia

É altamente recomendável que todos os dados confidenciais armazenados no Amazon S3 sejam criptografados, tanto em trânsito quanto em repouso.

Para criptografar os dados do Amazon S3 em trânsito, você pode usar os endpoint (endpoints) da API do Amazon S3 Secure Sockets Layer (SSL). Isso garante que todos os dados enviados para e do Amazon S3 sejam criptografados enquanto estiver em trânsito usando o protocolo HTTPS.

Para criptografar os dados do Amazon S3 em repouso, você pode usar várias variações de SSE (Server-Side Encryption).

O Amazon S3 criptografa seus dados no nível do objeto enquanto os grava em discos em seus datacenters e os descriptografa para você quando você os acessa.

Todo o SSE executado pelo Amazon S3 e pelo AWS Key Management Service (Amazon KMS) usa o Advanced Encryption Standard (AES) de 256 bits. Você também pode criptografar os dados do Amazon S3 em repouso usando o lado do cliente, criptografando seus dados no cliente antes de enviá-los para o Amazon S3.

SSE-S3 (chaves gerenciadas pela AWS)

Essa é uma solução de criptografia "estilo caixa de seleção" totalmente integrada, na qual a AWS lida com o gerenciamento e a proteção de chaves do Amazon S3. Cada objeto é criptografado com uma chave exclusiva.

A própria chave do objeto é então criptografada por uma chave mestra separada. Uma nova chave mestra é emitida pelo menos mensalmente, com a AWS rotacionando as chaves. Dados criptografados, chaves de criptografia e chaves mestras são todos armazenados separadamente em hosts seguros, aumentando ainda mais a proteção.

SSE-KMS (chaves AWS KMS)

Esta é uma solução totalmente integrada em que a Amazon lida com o gerenciamento e a proteção de chaves do Amazon S3, mas onde você gerencia as chaves.

O SSE-KMS oferece vários benefícios adicionais em comparação com o SSE-S3. Usando o SSE-KMS, existem permissões separadas para o uso da chave mestra, que fornecem proteção contra acesso não autorizado aos seus objetos armazenados no Amazon S3 e uma camada adicional de controle.

O AWS KMS também fornece auditoria, para que você possa ver quem usou sua chave para acessar qual objeto e quando eles tentaram acessar esse objeto. O AWS KMS também permite exibir quaisquer tentativas falhas de acessar dados de usuários que não tinham permissão para descriptografar os dados.

SSE-C (chaves fornecidas pelo cliente)

É usado quando você deseja manter suas próprias chaves de criptografia, mas não deseja gerenciar ou implementar sua própria biblioteca de criptografia do lado do cliente.

Com o SSE-C, a AWS faz a criptografia / descriptografia de seus objetos enquanto você mantém controle total das chaves usadas para criptografar / descriptografar os objetos no Amazon S3.

Criptografia do lado do cliente

Criptografia do lado do cliente refere-se à criptografia de dados no lado do cliente do seu aplicativo antes de enviá-los ao Amazon S3. Você tem as duas opções a seguir para usar chaves de criptografia de dados:

- Use uma chave mestra do cliente gerenciada pelo AWS KMS.
- Use uma chave mestra do lado do cliente.

Ao usar a criptografia do lado do cliente, você mantém o controle ponta a ponta do processo de criptografia, incluindo o gerenciamento das chaves de criptografia. Para máxima simplicidade e facilidade de uso, use a criptografia do servidor com chaves gerenciadas pela AWS (SSE-S3 ou SSE-KMS).

Versionamento

O controle de versão do Amazon S3 ajuda a proteger seus dados contra exclusão acidental ou mal-intencionada, mantendo várias versões de cada objeto no bucket, identificadas por um ID de versão exclusivo.

O controle de versão permite preservar, recuperar e restaurar todas as versões de todos os objetos armazenados no seu bucket do Amazon S3. Se um usuário fizer uma alteração acidental ou excluir um objeto maliciosamente no seu depósito S3,

você poderá restaurar o objeto ao seu estado original simplesmente referenciando o ID da versão, além do depósito e da chave do objeto.

O controle de versão está ativado no nível do bucket. Uma vez ativado, o controle de versão não pode ser removido de um bucket, só pode ser suspenso.

Excluisão por MFA (MFA Delete)

O MFA Delete adiciona outra camada de proteção de dados sobre a versão do bucket. A exclusão do MFA requer autenticação adicional para excluir permanentemente uma versão do objeto ou alterar o estado da versão de um bucket.

Além das credenciais de segurança normais, o MFA Delete exige um código de autenticação (uma senha temporária e única) gerada por um hardware ou dispositivo virtual de autenticação multifator (MFA). Observe que a exclusão MFA só pode ser ativado pela conta raiz.

URLs pré-assinadas

Todos os objetos do Amazon S3, por padrão, são privados, o que significa que apenas o proprietário tem acesso.

No entanto, o proprietário do objeto pode opcionalmente compartilhar objetos com outras pessoas, criando uma URL pré-assinada, usando suas próprias credenciais de segurança para conceder permissão por tempo limitado para fazer o download dos objetos.

Ao criar uma URL pré-assinada para seu objeto, você deve fornecer suas credenciais de segurança e especificar um nome de bloco, uma chave de objeto, o método HTTP (GET para fazer o download do objeto) e uma data e hora de vencimento.

Os URLs pré-assinados são válidos apenas pela duração especificada. Isso é particularmente útil para proteger contra a "captura de conteúdo" de conteúdo da Web, como arquivos de mídia armazenados no Amazon S3.

Upload de várias partes (Multipart Upload)

Para oferecer melhor suporte ao upload ou cópia de objetos grandes, o Amazon S3 fornece a API de upload com várias partes. Isso permite que você faça upload de objetos grandes como um conjunto de partes, o que geralmente oferece uma melhor utilização da rede (por meio de transferências paralelas), a capacidade de pausar e retomar e a capacidade de fazer upload de objetos em que o tamanho é inicialmente desconhecido.

O upload de várias partes é um processo de três etapas: iniciação, upload das peças e conclusão (ou abortamento). As peças podem ser carregadas independentemente em ordem arbitrária, com retransmissão, se necessário. Após o upload de todas as peças, o Amazon S3 reúne as peças para criar um objeto.

Em geral, você deve usar o upload de várias partes para objetos maiores que 100 Mbytes e o upload de várias partes para objetos maiores que 5 GB. Ao usar as APIs de baixo nível, você deve dividir o arquivo a ser carregado em partes e acompanhar as partes.

Ao usar as APIs de alto nível e os comandos Amazon S3 de alto nível na CLI da AWS (aws s3 cp, aws s3 mv e aws s3 sync), o upload de várias partes é executado automaticamente para objetos grandes.

Replicação entre regiões

A replicação entre regiões é um recurso do Amazon S3 que permite replicar de forma assíncrona todos os novos objetos no bucket de origem em uma região da AWS para um bucket de destino em outra região.

Todos os metadados e ACLs associados ao objeto também fazem parte da replicação.

Depois de configurar a replicação entre regiões no bloco de origem, qualquer alteração nos dados, metadados ou ACLs em um objeto aciona uma nova replicação no bloco de destino.

Para habilitar a replicação entre regiões, o controle de versão deve estar ativado para os buckets de origem e de destino e você deve usar uma política do IAM para conceder ao Amazon S3 permissão para replicar objetos em seu nome. A replicação entre regiões é comumente usada para reduzir a latência necessária para acessar objetos no Amazon S3, colocando objetos mais próximos de um conjunto de usuários ou para atender aos requisitos de armazenamento de dados de backup a uma certa distância dos dados de origem originais.

Se ativada em um bucket existente, a replicação entre regiões replicará apenas novos objetos. Os objetos existentes não serão replicados e devem ser copiados para o novo bucket por meio de um comando separado.

Logging

Para rastrear solicitações ao seu bucket do Amazon S3, você pode habilitar os logs de acesso do servidor Amazon S3.

O log está desativado por padrão, mas pode ser facilmente ativado. Quando você habilita o log para um bucket (o bucket de origem), deve escolher onde os logs serão armazenados (o bucket de destino).

Você pode armazenar logs de acesso no mesmo bucket ou em um bucket diferente. De qualquer forma, é opcional (mas uma prática recomendada) especificar um prefixo, como logs / ou nomedoseubucket / logs /, para que você possa identificar seus logs com mais facilidade.

Uma vez ativado, os logs são entregues com o máximo de esforço, com um pequeno atraso. Os logs incluem informações como:

- Conta do solicitante e endereço IP
- Nome do bloco
- Tempo de solicitação
- Ação (GET, PUT, LIST e assim por diante)
- Status da resposta ou código de erro
- Notificações de Eventos

Notificações por Eventos

As notificações de eventos do Amazon S3 podem ser enviadas em resposta a ações executadas em objetos carregados ou armazenados no Amazon S3. As notificações de eventos permitem executar fluxos de trabalho, enviar alertas ou executar outras ações em resposta a alterações em seus objetos armazenados no Amazon S3.

Você pode usar as notificações de eventos do Amazon S3 para configurar acionadores para executar ações, como transcodificar arquivos de mídia ao serem carregados, processar arquivos de dados quando estiverem disponíveis e sincronizar objetos do Amazon S3 com outros armazenamentos de dados.

As notificações de eventos do Amazon S3 são configuradas no nível do bucket e você pode configurá-las por meio do console do Amazon S3, da API REST ou do AWS SDK. O Amazon S3 pode publicar notificações quando novos objetos são criados (por uma conclusão de upload PUT, POST, COPY ou multipart), quando objetos são removidos (por um DELETE) ou quando o Amazon S3 detecta que um objeto RRS foi perdido.

Você também pode configurar notificações de eventos com base no nome do objeto prefixos e sufixos. As mensagens de notificação podem ser enviadas por meio do Amazon Simple Notification Service (Amazon SNS) ou do Amazon Simple Queue Service (Amazon SQS) ou entregues diretamente à AWS Lambda para invocar as funções da AWS Lambda.

Melhores Práticas, Padrões e Desempenho

É um padrão comum usar o armazenamento Amazon S3 em ambientes e aplicativos híbridos de TI.

Por exemplo, os dados em sistemas de arquivos locais, bancos de dados e arquivos de conformidade podem ser facilmente armazenados em backup pela Internet no Amazon S3 ou Amazon Glacier, enquanto o armazenamento principal do aplicativo ou banco de dados permanece no local.

Outro padrão comum é usar o Amazon S3 como armazenamento "blob" em massa para dados, mantendo um índice para esses dados em outro serviço, como o

Amazon DynamoDB ou o Amazon RDS. Isso permite pesquisas rápidas e consultas complexas nos nomes das chaves, sem listar as chaves continuamente.

O Amazon S3 será escalado automaticamente para suportar taxas de solicitação muito altas, reparticionando automaticamente seus buckets conforme necessário.

Se você precisar de taxas de solicitação superiores a 100 solicitações por segundo, convém revisar as diretrizes de boas práticas do Amazon S3 no Guia do desenvolvedor. Para oferecer suporte a taxas de solicitação mais altas, é melhor garantir algum nível de distribuição aleatória de chaves, por exemplo, incluindo um hash como prefixo dos nomes das chaves.

A Geleira da Amazon (Amazon Glacier)

O Amazon Glacier é um serviço de armazenamento de custo extremamente baixo que fornece armazenamento durável, seguro e flexível para arquivamento de dados e backup online.

Para manter os custos baixos, o Amazon Glacier foi projetado para dados acessados com pouca frequência, onde é necessário um tempo de recuperação de três a cinco horas aceitável. O Amazon Glacier pode armazenar uma quantidade ilimitada de praticamente qualquer tipo de dados, em qualquer formato.

Os casos de uso comuns do Amazon Glacier incluem a substituição de soluções de fita tradicionais para backup e arquivamento de longo prazo e armazenamento de dados necessários para fins de conformidade.

Na maioria dos casos, os dados armazenados no Amazon Glacier consistem em grandes arquivos TAR (Tape Archive) ou ZIP.

Como o Amazon S3, o Amazon Glacier é extremamente durável, armazenando dados em vários dispositivos em várias instalações de uma região. O Amazon Glacier foi projetado para oferecer uma durabilidade de 99,999999999% dos objetos em um determinado ano.

Arquivos

No Amazon Glacier, os dados são armazenados em arquivos. Um arquivo pode conter até 40 TB de dados e você pode ter um número ilimitado de arquivos.

Cada arquivo é atribuído um ID de arquivo único no momento da criação. (Ao contrário de uma chave de objeto do Amazon S3, você não pode especificar um código de nome do arquivo.) Todos os arquivos são criptografados automaticamente e os arquivos são imutáveis - depois que um arquivo é criado, ele não pode ser modificado.

Vaults

Os cofres são recipientes para arquivos. Cada conta da AWS pode ter até 1.000 cofres. Você pode controlar o acesso aos seus cofres e às ações permitidas usando políticas do IAM ou políticas de acesso ao cofre.

Vaults Locks

Você pode implantar e aplicar facilmente controles de conformidade para cofres individuais do Amazon Glacier com uma política de bloqueio de cofre. Você pode especificar controles como Write Once Read Many (WORM) em uma política de bloqueio de cofre e bloqueie a política de edições futuras. Uma vez bloqueada, a política não pode mais ser alterada.

Recuperação de dados

Você pode recuperar até 5% dos seus dados armazenados no Amazon Glacier gratuitamente a cada mês, calculados em uma base proporcional diária. Se você recuperar mais de 5%, incorrerá em taxas de recuperação com base na sua taxa máxima de recuperação.

Para eliminar ou minimizar essas taxas, você pode definir uma política de recuperação de dados em um cofre para limitar suas recuperações ao nível gratuito ou a uma taxa de dados especificada.

Amazon Glacier versus Amazon Simple Storage Service (Amazon S3)

O Amazon Glacier é semelhante ao Amazon S3, mas difere em vários aspectos principais. O Amazon Glacier suporta arquivos de 40 TB versus objetos de 5 TB no Amazon S3.

Os arquivos no Amazon Glacier são identificados por IDs de arquivamento gerados pelo sistema, enquanto o Amazon S3 permite usar nomes de chaves "amigáveis".

Os arquivos do Amazon Glacier são criptografados automaticamente, enquanto a criptografia em repouso é opcional no Amazon S3.

No entanto, usando o Amazon Glacier como uma classe de armazenamento do Amazon S3, juntamente com as políticas do ciclo de vida do objeto, você pode usar a interface do Amazon S3 para obter a maioria dos benefícios do Amazon Glacier sem aprender uma nova interface.

EC2 e EBS

Amazon Elastic Compute Cloud (Amazon EC2)

O Amazon EC2 é o principal serviço da Web da AWS que fornece capacidade de computação redimensionável na nuvem.

Noções básicas de Computação

Computação refere-se à quantidade de energia computacional necessária para atender sua carga de trabalho. Se sua carga de trabalho é muito pequena, como um site que recebe poucos visitantes, suas necessidades de computação são muito pequenas.

Uma grande carga de trabalho, como a seleção de dez milhões de compostos contra um alvo comum de câncer, pode exigir muita computação. A quantidade de computação necessária pode mudar drasticamente ao longo do tempo.

O Amazon EC2 permite adquirir computação por meio do lançamento de servidores virtuais chamados instâncias. Ao iniciar uma instância, você pode usar a computação como desejar, da mesma forma que faria com um servidor local.

Como você está pagando pelo poder de computação da instância, você é cobrado por hora enquanto a instância está em execução. Quando você interrompe a instância, não é mais cobrado.

Existem dois conceitos essenciais para iniciar instâncias na AWS: (1) a quantidade de hardware virtual dedicado à instância e (2) o software carregado na instância.

Essas duas dimensões de novas instâncias são controladas, respectivamente, pelo tipo de instância e pela AMI.

Tipos de Instância

O tipo de instância define o hardware virtual que suporta uma instância do Amazon EC2. Existem dezenas de tipos de instância disponíveis, variando nas seguintes dimensões:

- CPUs virtuais (vCPUs)
- Memória
- Armazenamento (tamanho e tipo)
- Desempenho de rede

Os tipos de instância são agrupados em famílias com base na proporção desses valores entre si. Por exemplo, a **família m5** fornece um equilíbrio de recursos de computação, memória e rede, e é uma boa opção para muitos aplicativos.

Dentro de cada família, existem várias opções que aumentam linearmente de tamanho.

Tamanho de instância	vCPU	Memória (GiB)	Armazenamento de instâncias (GiB)	Largura de banda de rede (Gbps)	Largura de banda do EBS (Mbps)
m5.large	2	8	Somente EBS	Até 10	Até 4.750
m5.xlarge	4	16	Somente EBS	Até 10	Até 4.750
m5.2xlarge	8	32	Somente EBS	Até 10	Até 4.750
m5.4xlarge	16	64	Somente EBS	Até 10	4.750
m5.8xlarge	32	128	Somente EBS	10	6.800
m5.12xlarge	48	192	Somente EBS	10	9.500
m5.16xlarge	64	256	Somente EBS	20	13.600
m5.24xlarge	96	384	Somente EBS	25	19.000

Observe que a proporção de vCPUs para a memória é constante conforme os tamanhos são dimensionados linearmente.

O preço por hora de cada tamanho também é linear. Por exemplo, uma instância m5.xlarge custa duas vezes mais que a instância m5.large.

Diferentes famílias de tipos de instância inclinam a proporção para acomodar diferentes tipos de cargas de trabalho, mas todas exibem esse comportamento linear de expansão dentro da família. Em resposta à demanda do cliente e para tirar proveito da nova tecnologia de processador, a AWS ocasionalmente apresenta novas famílias de instâncias.

Verifique o site da AWS para a lista atual: https://aws.amazon.com/pt/ec2/instance-types/

Outra variável a considerar ao escolher um tipo de instância é o desempenho da rede. Para a maioria dos tipos de instância, a AWS publica uma medida relativa do desempenho da rede: baixo, moderado ou alto. Alguns tipos de instância especificam um desempenho de rede de 10 Gbps. O desempenho da rede aumenta dentro de uma família à medida que o tipo de instância aumenta.

Para cargas de trabalho que exigem maior desempenho de rede, muitos tipos de instância suportam rede aprimorada. Rede aprimorada reduz o impacto da virtualização no desempenho de rede, ativando um recurso chamado Virtualização de E / S de Raiz Única (SR-IOV). Isso resulta em mais pacotes por segundo (PPS), menor latência e menos instabilidade.

A ativação da rede aprimorada em uma instância envolve garantir que os drivers corretos estejam instalados e modificar um atributo da instância.

AMIs (Amazon Machine Images)

A Amazon Machine Image (AMI) define o software inicial que estará em uma instância quando for lançado. Uma AMI define todos os aspectos do estado do software no lançamento da instância, incluindo:

- O sistema operacional (SO) e sua configuração
- O estado inicial de qualquer correção
- Aplicativo ou software do sistema

Todas as AMIs são baseadas em sistemas operacionais x86, Linux ou Windows.

Existem quatro fontes de AMIs:

Publicado pela AWS

A AWS publica AMIs com versões de muitos sistemas operacionais diferentes, Linux e Windows. Isso inclui várias distribuições do Linux (incluindo Ubuntu, Red Hat e distribuição própria da Amazon) e Windows 2008 e Windows 2012.

Iniciar uma instância com base em uma dessas AMIs resultará nas configurações padrão do sistema operacional, semelhante à instalação de um sistema operacional a partir da imagem ISO do sistema operacional padrão.

Como em qualquer instalação do sistema operacional, você deve aplicar imediatamente todos os patches apropriados ao iniciar.

AWS Marketplace

O AWS Marketplace é uma loja online que ajuda os clientes a encontrar, comprar e começar imediatamente a usar o software e os serviços executados no Amazon EC2. Muitos parceiros da AWS disponibilizaram seu software no AWS Marketplace.

Isso oferece dois benefícios: o cliente não precisa instalar o software e o contrato de licença é apropriado para a nuvem.

As instâncias iniciadas a partir de uma AMI do AWS Marketplace incorrem no custo horário padrão do tipo de instância, além de uma cobrança adicional por hora pelo software adicional (alguns AWS Marketplace de código aberto são pacotes que não têm custo adicional de software).

Gerado a partir de instâncias existentes

Uma AMI pode ser criada a partir de uma instância existente do Amazon EC2.

Essa é uma fonte muito comum de AMIs. Os clientes iniciam uma instância a partir de uma AMI publicada e, em seguida, a instância é configurada para atender a todos os padrões corporativos do cliente para atualizações, gerenciamento, segurança e assim por diante.

Uma AMI é gerada a partir da instância configurada e usada para gerar todas as instâncias desse SO. Dessa forma, todas as novas instâncias seguem o padrão corporativo e é mais difícil para projetos individuais iniciar instâncias não conformes.

Servidores virtuais carregados (Uploaded)

Usando o serviço AWS VM Import / Export, os clientes podem criar imagens a partir de vários formatos de virtualização, incluindo RAW, VHD, VMDK e OVA.

A lista atual de sistemas operacionais suportados (Linux e Windows) pode ser encontrada na documentação da AWS. Cabe aos clientes permanecer em conformidade com os termos de licenciamento do fornecedor do SO.

Usando uma instância com segurança

Uma vez iniciadas, as instâncias podem ser gerenciadas pela Internet. A AWS possui vários serviços e recursos para garantir que esse gerenciamento possa ser feito de maneira simples e segura.

Endereçando uma Instância

Existem várias maneiras pelas quais uma instância pode ser endereçada na Web após a criação:

- Nome do sistema de nome de domínio público (DNS) - Quando você inicia uma instância, a AWS cria um nome DNS que pode ser usado para acessar a instância. Este nome DNS é gerado automaticamente e não pode ser especificado pelo cliente. O nome pode ser encontrado na guia Descrição do AWS Management Console ou na interface da linha de comando (CLI) ou

Interface de programação de aplicativos (API). Esse nome DNS persiste apenas enquanto a instância está em execução e não pode ser transferido para outra instância.

- IP público - uma instância iniciada também pode ter um endereço IP público atribuído. Este endereço IP é atribuído a partir dos endereços reservados pela AWS e não pode ser especificado. Esse endereço IP é exclusivo na Internet, persiste apenas enquanto a instância está em execução e não pode ser transferido para outra instância.
- IP elástico - um endereço IP elástico é um endereço exclusivo na Internet que você reserva independentemente e associa a uma instância do Amazon EC2. Embora parecido com um IP público, existem algumas diferenças importantes. Esse endereço IP persiste até que o cliente o libere e não esteja vinculado ao tempo de vida ou estado de uma instância individual. Como ele pode ser transferido para uma instância de substituição no caso de uma falha da instância, é um endereço público que pode ser compartilhado externamente sem acoplar clientes a uma instância específica.

Acesso Inicial

O Amazon EC2 usa criptografia de chave pública para criptografar e descriptografar informações de login. A criptografia de chave pública usa uma chave pública para criptografar um dado e uma chave privada associada para descriptografar os dados.

Essas duas chaves juntas são chamadas de par de chaves. Os pares de chaves podem ser criados por meio do AWS Management Console, CLI ou API ou os clientes podem fazer upload de seus próprios pares de chaves. A AWS armazena a chave pública e a chave privada é mantida pelo cliente. A chave privada é essencial para adquirir acesso seguro a uma instância pela primeira vez.

Armazene suas chaves privadas com segurança. Quando o Amazon EC2 inicia uma instância do Linux, a chave pública é armazenada no arquivo /.ssh/authorized_keys na instância e um usuário inicial é criado.

O usuário inicial pode variar dependendo do sistema operacional. Por exemplo, o usuário inicial da distribuição do Amazon Linux é ec2-user. O acesso inicial à instância é obtido usando o usuário ec2 e a chave privada para efetuar login via SSH.

Nesse ponto, você pode configurar outros usuários e se inscrever em um diretório como LDAP.

Ao iniciar uma instância do Windows, o Amazon EC2 gera uma senha aleatória para a conta de administrador local e criptografa a senha usando a chave pública. O acesso inicial à instância é obtido descriptografando a senha com a chave privada, no console ou através da API. A senha descriptografada pode ser usada para efetuar login na instância com a conta de administrador local via RDP. Neste ponto, você pode criar outros usuários locais e / ou conectar-se a um domínio do Active Directory.

Proteção de Firewall Virtual

A AWS permite controlar o tráfego de entrada e saída de suas instâncias por meio de firewalls virtuais chamados grupos de segurança. Os grupos de segurança permitem controlar o tráfego com base na porta, protocolo e origem / destino.

Os grupos de segurança têm recursos diferentes, dependendo de estarem associados a um Amazon VPC ou Amazon EC2-Classic. Grupos de segurança são associados a instâncias quando são iniciados. Cada instância deve ter pelo menos um grupo de segurança, mas pode ter mais.

Um grupo de segurança tem a permissão padrão de negação (deny). Isto é, não permite nenhum tráfego que não seja explicitamente permitido por uma regra de grupo de segurança.

Uma regra é definida pelos três atributos na tabela abaixo. Quando uma instância é associada a vários grupos de segurança, as regras são agregadas e todo o tráfego permitido por cada um dos grupos individuais é permitido.

Atributo	Significado
Porta	O número da porta afetada por esta regra. Por exemplo, porta 80 para HTTP.
Protocolo	O padrão de comunicação para o tráfego afetado por esta regra.
Origem/Destino	Identifica a outra extremidade da comunicação, a origem das regras de tráfego de entrada ou o destino das regras de tráfego de saída.

Um grupo de segurança é um firewall com estado. O que isso significa? Uma mensagem de saída é lembrada para que a resposta seja permitida através do grupo de segurança sem que uma regra de entrada explícita seja necessária.

Os grupos de segurança são aplicados no nível da instância, em oposição a um firewall local tradicional que protege no perímetro.

O efeito disso é que, em vez de ter que violar um único perímetro para acessar todas as instâncias no seu grupo de segurança, um invasor precisará violar o grupo de segurança repetidamente para cada instância individual.

O ciclo de vida das instâncias

O Amazon EC2 possui vários recursos e serviços que facilitam o gerenciamento de instâncias do Amazon EC2 durante todo o ciclo de vida.

Lançamento

Existem vários serviços adicionais que são úteis ao iniciar novas instâncias do Amazon EC2.

Bootstrapping

Um grande benefício da nuvem é a capacidade de criar scripts de gerenciamento de hardware virtual de uma maneira que não é possível com o hardware local.

Para perceber o valor disso, é necessário que haja alguma maneira de configurar instâncias e instalar aplicativos programaticamente quando uma instância é iniciada. O processo de fornecer código a ser executado em uma instância no lançamento é chamado de inicialização.

Um dos parâmetros quando uma instância é iniciada é um valor de sequência chamado UserData. Essa sequência é transmitida ao sistema operacional para ser executada como parte do processo de inicialização na primeira vez em que a instância é inicializada.

Nas instâncias do Linux, isso pode ser um shell script e no Windows, pode ser um script em estilo de lote ou um script do PowerShell. O script pode executar tarefas como:

- Aplicando patches e atualizações no sistema operacional
- Registrando em um serviço de diretório
- Instalando o software aplicativo
- Copiando um script ou programa mais longo do armazenamento para ser executado na instância
- Instalando Chef ou Puppet e atribuindo uma função à instância para que o software de gerenciamento de configuração possa configurar a instância

UserData é armazenado com a instância e não é criptografado, portanto, é importante não incluir segredos, como senhas ou chaves, nos UserData.

Importação / Exportação de VM

Além de importar instâncias virtuais como AMIs, a Importação / Exportação de VM permite importar facilmente Máquinas Virtuais (VMs) do seu ambiente existente

como uma instância do Amazon EC2 e exportá-las de volta para o seu ambiente local.

Você pode exportar apenas instâncias do Amazon EC2 importadas anteriormente. Instâncias iniciadas na AWS pelas AMIs não podem ser exportadas.

Metadados da instância

Os metadados da instância são dados sobre sua instância que você pode usar para configurar ou gerenciar a instância em execução. Isso é único, pois é um mecanismo para obter as propriedades da instância da AWS de dentro do sistema operacional sem fazer uma chamada para a API da AWS.

Uma chamada HTTP para http://169.254.169.254/latest/meta-data/ retornará o nó superior da árvore de metadados da instância.

Os metadados da instância incluem uma ampla variedade de atributos, incluindo:

- Os grupos de segurança associados
- O ID da instância
- O tipo de instância
- A AMI usada para iniciar a instância

Isso começa apenas a arranhar a superfície das informações disponíveis nos metadados. Consulte a documentação da AWS para obter uma lista completa.

Gerenciando instâncias

Quando o número de instâncias na sua conta começa a subir, pode ser difícil administrá-las. As tags podem ajudar a gerenciar não apenas suas instâncias do Amazon EC2, mas também muitos de seus serviços da AWS.

Tags são pares de chave / valor que você pode associar à sua instância ou outro serviço. As tags podem ser usadas para identificar atributos de uma instância, como projeto, ambiente (desenvolvedor, teste etc.), departamento faturável etc. Você pode aplicar até 50 tags por instância.

Monitoramento de Instâncias

A AWS oferece um serviço chamado Amazon CloudWatch que fornece monitoramento e alerta para instâncias do Amazon EC2 e também outros serviços de infraestrutura da AWS.

Modificando uma Instância

Existem vários aspectos de uma instância que podem ser modificados após o lançamento.

Tipo de instância

A capacidade de alterar o tipo de instância contribui muito para a agilidade da execução de cargas de trabalho na nuvem. Em vez de se comprometer com uma determinada configuração de hardware meses antes do lançamento de uma carga de trabalho, a carga de trabalho pode ser iniciada usando uma melhor estimativa para o tipo de instância.

Se a computação precisar ser maior ou menor que o esperado, as instâncias poderão ser alteradas para um tamanho diferente, mais apropriado para a carga de trabalho.

As instâncias podem ser redimensionadas usando o AWS Management Console, a CLI ou a API. Para redimensionar uma instância, defina o estado como parado.

Escolha a função "Alterar tipo de instância" na ferramenta de sua escolha (o tipo de instância é listado como uma Configuração da instância no console e um Atributo da instância na CLI) e selecione o tipo de instância desejado. Reinicie a instância e o processo está completo.

Grupos de Segurança

Se uma instância estiver sendo executada em um Amazon VPC, você poderá alterar quais grupos de segurança estão associados a uma instância enquanto ela estiver em execução.

Para instâncias fora de um Amazon VPC (chamado EC2-Classic), a associação dos grupos de segurança não pode ser alterada após o lançamento.

Proteção de Exclusão (Termination Protection)

Quando uma instância do Amazon EC2 não é mais necessária, o estado pode ser definido como terminado e a instância será encerrada e removida da infraestrutura da AWS.

Para impedir a rescisão por meio do AWS Management Console, CLI ou API, a proteção de rescisão pode ser ativada para uma instância.

Enquanto ativada, as chamadas para finalizar a instância falharão até que a proteção de finalização seja desativada. Isso ajuda a impedir o encerramento acidental por erro humano.

Observe que isso apenas protege contra chamadas de término do console de gerenciamento, CLI ou API da AWS. Ele não impede a finalização acionada por um comando de encerramento do SO, a finalização de um grupo de Auto Scaling ou a finalização de uma Instância Spot devido a alterações no preço Spot.

Opções das Instâncias

Existem várias opções adicionais disponíveis no Amazon EC2 para melhorar a otimização de custos, segurança e desempenho que são importantes para o seu conhecimento.

Opções de preços

Você é cobrado pelas instâncias do Amazon EC2 por cada hora em que elas estão no estado de execução, mas o valor cobrado por hora pode variar com base em

três opções de preço: Instâncias sob demanda, Instâncias reservadas e Instâncias spot.

Instâncias sob demanda

O preço por hora para cada tipo de instância publicado no site da AWS representa o preço das Instâncias sob demanda. Essa é a opção de preço mais flexível, pois não requer compromisso inicial e o cliente tem controle sobre quando a instância é iniciada e quando é encerrada.

É a menos econômica das três opções de preço por hora de computação, mas sua flexibilidade permite que os clientes economizem ao provisionar um nível variável de computação para cargas de trabalho imprevisíveis.

Instâncias reservadas

A opção de preços de Instância reservada permite que os clientes façam reservas de capacidade para cargas de trabalho previsíveis. Ao usar Instâncias reservadas para essas cargas de trabalho, os clientes podem economizar até 75% sobre a taxa horária sob demanda.

Ao comprar uma reserva, o cliente especifica o tipo de instância e a Zona de Disponibilidade para essa Instância Reservada e obtém um preço horário efetivo mais baixo para essa instância durante a duração da reserva.

Um benefício adicional é que a capacidade nos datacenters da AWS é reservado para esse cliente. Existem dois fatores que determinam o custo da reserva: o termo compromisso e a opção de pagamento.

O termo compromisso é a duração da reserva e pode ser de um a três anos. Quanto maior o compromisso, maior o desconto.

Existem três opções de pagamento diferentes para instâncias reservadas:

- Tudo adiantado - pague antecipadamente por toda a reserva. Não há cobrança mensal para o cliente durante o período.
- Pagamento antecipado parcial - pague uma parte da cobrança da reserva antecipadamente e o restante em parcelas mensais pela duração do prazo.
- Sem adiantamento - pague o valor total da reserva em prestações mensais pela duração do prazo.

O valor do desconto é maior quanto mais o cliente paga antecipadamente.

Quando suas necessidades de computação mudam, você pode modificar suas instâncias reservadas e continuar a se beneficiar da sua reserva de capacidade.

A modificação não altera o prazo restante de suas instâncias reservadas, e as datas de término permanecem as mesmas. Não há taxa e você não recebe novas faturas ou faturas.

A modificação é separada da compra e não afeta a maneira como você usa, compra ou vende instâncias reservadas. Você pode modificar toda a sua reserva, ou apenas um subconjunto, de uma ou mais das seguintes maneiras:

Instâncias Spot

Para cargas de trabalho que não exigem tempo crítico e são tolerantes à interrupção, as Instâncias Spot oferecem o maior desconto. Com Instâncias Spot, os clientes especificam o preço que estão dispostos a pagar por um determinado tipo de instância.

Quando o preço de oferta do cliente estiver acima do preço Spot atual, o cliente receberá as instâncias solicitadas.

Essas instâncias funcionarão como todas as outras instâncias do Amazon EC2, e o cliente pagará apenas o preço spot pelas horas em que essas instâncias forem executadas.

As instâncias serão executadas até:

- O cliente as encerra.
- O preço spot fica acima do preço de oferta do cliente.
- Não há capacidade não utilizada suficiente para atender à demanda por instâncias spot.

Se o Amazon EC2 precisar encerrar uma Instância spot, a instância receberá um aviso de rescisão fornecendo um aviso de dois minutos antes do Amazon EC2 encerrar a instância.

Devido à possibilidade de interrupção, as Instâncias spot devem ser usadas apenas para cargas de trabalho tolerantes à interrupção. Isso pode incluir análises, modelagem financeira, big data, codificação, computação científica e testes.

Arquiteturas com Diferentes Modelos de Preços

Para modelos de arquiteturas complexos, que podem ter vários níveis e varias camadas para sua aplicação, é importante saber tirar proveito dos diferentes modelos de preços para criar uma arquitetura econômica.

Essa arquitetura pode incluir modelos de preços diferentes na mesma carga de trabalho. Por exemplo, um site que calcula a média de 5.000 visitas por dia, mas aumenta até 20.000 visitas por dia durante picos periódicos, pode comprar duas instâncias reservadas para lidar com o tráfego médio, mas depende de instâncias sob demanda para atender às necessidades de computação durante os horários de pico.

Opções de locação

Existem várias opções de locação para instâncias do Amazon EC2 que podem ajudar os clientes a atingir metas de segurança e conformidade.

- Locação compartilhada - A locação compartilhada é o modelo de locação padrão para todas as instâncias do Amazon EC2, independentemente do tipo de instância, modelo de precificação etc. Locação compartilhada significa que uma única máquina host pode hospedar instâncias de diferentes

clientes. Como a AWS não usa superprovisionamento e isola completamente instâncias de outras instâncias no mesmo host, esse é um modelo de locação segura.

- Instâncias dedicadas - As instâncias dedicadas são executadas em hardware dedicado a um único cliente. À medida que um cliente executa mais Instâncias Dedicadas, mais hardware subjacente pode ser dedicado à sua conta. Outras instâncias da conta (aquelas não designadas como dedicadas) serão executadas em locação compartilhada e serão isoladas no nível do hardware das Instâncias Dedicadas da conta.
- Host dedicado - Um host dedicado do Amazon EC2 é um servidor físico com capacidade de instância do Amazon EC2 totalmente dedicado ao uso de um único cliente. Hosts dedicados podem ajudá-lo a atender aos requisitos de licenciamento e reduzir custos, permitindo que você use o servidor existente em licenças de software. O cliente tem controle total sobre qual host específico executa em uma instância no lançamento. Isso difere das instâncias dedicadas, pois uma instância dedicada pode ser iniciada em qualquer hardware dedicado à conta.

Grupos de canais (Placement Groups)

Um grupo de canais é um agrupamento lógico de instâncias em uma única zona de disponibilidade. Os grupos de canais permitem que os aplicativos participem de uma rede de baixa latência e 10 Gbps.

Os grupos de canais são recomendados para aplicativos que se beneficiam de baixa latência da rede, alto rendimento da rede ou ambos. Lembre-se de que isso representa conectividade de rede entre instâncias. Para usar totalmente esse desempenho de rede para o seu grupo de canais, escolha um tipo de instância que suporte rede aprimorada e desempenho de rede de 10 Gbps.

Repositórios de Instâncias

Um armazenamento de instância (às vezes chamado de armazenamento efêmero) fornece armazenamento temporário em nível de bloco para sua instância.

Esse armazenamento está localizado em discos fisicamente conectados ao computador host. Um armazenamento de instância é ideal para armazenamento temporário de informações que muda frequentemente, como buffers, caches, dados temporários e outro conteúdo temporário ou para dados replicados em uma frota de instâncias, como um pool de servidores da Web com balanceamento de carga. O tamanho e o tipo de repositórios de instâncias disponíveis com uma instância do Amazon EC2 dependem do tipo de instância.

O tipo de instância também determina o tipo de hardware para os volumes de armazenamento da instância. Enquanto alguns fornecem repositórios de instâncias da unidade de disco rígido (HDD), outros tipos de instâncias usam unidades de estado sólido (SSDs) para fornecer altíssimas taxas de desempenho de E / S aleatório.

Os repositórios de instância estão incluídos no custo de uma instância do Amazon EC2, portanto, são uma solução muito econômica para cargas de trabalho apropriadas. O aspecto principal dos armazenamentos de instância é que eles são temporários. Os dados no armazenamento da instância são perdidos quando:

- A unidade de disco subjacente falha.
- A instância para (os dados persistirão se uma instância reiniciar).
- A instância termina.

Portanto, não confie nos repositórios de instâncias para obter dados valiosos e de longo prazo. Em vez disso, crie um grau de redundância via RAID ou use um sistema de arquivos que suporte redundância e tolerância a falhas, como o HDFS do Hadoop.

Faça backup dos dados para soluções de armazenamento de dados mais duráveis, como no Amazon Simple Storage Service (Amazon S3) ou Amazon EBS com frequência suficiente para atender aos objetivos do ponto de recuperação.

Amazon Elastic Block Store (Amazon EBS)

Embora os repositórios de instâncias sejam uma maneira econômica de atender às cargas de trabalho apropriadas, sua persistência limitada os torna inadequados para muitas outras cargas de trabalho.

Para cargas de trabalho que exigem armazenamento em bloco mais durável, a Amazon fornece o Amazon EBS.

Basico do Elastic Block Store

O Amazon EBS fornece volumes persistentes de armazenamento em nível de bloco para uso com instâncias do Amazon EC2.

Cada volume do Amazon EBS é replicado automaticamente dentro da sua Zona de disponibilidade para protegê-lo contra falhas de componentes, oferecendo alta disponibilidade e durabilidade. Os volumes do Amazon EBS estão disponíveis em vários tipos que diferem nas características de desempenho e no preço.

Vários volumes do Amazon EBS podem ser anexados a uma única instância do Amazon EC2, embora um volume possa ser anexado apenas a uma única instância de cada vez.

Tipos de volumes do Amazon EBS

Os volumes do Amazon EBS estão disponíveis em vários tipos diferentes. Os tipos variam em áreas como hardware, desempenho e custo subjacentes. É importante conhecer as propriedades dos diferentes tipos para que você possa especificar o tipo mais econômico que atenda às demandas de desempenho.

Volumes Magnéticos

Os volumes magnéticos têm as características de desempenho mais baixas de todos os tipos de volume do Amazon EBS. Como tal, eles custam o menor por

gigabyte. Eles são uma solução excelente e econômica para cargas de trabalho apropriadas.

Um volume magnético do Amazon EBS pode variar de 1 GB a 1 TB e terá em média 100 IOPS, mas tem a capacidade de estourar para centenas de IOPS.

Eles são mais adequados para:

- Cargas de trabalho em que os dados são acessados com pouca frequência
- Leituras sequenciais
- Situações em que o armazenamento de baixo custo é um requisito

Os volumes magnéticos são cobrados com base na quantidade de espaço de dados provisionado, independentemente da quantidade de dados que você realmente armazena no volume.

SSD de uso geral

Os volumes SSD de uso geral oferecem armazenamento econômico, ideal para uma ampla gama de cargas de trabalho. Eles oferecem alto desempenho a um preço moderado, adequado para uma ampla variedade de cargas de trabalho.

Um volume SSD de uso geral pode variar de 1 GB a 16 TB e fornece um desempenho básico de três IOPS por gigabyte provisionado, chegando a 10.000 IOPS.

Por exemplo, se você provisionar um volume de 1 TB, poderá esperar um desempenho básico de 3.000 IOPS. Um volume de 5 TB não fornecerá uma linha de base de 15.000 IOPS, pois atingiria o limite de 10.000 IOPS.

Os volumes SSD de uso geral com menos de 1 TB também apresentam a capacidade de estourar até 3.000 IOPS por longos períodos de tempo. Por exemplo, se você tiver um volume de 500 GB, poderá esperar uma linha de base de 1.500 IOPS.

Sempre que você não estiver usando esses IOPS, eles serão acumulados como créditos de E / S. Quando seu volume tiver tráfego intenso, ele usará os créditos de E / S a uma taxa de até 3.000 IOPS até que estejam esgotados.

Nesse ponto, seu desempenho é revertido para 1.500 IOPS. Com 1 TB, o desempenho da linha de base do volume já está em 3.000 IOPS; portanto, o comportamento explosivo não se aplica.

Os volumes SSD de uso geral são cobrados com base na quantidade de espaço de dados provisionado, independentemente da quantidade de dados que você realmente armazena no volume.

Eles são adequados para uma ampla gama de cargas de trabalho em que o desempenho mais alto do disco não é crítico, como:

- Volumes de inicialização do sistema operacional
- Bancos de dados de pequeno a médio porte
- Ambientes de desenvolvimento e teste

SSD IOPS provisionado

Os volumes de SSD de IOPS provisionados são projetados para atender às necessidades de cargas de trabalho intensivas de E / S, particularmente cargas de trabalho de banco de dados sensíveis ao desempenho de armazenamento e consistência na taxa de transferência de E / S de acesso aleatório.

Embora sejam o tipo de volume mais caro do Amazon EBS por gigabyte, eles fornecem o desempenho mais alto de qualquer tipo de volume do Amazon EBS de maneira previsível. Um volume de SSD IOPS provisionado pode variar em tamanho de 4 GB a 16 TB.

Ao provisionar um volume SSD de IOPS provisionado, você especifica não apenas o tamanho, mas também o número desejado de IOPS, até o menor número máximo de 30 vezes o número de GB do volume, ou 20.000 IOPS.

Você pode distribuir vários volumes juntos em uma configuração RAID 0 para obter um tamanho maior e um desempenho maior. O Amazon EBS fornece 10% do desempenho IOPS provisionado em 99,9% do tempo em um determinado ano.

Os preços são baseados no tamanho do volume e na quantidade de IOPS reservada. O custo por gigabyte é um pouco maior que o dos volumes SSD de uso geral e é

aplicado com base no tamanho do volume, não na quantidade do volume usado para armazenar dados.

Uma taxa mensal adicional é aplicada com base no número de IOPS provisionadas, sejam elas consumidas ou não. Os volumes IOPS SSD provisionados oferecem alto desempenho previsível e são adequados para:

- Aplicativos de negócios críticos que exigem desempenho IOPS sustentado
- Cargas de trabalho de banco de dados grandes

No momento que estou escrevendo, a AWS lançou dois novos tipos de volume de disco rígido: HDD com taxa de transferência otimizada e HDD frio.

Com o tempo, espera-se que esses novos tipos eclipsem o atual tipo de volume magnético, atendendo às necessidades de qualquer carga de trabalho que exija desempenho do HDD.

Os volumes de HDD otimizados pela taxa de transferência são volumes de HDD de baixo custo projetados para acesso frequente, cargas de trabalho intensivas na taxa de transferência, como big data, data warehouses e processamento de logs.

Os volumes podem ter até 16 TB, com IOPS máximo de 500 e taxa de transferência máxima de 500 MB / s. Esses volumes são significativamente mais baratos que os Volumes SSD de uso geral.

Os volumes HDD frios são projetados para cargas de trabalho acessadas com menos frequência, como dados mais frios que exigem menos digitalizações por dia. Os volumes podem ter até 16 TB, com IOPS máximo de 250 e taxa de transferência máxima de 250 MB / s. Esses volumes são significativamente mais baratos que os volumes de HDD com taxa de transferência otimizada.

Instâncias otimizadas para Amazon EBS

Ao usar qualquer tipo de volume diferente de magnético e de E / S do Amazon EBS, é importante usar instâncias otimizadas do Amazon EBS para garantir que a

instância do Amazon EC2 esteja preparada para tirar proveito da E / S do volume do Amazon EBS.

Uma instância otimizada do Amazon EBS usa uma pilha de configuração otimizada e fornece capacidade dedicada adicional para E / S do Amazon EBS. Essa otimização fornece o melhor desempenho para os volumes do Amazon EBS, minimizando a contenção entre a E / S do Amazon EBS e outro tráfego da sua instância.

Ao selecionar o Amazon EBS otimizado para uma instância, você paga uma taxa horária adicional por essa instância. Verifique a documentação da AWS para confirmar quais tipos de instância estão disponíveis como instância otimizada para Amazon EBS.

Protegendo dados

Durante o ciclo de vida de um volume Amazon EBS, existem várias práticas e serviços que você deve conhecer ao provisionar seus discos.

Backup / Recuperação (Snapshots)

Você pode fazer backup dos dados nos volumes do Amazon EBS, independentemente do tipo de volume, tirando snapshots pontuais. Os snapshots são backups incrementais, o que significa que apenas os blocos no dispositivo que foram alterados desde que o snapshot mais recente foi salvo.

Tirando snapshots

Você pode tirar snapshots de várias maneiras:

- Por meio do AWS Management Console
- Através da CLI
- Através da API
- Configurando uma programação de capturas instantâneas regulares

Os dados para o snapshots são armazenados usando a tecnologia Amazon S3. A ação de tirar uma captura instantânea é gratuita. Você paga apenas os custos de armazenamento pelos dados da captura de snapshots.

Quando você solicita uma captura de snapshots, a captura point-in-time é criada imediatamente e o volume pode continuar sendo usado, mas a captura de snapshots pode permanecer no status pendente até que todos os blocos modificados tenham sido transferidos para o Amazon S3.

É importante saber que, embora os snapshots sejam armazenados usando a tecnologia Amazon S3, eles são armazenados no armazenamento controlado pela AWS e não nos buckets do Amazon S3 da sua conta. Isso significa que você não pode manipulá-los como outros objetos do Amazon S3. Em vez disso, você deve usar os recursos de captura de snapshots do Amazon EBS para gerenciá-los.

Os snapshots são restritos à região em que são criados, o que significa que você pode usá-los para criar novos volumes apenas na mesma região. Se você precisar restaurar um snapshot em uma região diferente, poderá copiar para outra região.

Criando um volume a partir de um snapshot

Para usar um snapshot, você cria um novo volume do Amazon EBS a partir deste snapshot. Quando você faz isso, o volume é criado imediatamente, mas os dados são carregados lentamente. Isso significa que o volume pode ser acessado na criação e, se os dados solicitados ainda não foram restaurados, eles serão restaurados na primeira solicitação.

Por esse motivo, é uma boa prática inicializar um volume criado a partir de uma captura de snapshot acessando todos os blocos no volume.

Os snapshots também podem ser usados para aumentar o tamanho de um volume do Amazon EBS. Para aumentar o tamanho de um volume do Amazon EBS, tire uma captura de snapshot do volume e crie um novo volume do tamanho desejado a partir da captura de snapshot. Substitua então o volume original pelo novo volume.

Recuperando Volumes

Como os volumes do Amazon EBS persistem além do tempo de vida de uma instância, é possível recuperar dados se uma instância falhar.

Se uma instância suportada pelo Amazon EBS falhar e houver dados sobre ele na unidade de inicialização, é relativamente simples desanexar o volume da instância.

A menos que o sinalizador DeleteOnTermination do volume tenha sido definido como falso, o volume deve ser desanexado antes que a instância seja encerrada. O volume pode então ser anexado como um volume de dados a outra instância e os dados lidos e recuperados.

Opções de Criptografia

Muitas cargas de trabalho têm requisitos para que os dados sejam criptografados em repouso, devido a regulamentos de conformidade ou padrões corporativos internos.

O Amazon EBS oferece criptografia nativa em todos os tipos de volume. Quando você inicia um volume criptografado do Amazon EBS, a Amazon usa o KMS (Serviço de Gerenciamento de Chaves da AWS) para lidar com o gerenciamento de chaves.

Uma nova chave mestra será criada, a menos que você selecione uma chave mestra criada separadamente no serviço. Seus dados e chaves associadas são criptografados usando o algoritmo AES-256 padrão do setor.

A criptografia ocorre nos servidores que hospedam instâncias do Amazon EC2, portanto, os dados são realmente criptografados em trânsito entre o host e a mídia de armazenamento e também na mídia.

A criptografia é transparente, portanto, todo acesso a dados é igual aos volumes não criptografados, e você pode esperar o mesmo desempenho de IOPS em volumes criptografados como faria com volumes não criptografados, com um efeito mínimo na latência. As capturas de snapshots obtidas de volumes

criptografados são criptografadas automaticamente, assim como os volumes criados a partir de capturas de snapshots criptografadas.

Amazon VPC

O Amazon VPC é a camada de rede do Amazon Elastic Compute Cloud (Amazon EC2) e permite que você crie sua própria rede virtual na AWS.

Você controla vários aspectos do seu Amazon VPC, incluindo a seleção de seu próprio intervalo de endereços IP, criando suas próprias sub-redes, e configurando suas próprias tabelas de rotas, gateways de rede e configurações de segurança.

Em uma região, você pode criar vários Amazon VPCs e cada Amazon VPC é logicamente isolado, mesmo que compartilhe seu espaço de endereço IP. Ao criar um Amazon VPC, você deve especificar o intervalo de endereços IPv4 escolhendo um bloco CIDR (Classless Inter-Domain Routing), como 10.0.0.0/16. O intervalo de endereços do Amazon VPC não pode ser alterado após a criação do Amazon VPC.

Um intervalo de endereços do Amazon VPC pode ser tão grande quanto / 16 (65.536 endereços disponíveis) ou tão pequeno quanto / 28 (16 endereços disponíveis) e não deve se sobrepor a nenhuma outra rede à qual eles devem ser conectados.

O serviço Amazon VPC foi lançado após o serviço Amazon EC2, por isso, existem duas plataformas de rede diferentes disponíveis na AWS: EC2-Classic e EC2-VPC.

O Amazon EC2 foi lançado originalmente com uma rede única e plana compartilhada com outros clientes da AWS, chamada EC2-Classic. Dessa forma, as contas da AWS criadas antes da chegada do serviço Amazon VPC podem iniciar instâncias na rede EC2-Classic e EC2-VPC.

As contas da AWS criadas após dezembro de 2013 suportam apenas o lançamento de instâncias usando o EC2-VPC. As contas da AWS que oferecem suporte ao EC2-VPC terão uma VPC padrão criada em cada região com uma sub-rede padrão criada em cada Zona de Disponibilidade. O bloco CIDR atribuído da VPC será 172.31.0.0/16.

Um Amazon VPC consiste nos seguintes componentes:

- Sub-redes
- Tabelas de rotas
- Conjuntos de opções do protocolo DHCP (Dynamic Host Configuration Protocol)
- Grupos de segurança
- Listas de controle de acesso à rede (ACLs)

Um Amazon VPC possui os seguintes componentes opcionais:

- Gateways da Internet (IGWs)
- Endereços de IP elástico (EIP)
- Interfaces de rede elástica (ENIs)
- Endpoints
- Peering
- Instâncias de conversão de endereço de rede (NATs) e gateways NAT
- Gateway Privado Virtual (VPG), Gateways de Cliente (CGWs) e Redes Privadas Virtuais (VPNs)

Sub-redes

Uma sub-rede é um segmento do intervalo de endereços IP de um Amazon VPC onde você pode iniciar instâncias do Amazon EC2, bancos de dados do Amazon Relational Database Service (Amazon RDS) e outros recursos da AWS. Os blocos CIDR definem sub-redes (por exemplo, 10.0.1.0/24 e 192.168.0.0/24).

A menor sub-rede que você pode criar é um / 28 (16 endereços IP). A AWS reserva os quatro primeiros endereços IP e o último endereço IP de cada sub-rede para fins de rede interna. Por exemplo, uma sub-rede definida como / 28 possui 16 endereços IP disponíveis; subtraia os 5 IPs necessários pela AWS para gerar 11 endereços IP para seu uso na sub-rede.

Após criar um Amazon VPC, você pode adicionar uma ou mais sub-redes em cada zona de disponibilidade. As sub-redes residem em uma zona de disponibilidade e não podem abranger zonas. Esse é um ponto importante que você deve entender. Lembre-se de que uma sub-rede é igual a uma zona de disponibilidade. No entanto, você pode ter várias sub-redes em uma zona de disponibilidade.

As sub-redes podem ser classificadas como públicas, privadas ou somente VPN. Uma sub-rede pública é aquela na qual a tabela de rotas associada (discutida posteriormente) direciona o tráfego da sub-rede para o IGW do Amazon VPC (também discutido mais adiante).

Uma sub-rede privada é aquela em que a tabela de rotas associada não direciona o tráfego da sub-rede para o IGW do Amazon VPC. Uma sub-rede somente VPN é aquela em que a tabela de rotas associada direciona o tráfego da sub-rede para o VPG do Amazon VPC (discutido mais adiante) e não possui uma rota para o IGW. Independentemente do tipo de sub-rede, o IP interno e seu intervalo de endereços da sub-rede é sempre privado (ou seja, não roteável na Internet).

Os Amazon VPCs padrão contêm uma sub-rede pública em todas as zonas de disponibilidade da região, com uma máscara de rede de / 20.

Tabelas de Rotas

Uma tabela de rota é uma construção lógica dentro de um Amazon VPC que contém um conjunto de regras (chamadas rotas) que são aplicadas à sub-rede e usadas para determinar para onde o tráfego de rede é direcionado. As rotas de uma tabela de rotas são o que permite que instâncias do Amazon EC2 em diferentes sub-redes de um Amazon VPC se comuniquem entre si.

Você pode modificar as tabelas de rotas e adicionar suas próprias rotas personalizadas. Você também pode usar tabelas de rotas para especificar quais sub-redes são públicas (direcionando o tráfego da Internet para o IGW) e quais sub-redes são privadas (não tendo uma rota que direcione o tráfego para o IGW).

Cada tabela de rota contém uma rota padrão chamada rota local, que permite a comunicação no Amazon VPC, e essa rota não pode ser modificada ou removida.

Rotas adicionais podem ser adicionadas ao tráfego direto para sair do Amazon VPC por meio do IGW (discutido mais tarde), do VPG (discutido mais tarde) ou da instância NAT (discutida mais adiante). Você deve se lembrar dos seguintes pontos sobre tabelas de rotas:

- Sua VPC possui um roteador implícito.
- Seu VPC vem automaticamente com uma tabela de rotas principal que você pode modificar.
- Você pode criar tabelas de rotas personalizadas adicionais para sua VPC.
- Cada sub-rede deve estar associada a uma tabela de rotas, que controla o roteamento da sub-rede. Se você não associar explicitamente uma sub-rede a uma tabela de rotas específica, a sub-rede usará a tabela de rotas principal.
- Você pode substituir a tabela de rotas principal por uma tabela personalizada criada para que cada nova sub-rede seja automaticamente associada a ela.
- Cada rota em uma tabela especifica um CIDR de destino e um destino; por exemplo, o tráfego destinado a 172.16.0.0/12 é direcionado para o VPG.
- A AWS usa a rota mais específica que corresponde ao tráfego para determinar como rotear o tráfego.

Gateways da Internet

Um Internet Gateway (IGW) é um componente do Amazon VPC redundante e altamente disponível que permite a comunicação entre instâncias do Amazon VPC e da Internet.

Um IGW fornece um destino nas tabelas de rotas do Amazon VPC para tráfego roteável pela Internet e executa a conversão de endereços de rede para instâncias às quais foram atribuídos endereços IP públicos. As instâncias do Amazon EC2 em uma VPC estão cientes apenas de seus endereços IP privados.

Quando o tráfego é enviado da instância para a Internet, o IGW converte o endereço de resposta no endereço IP público da instância (ou endereço EIP, coberto posteriormente) e mantém o mapa individual do endereço IP privado e do endereço IP público da instância.

Quando uma instância recebe tráfego da Internet, o IGW converte o endereço de destino (endereço IP público) no endereço IP privado da instância e encaminha o tráfego para a VPC.

Você deve fazer o seguinte para criar uma sub-rede pública com acesso à Internet:

- Anexe um IGW ao seu VPC.
- Crie uma regra da tabela de rota de sub-rede para enviar todo o tráfego não local (0.0.0.0/0) para o IGW.
- Configure as ACLs da rede e as regras do grupo de segurança para permitir que o tráfego relevante flua para e da sua instância.
- Você deve fazer o seguinte para permitir que uma instância do Amazon EC2 envie e receba tráfego da Internet:
    - Atribua um endereço IP público ou endereço EIP.

Você pode definir o escopo da rota para todos os destinos que não sejam explicitamente conhecidos na tabela de rotas (0.0.0.0/0), ou pode definir a rota para um intervalo mais restrito de endereços IP, como os endereços IP públicos dos endpoint públicos da sua empresa fora de Endereços AWS ou EIP de outras instâncias do Amazon EC2 fora do seu Amazon VPC.

DHCP (Dynamic Host Configuration Protocol)

O DHCP (Dynamic Host Configuration Protocol) fornece um padrão para passar informações de configuração para hosts em uma rede TCP / IP.

O campo de opções de uma mensagem DHCP contém os parâmetros de configuração. Alguns desses parâmetros são o nome de domínio, o servidor de nomes de domínio e o tipo netbios-node-type.

A AWS cria e associa automaticamente uma opção DHCP definida para o Amazon VPC na criação e define duas opções: servidores de nome de domínio (padrão no AmazonProvidedDNS) e nome de domínio (padrão no nome de domínio da sua região).

O AmazonProvidedDNS é um servidor DNS (Sistema de nomes de domínio da Amazon), e essa opção habilita o DNS para instâncias que precisam se comunicar pelo IGW do Amazon VPC. A opção DHCP define o elemento de um Amazon VPC permite direcionar as atribuições de nome de host do Amazon EC2 para seus próprios recursos.

Para atribuir seu próprio nome de domínio às suas instâncias, crie um conjunto de opções DHCP personalizado e atribua-o ao seu Amazon VPC.

Você pode configurar os seguintes valores em um conjunto de opções DHCP:

- servidores de nome de domínio - os endereços IP de até quatro servidores de nome de domínio, separados por vírgulas. O padrão é AmazonProvidedDNS.
- nome do domínio - especifique aqui o nome de domínio desejado (por exemplo, minhaempresa.com.br).
- servidores ntp - os endereços IP de até quatro servidores NTP (Network Time Protocol), separados por vírgulas
- netbios-name-servers - Os endereços IP de até quatro servidores de nomes NetBIOS, separados por vírgulas
- netbios-node-type - defina esse valor como 2.

Toda VPC deve ter apenas um conjunto de opções DHCP atribuído a ele.

Endereços IP Elásticos (EIPs)

A AWS mantém um conjunto de endereços IP públicos em cada região e os disponibiliza para você associar a recursos em seus Amazon VPCs.

Um endereço IP elástico (EIP) é um endereço IP público estático no pool para a região que você pode alocar para sua conta (extrair do pool) e liberar (retornar ao pool).

Os EIPs permitem manter um conjunto de endereços IP que permanecem fixos enquanto a infraestrutura subjacente pode mudar com o tempo. Aqui estão os pontos importantes para entender:

- Você deve primeiro alocar um EIP para uso em uma VPC e depois atribuí-lo a uma instância.
- Os EIPs são específicos para uma região (ou seja, um EIP em uma região não pode ser atribuído a uma instância dentro de um Amazon VPC em uma região diferente).
- Há um relacionamento individual entre interfaces de rede e EIPs.
- Você pode mover EIPs de uma instância para outra, no mesmo Amazon VPC ou em um Amazon VPC diferente na mesma região.
- Os EIPs permanecem associados à sua conta da AWS até que você os libere explicitamente.
- Existem cobranças por EIPs alocados à sua conta, mesmo quando não estão associados a um recurso.

Interfaces de Rede Elástica (ENIs)

Uma interface de rede elástica (ENI) é uma interface de rede virtual que você pode conectar a uma instância em um Amazon VPC.

As ENIs estão disponíveis apenas em uma VPC e são associadas a uma sub-rede na criação. Eles podem ter um endereço IP público e vários endereços IP privados.

Se houver vários endereços IP privados, um deles é primário. A atribuição de uma segunda interface de rede a uma instância por meio de uma ENI permite que ela seja de hospedagem dupla (tenha presença de rede em diferentes sub-redes).

Uma ENI criada independentemente de uma instância específica persiste, independentemente do tempo de vida de qualquer instância à qual está anexada. Se uma instância subjacente falhar, o endereço IP pode ser preservado anexando o ENI a uma instância de substituição. As ENIs permitem criar uma rede de gerenciamento, usar dispositivos de rede e segurança no Amazon VPC, criar instâncias de hospedagem dupla com cargas de trabalho / funções em sub-redes distintas ou criar uma solução de baixo orçamento e alta disponibilidade.

Endpoints

Um endpoint da Amazon VPC permite criar uma conexão privada entre o Amazon VPC e outro serviço da AWS sem exigir acesso pela Internet ou por meio de uma instância NAT, conexão VPN ou AWS Direct Connect.

Você pode criar vários endpoints para um único serviço e pode usar tabelas de rotas diferentes para impor políticas de acesso diferentes a partir de sub-redes diferentes para o mesmo serviço.

Atualmente, os endpoints da Amazon VPC oferecem suporte à comunicação com o Amazon Simple Storage Service (Amazon S3), e espera-se que outros serviços sejam adicionados no futuro. Você deve fazer o seguinte para criar um endpoint:

- Especifique o Amazon VPC.
- Especifique o serviço. Um serviço é identificado por uma lista de prefixos no formato com.amazonaws. <Região>. <Serviço>.
- Especifique a política. Você pode permitir acesso completo ou criar uma política personalizada. Esta política pode ser alterada a qualquer momento.

- Especifique as tabelas de rotas. Uma rota será adicionada a cada tabela de rotas especificada, que indicará o serviço como destino e o terminal como destino.

Peering

Uma conexão de emparelhamento do Amazon VPC é uma conexão de rede entre dois Amazon VPCs que permite que instâncias no Amazon VPC se comuniquem entre si como se estivessem na mesma rede.

Você pode criar uma conexão de peering do Amazon VPC entre o seu Amazon VPCs ou com um Amazon VPC em outra conta da AWS em uma única região.

Uma conexão de mesmo nível não é um gateway nem uma conexão VPN da Amazon e não apresenta um único ponto de falha para a comunicação. As conexões de peering são criadas por meio de um protocolo de solicitação / aceitação.

O proprietário do Amazon VPC solicitante envia uma solicitação ao mesmo nível para o proprietário do Amazon VPC do mesmo nível. Se o Amazon VPC de mesmo nível estiver na mesma conta, ele será identificado pelo seu ID de VPC. Se a VPC de mesmo nível for em uma conta diferente, ele é identificado pelo ID da conta e pelo ID da VPC.

O proprietário do Amazon VPC de mesmo nível tem uma semana para aceitar ou rejeitar a solicitação de mesmo nível com o Amazon VPC solicitante antes que a solicitação de peering expire.

Um Amazon VPC pode ter várias conexões de peering, e o emparelhamento é um relacionamento individual entre os Amazon VPCs, o que significa que dois Amazon VPCs não podem ter dois acordos de peering entre eles. Além disso, as conexões de mesmo nível não oferecem suporte ao roteamento transitivo.

Grupos de Segurança

Um grupo de segurança é um firewall com estado virtual que controla o tráfego de rede de entrada e saída para recursos da AWS e instâncias do Amazon EC2.

Todas as instâncias do Amazon EC2 devem ser iniciadas em um grupo de segurança. Se um grupo de segurança não for especificado na inicialização, a instância será iniciada no grupo de segurança padrão do Amazon VPC.

O grupo de segurança padrão permite a comunicação entre todos os recursos dentro do grupo de segurança, permite todo o tráfego de saída e nega todo o outro tráfego. Aqui estão os pontos importantes para entender sobre grupos de segurança:

- Você pode criar até 500 grupos de segurança para cada Amazon VPC.

- Você pode adicionar até 50 regras de entrada e 50 de saída a cada grupo de segurança. Se você precisar aplicar mais de 100 regras a uma instância, poderá associar até cinco grupos de segurança a cada interface de rede.
- Você pode especificar regras de permissão, mas não negar regras. Essa é uma diferença importante entre grupos de segurança e ACLs.
- Você pode especificar regras separadas para o tráfego de entrada e saída.
- Por padrão, nenhum tráfego de entrada é permitido até você adicionar regras de entrada ao grupo de segurança.
- Por padrão, os novos grupos de segurança têm uma regra de saída que permite todo o tráfego de saída.
- Você pode remover a regra e adicionar regras de saída que permitem apenas tráfego de saída específico.
- Grupos de segurança são stateful. Isso significa que as respostas ao tráfego de entrada permitido podem fluir de saída, independentemente das regras de saída e vice-versa. Essa é uma diferença importante entre grupos de segurança e ACLs de rede.

- As instâncias associadas ao mesmo grupo de segurança não podem se comunicar a menos que você adicione regras que permitam (com a exceção o grupo de segurança padrão).
- Você pode alterar os grupos de segurança aos quais uma instância está associada após o lançamento, e as alterações entrarão em vigor imediatamente.

Listas de Controle de Acesso à Rede (ACLs)

Uma lista de controle de acesso à rede (ACL) é outra camada de segurança que atua como um firewall sem estado no nível de sub-rede.

Uma ACL de rede é uma lista numerada de regras que a AWS avalia em ordem, começando com a regra numerada mais baixa, para determinar se o tráfego é permitido dentro ou fora de qualquer sub-rede associada à ACL da rede.

Os Amazon VPCs são criados com uma ACL de rede padrão modificável associada a todas as sub-redes que permitem todo o tráfego de entrada e saída.

Quando você cria uma ACL de rede personalizada, sua configuração inicial negará todo o tráfego de entrada e saída até você criar regras que permitam o contrário.

Você pode configurar ACLs de rede com regras semelhantes aos seus grupos de segurança para adicionar uma camada de segurança ao Amazon VPC ou pode optar por usar a ACL de rede padrão que não filtra o tráfego que atravessa o limite da sub-rede. No geral, todas as sub-redes devem estar associadas a uma ACL da rede.

Instâncias NAT e NAT Gateways

Por padrão, qualquer instância iniciada em uma sub-rede privada em um Amazon VPC não pode se comunicar com a Internet por meio do IGW. Isso é problemático se as instâncias nas sub-redes privadas precisarem de acesso direto à Internet a partir do Amazon VPC para aplicar atualizações de segurança, baixar patches ou atualizar o aplicativo.

A AWS fornece instâncias NAT e gateways NAT para permitir que instâncias implantadas em sub-redes privadas obtenham acesso à Internet.

Para casos de uso comuns, recomendamos que você use um gateway NAT em vez de uma instância NAT. O gateway NAT fornece melhor disponibilidade e maior largura de banda e requer menos esforço administrativo do que as instâncias NAT.

Instância NAT

Uma instância de conversão de endereço de rede (Instância NAT) é uma AMI (Amazon Machine Image) da Amazon Linux projetada para aceitar tráfego de instâncias em uma sub-rede privada, converter o endereço IP de origem no endereço IP público da instância NAT e encaminhar o tráfego para o IGW.

Além disso, a instância do NAT mantém o estado do tráfego encaminhado para retornar o tráfego de resposta da Internet para a instância apropriada na sub-rede privada. Essas instâncias têm a cadeia amzn-ami-vpc-nat em seus nomes, que pode ser pesquisada no console do Amazon EC2.

Para permitir que instâncias em uma sub-rede privada acessem recursos da Internet através do IGW por meio de uma instância NAT, faça o seguinte:

- Crie um grupo de segurança para o NAT com regras de saída que especificam os recursos necessários da Internet por porta, protocolo e endereço IP.
- Inicie uma AMI NAT do Amazon Linux como uma instância em uma sub-rede pública e associe-a ao grupo de segurança NAT.
- Desative o atributo Source / Destination Check do NAT.
- Configure a tabela de rotas associada a uma sub-rede privada para direcionar o tráfego vinculado à Internet para a instância NAT (por exemplo, i-1a2b3c4d).
- Aloque um EIP e associe-o à instância NAT.

Essa configuração permite que instâncias em sub-redes privadas enviem comunicação de saída da Internet, mas impede que as instâncias recebam tráfego de entrada iniciado por alguém na Internet.

Gateway NAT

Um gateway NAT é um recurso gerenciado da Amazon, projetado para operar como uma instância NAT, mas é mais simples de gerenciar e altamente disponível em uma Zona de Disponibilidade.

Para permitir que instâncias em uma sub-rede privada acessem os recursos da Internet por meio do IGW por meio de um gateway NAT, faça o seguinte:

- Configure a tabela de rotas associada à sub-rede privada para direcionar o tráfego vinculado à Internet para o gateway NAT (por exemplo, nat-1a2b3c4d).

- Aloque um EIP e associe-o ao gateway NAT.

Como uma instância NAT, esse serviço gerenciado permite a comunicação de saída da Internet e impede que as instâncias recebam tráfego de entrada iniciado por alguém na Internet.

Para criar uma arquitetura independente da zona de disponibilidade, crie um gateway NAT em cada zona de disponibilidade e configure seu roteamento para garantir que os recursos usem o gateway NAT na mesma zona de disponibilidade.

Gateways Privados Virtuais (VPGs), Gateways de Cliente (CGWs), e Redes Privadas Virtuais (VPNs)

Você pode conectar um datacenter existente ao Amazon VPC usando conexões VPN de hardware ou software, o que tornará o Amazon VPC uma extensão do datacenter.

O Amazon VPC oferece duas maneiras de conectar uma rede corporativa a um VPC: VPG e CGW.

Um gateway privado virtual (VPG) é o concentrador de rede virtual privada (VPN) no lado da AWS da conexão VPN entre as duas redes.

Um gateway de cliente (CGW) representa um dispositivo físico ou um aplicativo de software no lado do cliente da conexão VPN.

Após a criação desses dois elementos de um Amazon VPC, a última etapa é criar um túnel VPN. O túnel da VPN é estabelecido depois que o tráfego é gerado a partir do lado do cliente da conexão VPN. Você deve especificar o tipo de roteamento que planeja usar ao criar uma conexão VPN.

Se o CGW suportar o BGP (Border Gateway Protocol), configure a conexão VPN para roteamento dinâmico. Caso contrário, configure as conexões para roteamento estático. Se você estiver usando roteamento estático, deverá inserir as rotas da sua rede que devem ser comunicadas ao VPG.

As rotas serão propagadas para o Amazon VPC para permitir que seus recursos direcionem o tráfego de rede de volta para a rede corporativa através do VGW e através do túnel VPN.

O Amazon VPC também oferece suporte a vários CGWs, cada um com uma conexão VPN com um único VPG (design muitos-para-um). Para suportar essa topologia, os endereços IP do CGW devem ser exclusivos na região.

O Amazon VPC fornecerá as informações necessárias ao administrador da rede para configurar o CGW e estabelecer a conexão VPN com o VPG.

A conexão VPN consiste em dois túneis IPSec (Internet Protocol Security) para maior disponibilidade ao Amazon VPC. A seguir, estão os pontos importantes a serem entendidos sobre VPGs, CGWs e VPNs:

- O VPG é o final da AWS do túnel da VPN.
- O CGW é um aplicativo de hardware ou software no lado do cliente do túnel da VPN.
- Você deve iniciar o túnel VPN do CGW para o VPG.
- Os VPGs suportam roteamento dinâmico com BGP e roteamento estático.
- A conexão VPN consiste em dois túneis para maior disponibilidade da VPC.

Elastic Load Balancing, Amazon CloudWatch, e Auto Scaling

Balanceamento de Carga Elástico

Uma vantagem de ter acesso a um grande número de servidores na nuvem, como instâncias do Amazon EC2 na AWS, é a capacidade de fornecer uma experiência mais consistente para o usuário final.

Uma maneira de garantir consistência é equilibrar a carga da solicitação em mais de um servidor. Um balanceador de carga é um mecanismo que distribui automaticamente o tráfego entre várias instâncias do Amazon EC2.

Você pode gerenciar seus próprios balanceadores de carga virtuais nas instâncias do Amazon EC2 ou aproveitar um serviço da AWS chamado Elastic Load Balancing, que fornece um balanceador de carga gerenciado para você.

O serviço Elastic Load Balancing permite distribuir o tráfego entre um grupo de instâncias do Amazon EC2 em uma ou mais zonas de disponibilidade, permitindo obter alta disponibilidade em seus aplicativos. O Balanceamento de carga elástico oferece suporte ao roteamento e balanceamento de carga do HTTP (Hypertext Transfer Protocol), HTTPS (Hypertext Transfer Protocol Secure), Transmission Control Protocol (TCP) e Secure Sockets Layer (SSL) para instâncias do Amazon EC2.

O Elastic Load Balancing fornece um ponto de entrada único e estável de registro de nome canônico (CNAME) para a configuração do DNS (Sistema de Nome de Domínio) e suporta balanceadores de carga voltados para a Internet e internos para aplicativos.

O Elastic Load Balancing suporta verificações de integridade para instâncias do Amazon EC2 para garantir que o tráfego não seja roteado para instâncias não íntegras ou com falha. Além disso, o Elastic Load Balancing pode ser dimensionado automaticamente com base nas métricas coletadas.

Existem várias vantagens em usar o Elastic Load Balancing. Como o Elastic Load Balancing é um serviço gerenciado, ele entra e sai automaticamente para atender às demandas do aumento do tráfego de aplicativos e está altamente disponível na própria região como serviço.

O Elastic Load Balancing ajuda a obter alta disponibilidade para seus aplicativos, distribuindo o tráfego entre instâncias íntegras em várias zonas de disponibilidade. Além disso, o Elastic Load Balancing se integra perfeitamente ao serviço Auto Scaling para dimensionar automaticamente as instâncias do Amazon EC2 atrás do balanceador de carga.

Por fim, o Elastic Load Balancing é seguro, trabalhando com o Amazon Virtual Private Cloud (Amazon VPC) para rotear o tráfego internamente entre as camadas de aplicativos, permitindo expor apenas endereços IP públicos da Internet. O Elastic Load Balancing também suporta gerenciamento de certificado integrado e terminação SSL.

Tipos de Balanceadores de Carga

O Elastic Load Balancing fornece vários tipos de balanceadores de carga para lidar com diferentes tipos de conexões, incluindo balanceadores voltados para a Internet, internos e que suportam conexões criptografadas.

Balanceadores de Carga Voltados para a Internet

Um balanceador de carga voltado para a Internet é, como o nome indica, um balanceador de carga que recebe solicitações de clientes pela Internet e as distribui para instâncias do Amazon EC2 que são registrado com o balanceador de carga.

Quando você configura um balanceador de carga, ele recebe um nome DNS público que os clientes podem usar para enviar solicitações ao seu aplicativo. Os servidores DNS resolvem o nome DNS para o endereço IP público do seu balanceador de carga, que pode ser visível para aplicativos clientes.

Uma prática recomendada da AWS é sempre fazer referência a um balanceador de carga pelo nome DNS, em vez do endereço IP do balanceador de carga, a fim de fornecer um ponto de entrada único e estável. Como o Elastic Load Balancing é expandido para atender à demanda de tráfego, não é recomendável vincular um aplicativo a um endereço IP que pode não fazer mais parte do pool de recursos de um balanceador de carga. Pontos importantes a serem lembrados:

- O Elastic Load Balancing no Amazon VPC suporta apenas endereços IPv4.
- O Balanceamento de carga elástico no EC2-Classic suporta endereços IPv4 e IPv6.

Balanceadores de Carga Internos

Em um aplicativo de várias camadas, geralmente é útil carregar o equilíbrio entre as camadas do aplicativo.

Por exemplo, um balanceador de carga voltado para a Internet pode receber e equilibrar o tráfego externo para a apresentação ou camada da Web cujas

instâncias do Amazon EC2 enviam suas solicitações para um balanceador de carga sentado na frente da camada de aplicativo. Você pode usar balanceadores de carga internos para rotear o tráfego para suas instâncias do Amazon EC2 em VPCs com sub-redes privadas.

Balanceadores de Carga HTTPS

Você pode criar um balanceador de carga que use o protocolo SSL / Transport Layer Security (TLS) para conexões criptografadas (também conhecidas como descarga SSL). Esse recurso permite a criptografia de tráfego entre o seu balanceador de carga e os clientes que iniciam sessões HTTPS e para conexões entre seu balanceador de carga e suas instâncias de back-end.

O Elastic Load Balancing fornece políticas de segurança que têm configurações predefinidas de negociação SSL a serem usadas para negociar conexões entre clientes e o balanceador de carga.

Para usar o SSL, você deve instalar um certificado SSL no balanceador de carga usado para encerrar a conexão e descriptografar solicitações de clientes antes de enviar solicitações para as instâncias de backend do Amazon EC2. Opcionalmente, você pode optar por ativar a autenticação em suas instâncias de back-end.

O Elastic Load Balancing não suporta SNI (Server Name Indication) no seu balanceador de carga. Isso significa que, se você deseja hospedar vários sites em uma frota de instâncias do Amazon EC2 por trás do Elastic Load Balancing com um único certificado SSL, será necessário adicionar um nome alternativo do assunto (SAN) para cada site ao certificado para evitar que os usuários do site vejam uma mensagem de aviso quando o site é acessado.

Listeners (Ouvintes)

Todo balanceador de carga deve ter um ou mais listeners configurados. Um listener é um processo que verifica solicitações de conexão - por exemplo, um CNAME configurado com o nome de registro A de balanceador de carga.

Todo listener é configurado com um protocolo e uma porta (cliente para balanceador de carga) para uma conexão front-end e um protocolo e uma porta para a conexão back-end (balanceador de carga para instância do Amazon EC2).

O Elastic Load Balancing suporta os seguintes protocolos:

- HTTP
- HTTPS
- TCP
- SSL

O Elastic Load Balancing suporta protocolos que operam em duas camadas diferentes de OSI (Open System Interconnection).

No modelo OSI, a Camada 4 é a camada de transporte que descreve a conexão TCP entre o cliente e sua instância de back-end por meio do balanceador de carga. A camada 4 é o nível mais baixo configurável para o seu balanceador de carga.

A camada 7 é a camada de aplicativo que descreve o uso de conexões HTTP e HTTPS dos clientes para o balanceador de carga e do balanceador de carga para sua instância de back-end. O protocolo SSL é usado principalmente para criptografar dados confidenciais em redes inseguras, como a Internet.

O protocolo SSL estabelece uma conexão segura entre um cliente e o servidor back-end e garante que todos os dados passados entre o cliente e o servidor sejam privados.

Configurando o Balanceamento de Carga Elástico

O Elastic Load Balancing permite configurar muitos aspectos do balanceador de carga, incluindo tempo limite de conexão inativa, balanceamento de carga entre zonas, drenagem de conexão, protocolo proxy, sessões permanentes e verificações de integridade. As definições de configuração podem ser modificadas usando o AWS Management Console ou uma interface de linha de comando (CLI). Algumas das opções são descritas a seguir.

Tempo limite de conexão inativa

Para cada solicitação que um cliente faz por meio de um balanceador de carga, o balanceador de carga mantém duas conexões. Uma conexão é com o cliente e a outra é com a instância de back-end.

Para cada conexão, o balanceador de carga gerencia um tempo limite inativo que é acionado quando nenhum dado é enviado pela conexão por um período especificado.

Após o período de tempo limite inativo, se nenhum dado foi enviado ou recebido, o balanceador de carga fecha a conexão. Por padrão, o Elastic Load Balancing define o tempo limite inativo para 60 segundos nas duas conexões.

Se uma solicitação HTTP não for concluída dentro do período de tempo limite inativo, o balanceador de carga fechará a conexão, mesmo se os dados ainda estiverem sendo transferidos.

Você pode alterar a configuração de tempo limite ocioso das conexões para garantir que operações demoradas, como upload de arquivos, tenham tempo para serem concluídas. Se você usa listeners HTTP e HTTPS, recomendamos que você ative a opção keep-alive para suas instâncias do Amazon EC2.

Você pode ativar o keep-alive nas configurações do servidor da web ou nas configurações do kernel para as instâncias do Amazon EC2. Manter ativo (keepalive), quando ativado, permite que o balanceador de carga reutilize conexões com sua instância de back-end, o que reduz a utilização da CPU.

Para garantir que o balanceador de carga seja responsável por fechar as conexões com sua instância de backend, verifique se o valor definido para o tempo de manutenção é maior que a configuração de tempo limite inativo no seu balanceador de carga

Balanceamento de Carga entre Zonas (CrossZone)

Para garantir que o tráfego de solicitação seja roteado uniformemente em todas as instâncias de back-end para o seu balanceador de carga, independentemente da zona de disponibilidade em que elas estão localizadas, você deve habilitar o balanceamento de carga entre zonas no seu balanceador de carga.

O balanceamento de carga entre zonas reduz a necessidade de manter números equivalentes de instâncias de back-end em cada zona de disponibilidade e melhora a capacidade do seu aplicativo de lidar com a perda de uma ou mais instâncias de back-end. No entanto, ainda é recomendável que você mantenha um número aproximadamente equivalente de instâncias em cada zona de disponibilidade para obter maior tolerância a falhas.

Para ambientes em que os clientes armazenam em cache pesquisas de DNS, as solicitações recebidas podem favorecer uma das zonas de disponibilidade. Usando o balanceamento de carga entre zonas, esse desequilíbrio na carga de solicitação é espalhado por todas as instâncias de back-end disponíveis na região, reduzindo o impacto de clientes mal configurados.

Drenagem de Conexão (Connection Draining)

Você deve ativar a drenagem de conexão para garantir que o balanceador de carga pare de enviar solicitações para instâncias que estão com cancelamento de registro ou não estão íntegras, mantendo as conexões existentes abertas. Isso permite que o balanceador de carga conclua as solicitações de bordo feitas para essas instâncias.

Ao ativar a drenagem de conexão, é possível especificar um tempo máximo para o balanceador de carga manter as conexões ativas antes de relatar a instância como cancelada o registro.

O valor máximo do tempo limite pode ser definido entre 1 e 3.600 segundos (o padrão é 300 segundos). Quando o limite de tempo máximo é atingido, o balanceador de carga fecha forçosamente as conexões com a instância de cancelamento de registro.

Protocolo Proxy

Quando você usa TCP ou SSL para conexões de front-end e back-end, seu balanceador de carga encaminha solicitações para as instâncias de back-end sem modificar os cabeçalhos de solicitação.

Se você ativar o Protocolo Proxy, um cabeçalho legível por humanos será adicionado ao cabeçalho da solicitação com informações de conexão, como endereço IP de origem, endereço IP de destino e números de porta. O cabeçalho é então enviado para a instância de back-end como parte da solicitação.

Antes de usar o Proxy Protocol, verifique se o seu balanceador de carga não está atrás de um servidor proxy com o Proxy Protocol ativado. Se o Protocolo Proxy estiver ativado no servidor proxy e no balanceador, o balanceador de carga adiciona outro cabeçalho à solicitação, que já possui um cabeçalho do servidor proxy.

Dependendo de como sua instância de back-end estiver configurada, essa duplicação pode resultar em erros.

Sessões de aderência (Sticky Sessions)

Por padrão, um balanceador de carga roteia cada solicitação independentemente para a instância registrada com a menor carga.

No entanto, você pode usar o recurso de sessão permanente (também conhecido como afinidade da sessão), que permite ao balanceador de carga vincular a sessão de um usuário a uma instância específica. Isso garante que todas as solicitações do usuário durante a sessão sejam enviadas para a mesma instância.

A chave para gerenciar sessões persistentes é determinar por quanto tempo seu balanceador de carga deve encaminhar consistentemente a solicitação do usuário para a mesma instância.

Se seu aplicativo tiver seu próprio cookie de sessão, você poderá configurar o Elastic Load Balancing para que o cookie de sessão siga a duração especificada pelo cookie de sessão do aplicativo. Se seu aplicativo não tiver seu próprio cookie de

sessão, você poderá configurar o Elastic Load Balancing para criar um cookie de sessão especificando sua própria duração de aderência.

O Elastic Load Balancing cria um cookie chamado AWSELB que é usado para mapear a sessão para a instância.

Verificações de Saúde

O Elastic Load Balancing suporta verificações de integridade para testar o status das instâncias do Amazon EC2 atrás de um balanceador de carga do Elastic Load Balancing.

O status das instâncias que estão em boas condições no momento da verificação de integridade é InService.

O status de todas as instâncias que não são íntegras no momento da verificação de integridade é OutOfService.

O balanceador de carga executa verificações de integridade em todas as instâncias registradas para determinar se a instância está em um estado íntegro ou não íntegro.

Uma verificação de integridade é um ping, uma tentativa de conexão ou uma página que é verificada periodicamente. Você pode definir o intervalo de tempo entre as verificações de integridade e também o tempo de espera para responder, caso a página de verificação de integridade inclua um aspecto computacional.

Por fim, você pode definir um limite para o número de falhas consecutivas na verificação de integridade antes que uma instância seja marcada como não íntegra.

Atualizações por trás de um Balanceador de Carga Elástico

Aplicativos de execução demorada precisarão ser mantidos e atualizados com uma versão mais recente do aplicativo.

Ao usar instâncias do Amazon EC2 executando atrás de um balanceador de carga do Elastic Load Balancing, você pode cancelar o registro manual dessas instâncias

de longa duração do Amazon EC2 associadas a um balanceador de carga e registrar manualmente as instâncias do Amazon EC2 recém-iniciadas que você iniciou com as novas atualizações instaladas.

Amazon CloudWatch

O Amazon CloudWatch é um serviço que você pode usar para monitorar seus recursos e aplicativos da AWS em tempo real. Com o Amazon CloudWatch, você pode coletar e rastrear métricas, criar alarmes que enviam notificações e fazem alterações nos recursos monitorados com base nas regras definidas por você.

Por exemplo, você pode optar por monitorar a utilização da CPU para decidir quando adicionar ou remover instâncias do Amazon EC2 em uma camada de aplicativo.

Ou, se uma métrica específica de aplicativo específica que não estiver visível para a AWS for o melhor indicador para avaliar suas necessidades de dimensionamento, você poderá executar uma solicitação PUT para inserir essa métrica no Amazon CloudWatch.

Você pode usar essa métrica personalizada para gerenciar a capacidade. Você pode especificar parâmetros para uma métrica durante um período de tempo e configurar alarmes e ações automatizadas quando um limite for atingido.

O Amazon CloudWatch oferece suporte a vários tipos de ações, como enviar uma notificação para um tópico do Amazon SNS (Serviço de Notificação Simples da Amazon) ou executar uma política de Auto Scaling.

O Amazon CloudWatch oferece monitoramento básico ou detalhado dos produtos suportados da AWS. O monitoramento básico envia pontos de dados ao Amazon CloudWatch a cada cinco minutos, para um número limitado de métricas pré-selecionadas, sem nenhum custo.

O monitoramento detalhado envia pontos de dados para o Amazon CloudWatch a cada minuto e permite a agregação de dados por uma taxa adicional. Se você deseja usar o monitoramento detalhado, deve habilitá-lo - básico é o padrão.

O Amazon CloudWatch suporta monitoramento e métricas específicas para a maioria dos serviços da AWS, incluindo:

- Auto Scaling
- Amazon CloudFront
- Amazon CloudSearch
- Amazon DynamoDB
- Amazon EC2
- Amazon EC2 Container Service (Amazon ECS)
- Amazon ElastiCache
- Amazon ElastiCache
- Amazon Elastic Block Store (Amazon EBS)
- Elastic Load Balancing
- Amazon Elastic MapReduce (Amazon EMR)
- Amazon Elasticsearch Service
- Amazon Kinesis Streams
- Amazon Kinesis Firehose
- AWS Lambda
- Amazon Machine Learning
- AWS OpsWorks
- Amazon Redshift
- Amazon Relational Database Service (Amazon RDS)
- Amazon Route 53
- Amazon SNS
- Serviço de fila simples da Amazon (Amazon SQS)
- Amazon S3
- Serviço de fluxo de trabalho simples da AWS (Amazon SWF)

- AWS Storage Gateway
- AWS WAF
- Amazon WorkSpaces.

As métricas do Amazon CloudWatch podem ser recuperadas executando uma solicitação GET. Ao usar o monitoramento detalhado, você também pode agregar métricas por um período especificado.

O Amazon CloudWatch não agrega dados entre regiões, mas pode agregar zonas de disponibilidade em uma região.

A AWS fornece um rico conjunto de métricas incluídas em cada serviço, mas você também pode definir métricas personalizadas para monitorar recursos e eventos que a AWS não tem visibilidade - por exemplo, consumo de memória da instância do Amazon EC2 e métricas de disco visíveis para o sistema operacional de a instância do Amazon EC2, mas não visível para a AWS ou limites específicos de aplicativos em execução em instâncias que não são conhecidas pela AWS.

O Amazon CloudWatch oferece suporte a uma API (Interface de programação de aplicativos) que permite que programas e scripts coloquem métricas no Amazon CloudWatch como pares de nome e valor que podem ser usados para criar eventos e acionar alarmes da mesma maneira que as métricas padrão do Amazon CloudWatch.

Os logs do Amazon CloudWatch podem ser usados para monitorar, armazenar e acessar arquivos de log de instâncias do Amazon EC2, AWS CloudTrail e outras fontes.

Você pode recuperar os dados do log e monitorar em tempo real os eventos - por exemplo, pode rastrear o número de erros nos logs do aplicativo e enviar uma notificação se uma taxa de erro exceder um limite. O Amazon CloudWatch Logs também pode ser usado para armazenar seus logs no Amazon S3 ou Amazon Glacier.

Os logs podem ser retidos indefinidamente ou de acordo com uma política antiga que excluirá os logs mais antigos quando não for mais necessário.

Está disponível um agente do CloudWatch Logs que fornece uma maneira automatizada de enviar dados de log para o CloudWatch Logs para instâncias do Amazon EC2 executando o Amazon Linux ou Ubuntu. Você pode usar o instalador do agente Amazon CloudWatch Logs em uma instância existente do Amazon EC2 para instalar e configurar o agente CloudWatch Logs. Após a conclusão da instalação, o agente confirma que foi iniciado e permanece em execução até você desativá-lo.

O Amazon CloudWatch possui alguns limites que você deve ter em mente ao usar o serviço. Cada conta da AWS é limitada a 5.000 alarmes por conta da AWS, e os dados das métricas são retidos por duas semanas por padrão (no momento da redação deste). Se você deseja manter os dados por mais tempo, precisará mover os logs para um armazenamento persistente como o Amazon S3 ou o Amazon Glacier.

Você deve se familiarizar com os limites do Amazon CloudWatch no Guia do desenvolvedor do Amazon CloudWatch.

Escala automática (Auto Scaling)

Uma vantagem distinta da implantação de aplicativos na nuvem é a capacidade de iniciar e liberar servidores em resposta a cargas de trabalho variáveis.

Provisionar servidores sob demanda e liberá-los quando não forem mais necessários pode proporcionar uma economia significativa de custos para cargas de trabalho que não são estáveis.

Os exemplos incluem um site para um evento esportivo específico, um sistema de entrada de dados no final do mês, um site de compras de varejo com suporte a vendas instantâneas, um site de artistas da música durante o lançamento de novas

músicas, um site da empresa que anuncia ganhos bem-sucedidos ou um evento noturno, ou processamento executado para calcular a atividade diária.

O Auto Scaling é um serviço que permite dimensionar sua capacidade do Amazon EC2 automaticamente, dimensionando e dimensionando de acordo com os critérios definidos por você.

Com o Auto Scaling, você pode garantir que o número de instâncias do Amazon EC2 em execução aumente durante picos de demanda ou períodos de pico de demanda para manter o desempenho do aplicativo e diminua automaticamente durante períodos de pausa ou calha para minimizar os custos.

Sem Medo de Picos

Muitos aplicativos da Web têm aumentos de carga não planejados com base em eventos fora do seu controle.

Por exemplo, sua empresa pode ser mencionada em um blog ou programa de televisão popular, levando muito mais pessoas a visitar seu site do que o esperado.

Configurando Auto Scaling com antecedência permitirá que você adote e sobreviva a esse tipo de aumento rápido no número de solicitações. O Auto Scaling aumentará o tamanho do seu site para atender à crescente demanda e, em seguida, reduza a escala quando o evento desaparecer.

Planos de Dimensionamento Automático

O Auto Scaling possui vários esquemas ou planos que você pode usar para controlar como deseja que o Auto Scaling execute.

Manter os níveis da instância atual

Você pode configurar seu grupo de Auto Scaling para manter um número mínimo ou especificado de instâncias em execução o tempo todo.

Para manter os níveis de instância atuais, o Auto Scaling executa uma verificação periódica da integridade das instâncias em execução em um grupo de Auto Scaling. Quando o Auto Scaling encontra uma instância não íntegra, ele termina essa instância e inicia uma nova.

As cargas de trabalho em estado estacionário que precisam sempre de um número consistente de instâncias do Amazon EC2 podem usar o Auto Scaling para monitorar e manter esse número específico de instâncias do Amazon EC2 em execução.

Escala manual

A escala manual é a maneira mais básica de dimensionar seus recursos. Você só precisa especificar a alteração na capacidade máxima, mínima ou desejada do seu grupo de Auto Scaling.

O Auto Scaling gerencia o processo de criação ou encerramento de instâncias para manter a capacidade atualizada.

A expansão manual pode ser muito útil para aumentar os recursos de um evento pouco frequente, como o lançamento de uma nova versão do jogo que estará disponível para download e requer um registro do usuário. Para eventos de escala extremamente grande, mesmo os balanceadores de carga do Elastic Load Balancing podem ser pré-aquecidos trabalhando com suas soluções locais ou suporte da AWS.

Escalonamento Agendado

Às vezes, você sabe exatamente quando precisará aumentar ou diminuir o número de instâncias no seu grupo, simplesmente porque essa necessidade surge em um cronograma previsível.

Os exemplos incluem eventos periódicos, como processamento de final de mês, final de trimestre ou final de ano, e outros eventos recorrentes previsíveis.

Escalonamento agendado significa que as ações de escalonamento são executadas automaticamente em função da hora e da data.

Eventos recorrentes, como processamento de final de mês, trimestre ou ano, ou testes automáticos agendados e recorrentes de desempenho e carga, podem ser antecipados e o Auto Scaling pode ser aumentado adequadamente no momento do evento agendado.

Escala Dinâmica

O dimensionamento dinâmico permite definir parâmetros que controlam o processo de dimensionamento automático em uma política de dimensionamento.

Por exemplo, você pode criar uma política que adicione mais instâncias do Amazon EC2 à camada da Web quando a largura de banda da rede, medida pelo Amazon CloudWatch, atingir um determinado limite.

Componentes de Dimensionamento Automático

O Auto Scaling possui vários componentes que precisam ser configurados para funcionar corretamente: uma configuração de inicialização, um grupo de Auto Scaling e uma política de dimensionamento opcional.

Configuração de Inicialização

Uma configuração de inicialização é o modelo usado pelo Auto Scaling para criar novas instâncias e é composto pelo nome da configuração, AMI (Amazon Machine Image), tipo de instância do Amazon EC2, grupo de segurança e par de chaves da instância. Cada grupo do Auto Scaling pode ter apenas uma configuração de inicialização por vez.

Grupos de segurança para instâncias iniciadas no EC2-Classic podem ser referenciados pelo nome do grupo de segurança, como "SSH" ou "Web", se é assim que eles são nomeados, ou você pode fazer referência aos IDs do grupo de segurança, como sg-f57cde9d.

Se você iniciou as instâncias no Amazon VPC, recomendado, você deve usar os IDs do grupo de segurança para fazer referência aos grupos de segurança que deseja associar às instâncias em uma configuração de inicialização do Auto Scaling.

O limite padrão para configurações de inicialização é de 100 por região. Se você exceder esse limite, a chamada para create-launch-configuration falhará. Você pode visualizar e atualizar esse limite executando a descrição-conta-limites na linha de comando, conforme mostrado aqui.

```
> aws autoscaling describe-account-limits
```

O Auto Scaling pode fazer com que você alcance limites de outros serviços, como o número padrão de instâncias do Amazon EC2 que você pode iniciar atualmente em uma região, que é 20. Ao criar arquiteturas mais complexas com a AWS, é importante ter em mente os limites de serviço para todos os serviços da AWS Cloud que você está usando.

Quando você executa um comando usando a CLI e ele falha, verifique sua sintaxe primeiro. Se isso ocorrer, verifique os limites do comando que você está tentando e verifique se você não excedeu um limite.

Alguns limites podem ser aumentados e geralmente têm um valor razoável para limitar uma condição de corrida, um script incorreto em execução em um loop ou outra automação semelhante que pode causar alto uso não intencional e cobrança dos recursos da AWS.

Os limites de serviço da AWS podem ser visualizados no Guia de referência geral da AWS, em Limites de serviço da AWS.

Grupo de Dimensionamento Automático

Um grupo de Auto Scaling é uma coleção de instâncias do Amazon EC2 gerenciadas pelo serviço Auto Scaling. Cada grupo do Auto Scaling contém opções de

configuração que controlam quando o Auto Scaling deve iniciar novas instâncias e encerrar instâncias existentes.

Um grupo de dimensionamento automático deve conter um nome e um número mínimo e máximo de instâncias que possam estar no grupo. Opcionalmente, você pode especificar a capacidade desejada, que é o número de instâncias que o grupo deve ter o tempo todo.

Se você não especificar uma capacidade desejada, a capacidade desejada padrão será o número mínimo de instâncias que você especificar.

Um grupo de Auto Scaling pode usar Instâncias sob demanda ou Spot como as instâncias do Amazon EC2 que gerencia.

On-Demand é o padrão, mas as Instâncias Spot podem ser usadas referenciando um preço máximo de oferta na configuração de inicialização (—spot-price "0,15") associada ao grupo Auto Scaling.

Você pode alterar o preço da oferta criando uma nova configuração de lançamento com o novo preço da oferta e associando-a ao seu grupo de Auto Scaling.

Se houver instâncias disponíveis no preço do lance ou abaixo dele, elas serão lançadas no seu grupo Auto Scaling.

As Instâncias Spot em um grupo de Auto Scaling seguem as mesmas diretrizes das Instâncias Spot fora de um grupo de Auto Scaling e exigem aplicativos flexíveis e que possam tolerar instâncias do Amazon EC2 encerradas com aviso prévio, por exemplo, quando o preço Spot subir acima do preço da oferta você define na configuração de inicialização.

Uma configuração de ativação pode fazer referência a Instâncias On Demand ou Instâncias Spot, mas não as duas.

Vamos de Spot

O Auto Scaling suporta o uso de instâncias spot. Isso pode ser muito útil quando você está hospedando sites nos quais deseja fornecer capacidade de computação adicional, mas com restrições de preço.

Um exemplo é um modelo de site "freemium", no qual você pode oferecer algumas funcionalidade básica aos usuários gratuitamente e funcionalidade adicional para usuários premium que pagam pelo uso. Instâncias spot podem ser usadas para fornecer a funcionalidade básica quando disponível referenciando um preço máximo de oferta na configuração de lançamento ("preço do ponto" 0,15 ") associado ao grupo Auto Scaling.

Política de Dimensionamento

Você pode associar alarmes e políticas de dimensionamento do Amazon CloudWatch a um grupo de dimensionamento automático para ajustar dinamicamente o dimensionamento automático.

Quando um limite é ultrapassado, o Amazon CloudWatch envia alarmes para acionar alterações (aumento ou redução da escala) no número de instâncias do Amazon EC2 atualmente recebendo tráfego atrás de um balanceador de carga.

Depois que o alarme do Amazon CloudWatch envia uma mensagem ao grupo Auto Scaling, o Auto Scaling executa a política associada para dimensionar seu grupo. A política é um conjunto de instruções que informam ao Auto Scaling se deve ser expandido, iniciando novas instâncias do Amazon EC2 referenciadas na configuração de inicialização associada ou se é dimensionado e finalizado.

Existem várias maneiras de configurar uma política de dimensionamento: Você pode aumentar ou diminuir em um número específico de instâncias, como adicionar duas instâncias; você pode segmentar um número específico de instâncias, como no máximo cinco instâncias totais do Amazon EC2, ou você pode ajustar com base em uma porcentagem.

Você também pode escalar por etapas e aumentar ou diminuir a capacidade atual do grupo com base em um conjunto de ajustes de escala que variam com base no tamanho do acionador do limite de alarme.

Você pode associar mais de uma política de dimensionamento a um grupo de dimensionamento automático. Por exemplo, você pode criar uma política usando o gatilho para utilização da CPU, chamado CPULoad, e a métrica CloudWatch CPUUtilization para especificar a expansão se a utilização da CPU for maior que 75% para dois minutos.

Você pode anexar outra política ao mesmo grupo de Auto Scaling para escalar se a utilização da CPU for inferior a 40% por 20 minutos.

Uma prática recomendada é escalar rapidamente e escalar lentamente para que você possa responder a picos ou explosões, mas evite encerrar inadvertidamente instâncias do Amazon EC2 com muita rapidez, apenas tendo que iniciar mais instâncias do Amazon EC2 se a explosão for sustentada.

O Auto Scaling também suporta um período de espera, que é uma configuração configurável que determina quando suspender as atividades de dimensionamento por um curto período de tempo para um grupo de Auto Scaling.

Se você iniciar uma instância do Amazon EC2, será cobrado por uma hora inteira de tempo de execução. As horas parciais da instância consumidas são cobradas como horas completas.

Isso significa que, se você tiver uma política de dimensionamento permissiva que inicie, encerre e reinicie várias instâncias por hora, estará cobrando uma hora inteira para cada instância iniciada, mesmo que encerre algumas dessas instâncias em menos de uma hora.

Uma prática recomendada para rentabilidade é aumentar rapidamente quando necessário, mas diminuir mais lentamente para evitar a necessidade de reiniciar novas e separadas instâncias do Amazon EC2 para um aumento na demanda de carga de trabalho que flutua para cima e para baixo em minutos, mas geralmente continua a precisar de mais recursos dentro de uma hora.

É importante considerar a inicialização de instâncias do Amazon EC2 iniciadas usando o Auto Scaling. Leva tempo para configurar cada instância recém-lançada do Amazon EC2 antes que a instância esteja íntegra e capaz de aceitar tráfego.

Instâncias iniciadas e disponíveis para carregamento mais rápido podem ingressar no pool de capacidade mais rapidamente.

AWS Identity and Access Management (IAM)

O IAM é um serviço poderoso que permite controlar como as pessoas e os programas têm permissão para manipular sua infraestrutura da AWS. O IAM usa conceitos tradicionais de identidade, como usuários, grupos e políticas de controle de acesso para controlar quem pode usar sua conta da AWS, quais serviços e recursos eles podem usar e como eles podem usá-los.

O controle fornecido pelo IAM é granular o suficiente para limitar um único usuário à capacidade de executar uma única ação em um recurso específico a partir de um endereço IP específico durante uma janela de tempo específica.

Os aplicativos podem ter acesso aos recursos da AWS, estejam eles executando localmente ou na nuvem. Essa flexibilidade cria um sistema muito poderoso que lhe dará todo o poder necessário para garantir que os usuários da sua conta da AWS tenham a capacidade de atender às suas necessidades de negócios e, ao mesmo tempo, abordar todas as preocupações de segurança da sua organização.

Lembre-se de que, no modelo de responsabilidade compartilhada, você controla o console e a configuração do sistema operacional.

Qualquer mecanismo que você usa atualmente para controlar o acesso à infraestrutura do servidor continuará funcionando nas instâncias do Amazon Elastic Compute Cloud (Amazon EC2), seja gerenciando contas de login de máquinas individuais ou um serviço de diretório como o Active Directory ou o LDAP. Você pode executar um servidor Active Directory ou LDAP no Amazon EC2 ou pode estender seu sistema onpremises para a nuvem.

O Serviço de Diretório da AWS também funcionará bem para fornecer a funcionalidade do Active Directory na nuvem como um serviço, independente ou integrado ao seu Active Directory existente.

Principais do IAM

O primeiro conceito do IAM a entender é o principal. Um principal é uma entidade do IAM que tem permissão para interagir com os recursos da AWS. Um principal pode ser permanente ou temporário e pode representar um ser humano ou um

aplicativo. Existem três tipos de entidades: usuários raiz, usuários IAM e funções / tokens de segurança temporários.

Usuário raiz

Quando você cria uma conta da AWS pela primeira vez, começa com apenas uma entidade de login único que tem acesso completo a todos os serviços e recursos da nuvem da AWS na conta. Esse princípio é chamado de usuário root. Contanto que você tenha uma conta aberta na AWS, o usuário raiz desse relacionamento persistirá. O usuário root pode ser usado para o acesso do console e do programa aos recursos da AWS.

O usuário root é semelhante em conceito à raiz UNIX ou à conta de administrador do Windows - possui privilégios totais para fazer qualquer coisa na conta, incluindo o fechamento da conta.

É altamente recomendável que você não use o usuário root nas tarefas diárias, mesmo as administrativas. Em vez disso, siga as práticas recomendadas de usar o usuário raiz apenas para criar seu primeiro usuário do IAM e, em seguida, bloquear com segurança as credenciais do usuário raiz.

Usuários do IAM

Os usuários são identidades persistentes configuradas pelo serviço IAM para representar pessoas ou aplicativos individuais. Você pode criar usuários separados do IAM para cada membro da sua equipe de operações, para que eles possam interagir com o console e usar a CLI.

Você também pode criar usuários de desenvolvimento, teste e produção para aplicativos que precisam acessar os serviços em nuvem da AWS (embora você veja mais adiante neste capítulo que as funções do IAM podem ser uma solução melhor para esse caso de uso).

Os usuários do IAM podem ser criados por entidades com privilégios administrativos do IAM a qualquer momento, por meio do Console de

Gerenciamento da AWS, CLI ou SDKs. Os usuários são persistentes, pois não há período de expiração; elas são entidades permanentes que existem até que um administrador do IAM faça uma ação para excluí-los.

Os usuários são uma excelente maneira de aplicar o princípio do menor privilégio, ou seja, o conceito de permitir que uma pessoa ou processo que interaja com os recursos da AWS execute exatamente as tarefas de que precisa, mas nada mais.

Os usuários podem ser associados a políticas muito granulares que definem essas permissões.

Funções / Tokens de Segurança Temporários

Funções e tokens de segurança temporários são muito importantes para o uso avançado do IAM, mas muitos usuários da AWS os consideram confusos.

As funções são usadas para conceder privilégios específicos a atores específicos por um período de tempo definido.

Esses atores podem ser autenticados pela AWS ou por algum sistema externo confiável. Quando um desses atores assume uma função, a AWS fornece ao agente um token de segurança temporário do AWS Security Token Service (STS) que o ator pode usar para acessar os serviços da AWS Cloud.

A solicitação de um token de segurança temporário requer a especificação de quanto tempo o token existirá antes de expirar. O intervalo de uma vida útil temporária do token de segurança é de 15 minutos a 36 horas.

As funções e os tokens de segurança temporários permitem vários casos de uso:

- Funções do Amazon EC2 - concessão de permissões para aplicativos em execução em uma instância do Amazon EC2.
- Acesso entre contas - conceder permissões a usuários de outras contas da AWS, independentemente de você controlar essas contas ou não.

- Federação - Concedendo permissões a usuários autenticados por um sistema externo confiável.

Funções do Amazon EC2

A concessão de permissões para um aplicativo é sempre complicada, pois geralmente requer a configuração do aplicativo com algum tipo de credencial na instalação.

Isso leva a problemas relacionados ao armazenamento seguro da credencial antes do uso, como acessá-la com segurança durante a instalação e como protegê-la na configuração.

Suponha que um aplicativo em execução em uma instância do Amazon EC2 precise acessar um bucket do Amazon Simple Storage Service (Amazon S3). Uma política que concede permissão para ler e gravar esse bucket pode ser criada e atribuída a um usuário do IAM, e o aplicativo pode usar a chave de acesso desse usuário do IAM para acessar o bucket do Amazon S3.

O problema dessa abordagem é que a chave de acesso do usuário deve estar acessível ao aplicativo, provavelmente armazenando-a em algum tipo de arquivo de configuração. O processo para obter a chave de acesso e armazená-la criptografada na configuração geralmente é complicado e dificulta o desenvolvimento ágil.

Além disso, a chave de acesso corre risco ao ser distribuída. Finalmente, quando chega a hora de girar a tecla de acesso, a rotação envolve executar todo o processo novamente.

O uso de funções do IAM no Amazon EC2 elimina a necessidade de armazenar credenciais da AWS em um arquivo de configuração.

Uma alternativa é criar uma função do IAM que conceda o acesso necessário ao bucket do Amazon S3. Quando a instância do Amazon EC2 é iniciada, a função é atribuída à instância.

Quando o aplicativo em execução na instância usa a API (Interface de Programação de Aplicativos) para acessar o bucket do Amazon S3, ele assume a função atribuída à instância e obtém um token temporário que envia à API.

O processo de obter o token temporário e passá-lo para a API é tratado automaticamente pela maioria dos SDKs da AWS, permitindo o aplicativo para fazer uma chamada para acessar o bucket do Amazon S3 sem se preocupar com autenticação.

Além de ser fácil para o desenvolvedor, isso elimina a necessidade de armazenar uma chave de acesso em um arquivo de configuração. Além disso, como o acesso à API usa um token temporário, não há chave de acesso fixo que precise ser rotacionada.

Acesso entre contas (Cross Account)

Outro caso de uso comum para funções do IAM é conceder acesso aos recursos da AWS para usuários do IAM em outras contas da AWS.

Essas contas podem ser outras contas da AWS controladas por sua empresa ou agentes externos, como clientes ou fornecedores.

Você pode configurar uma função do IAM com as permissões que deseja conceder aos usuários da outra conta, para que os usuários da outra conta possam assumir essa função para acessar seus recursos.

Isso é altamente recomendado como uma prática recomendada, em vez de distribuir chaves de acesso fora da sua organização.

Federação

Muitas organizações já possuem um repositório de identidade fora da AWS e preferem aproveitar esse repositório a criar um repositório novo e amplamente duplicado de usuários do IAM.

Da mesma forma, os aplicativos baseados na Web podem querer aproveitar identidades baseadas na Web, como Facebook, Google ou Login com a Amazon.

Os provedores de identidade do IAM oferecem a capacidade de associar essas identidades externas ao IAM e atribuir privilégios aos usuários autenticados fora do IAM. O IAM pode integrar-se a dois tipos diferentes de Provedores de Identidade externos (IdP).

Para federar identidades da web como Facebook, Google ou Login com Amazon, o IAM oferece suporte à integração via OpenID Connect (OIDC).

Isso permite que o IAM conceda privilégios a usuários autenticados com alguns dos principais IdPs baseados na Web. Para federar identidades internas, como Active Directory ou LDAP, o IAM oferece suporte à integração via SAML (Security Assertion Markup Language 2.0).

Um IdP compatível com SAML, como os Serviços de Federação do Active Directory (ADFS), é usado para associar o diretório interno ao IAM. (Instruções para configurar muitos produtos compatíveis podem ser encontradas no site da AWS.)

Em cada caso, a federação funciona retornando um token temporário associado a uma função para o IdP para a identidade autenticada a ser usada nas chamadas para a API da AWS.

A função real retornada é determinada por meio de informações recebidas do IdP, atributos do usuário no armazenamento de identidade local ou o nome de usuário e o serviço de autenticação do armazenamento de identidade da web.

Autenticação

Existem três maneiras pelas quais o IAM autentica uma entidade:

- Nome de usuário / senha - Quando um principal representa um humano que interage com o console, o humano fornecerá um par de nome de usuário / senha para verificar sua identidade. O IAM permite criar uma política de senha que imponha complexidade e expiração de senha.
- Chave de acesso - Uma chave de acesso é uma combinação de um ID da chave de acesso (20 caracteres) e uma chave secreta de acesso (40

caracteres). Quando um programa está manipulando a infraestrutura da AWS por meio da API, ele usa esses valores para assinar as chamadas REST subjacentes aos serviços. Os SDKs e ferramentas da AWS lidam com todos os meandros da assinatura de chamadas REST. Portanto, usar uma chave de acesso quase sempre será uma questão de fornecer os valores ao SDK ou à ferramenta.

- Chave de acesso / token de sessão - Quando um processo opera sob uma função assumida, o token de segurança temporário fornece uma chave de acesso para autenticação. Além da chave de acesso (lembre-se de que ela consiste em duas partes), o token também inclui um token de sessão.

As chamadas para a AWS devem incluir a chave de acesso em duas partes e o token da sessão para autenticação.

É importante observar que, quando um usuário do IAM é criado, ele não possui uma chave de acesso nem uma senha, e o administrador do IAM pode configurar um ou ambos.

Isso adiciona uma camada extra de segurança, pois os usuários do console não podem usar suas credenciais para executar um programa que acessa sua infraestrutura da AWS.

Autorização

Depois que o IAM autentica um principal, ele deve gerenciar o acesso desse principal para proteger sua infraestrutura da AWS.

O processo de especificar exatamente quais ações um diretor pode ou não executar é chamado de autorização.

A autorização é tratada no IAM, definindo privilégios específicos nas políticas e associando essas políticas aos principais.

Políticas

O entendimento de como o gerenciamento de acesso funciona no IAM começa com o entendimento de políticas. Uma política é um documento JSON que define completamente um conjunto de permissões para acessar e manipular os recursos da AWS.

Os documentos de política contêm uma ou mais permissões, com cada permissão definindo:

- Efeito - uma única palavra: Permitir ou Negar.
- Serviço - Para que serviço essa permissão se aplica? A maioria dos serviços em nuvem da AWS oferece suporte à concessão de acesso por meio do IAM, incluindo o próprio IAM.
- Recurso - o valor do recurso especifica a infraestrutura específica da AWS à qual essa permissão se aplica.

Isso é especificado como um ARN (Amazon Resource Name). O formato para um ARN varia um pouco entre os serviços, mas o formato básico é: "arn: aws: service: region: account-id: [resourcetype:] resource"

Para alguns serviços, valores curinga são permitidos; por exemplo, um Amazon S3 ARN poderia ter um recurso de foldername \ * para indicar todos os objetos na pasta especificada

Associando Políticas a Principais (recursos demandantes)

Existem várias maneiras de associar uma política a um usuário do IAM; Esta seção cobrirá apenas os mais comuns. Uma política pode ser associada diretamente a um usuário do IAM de duas maneiras:

- Política do usuário - Essas políticas existem apenas no contexto do usuário ao qual estão anexadas. No console, uma política de usuário é inserida na interface do usuário na página de usuário do IAM.
- Políticas gerenciadas - Essas políticas são criadas na guia Políticas na página do IAM (ou através da CLI e assim por diante) e existem independentemente de qualquer usuário individual. Dessa maneira, a mesma política pode ser associada a muitos usuários ou grupos de usuários.

Há um grande número de políticas gerenciadas predefinidas que você pode revisar na guia Políticas da página IAM no AWS Management Console. Além disso, você pode escrever suas próprias políticas específicas para seus casos de uso.

O uso de políticas gerenciadas predefinidas garante que, quando novas permissões forem adicionadas para novos recursos, seus usuários ainda tenham o acesso correto.

O outro método comum para associar políticas a usuários é o recurso de grupos do IAM. Os grupos simplificam o gerenciamento de permissões para um grande número de usuários. Depois que uma política é atribuída a um grupo, qualquer usuário que seja membro desse grupo assume essas permissões.

Isso simplifica a atribuição de políticas a uma equipe inteira em sua organização.

Por exemplo, se você criar um grupo "Operações" com todos os usuários do IAM para sua equipe de operações atribuídos a esse grupo, é simples associar as permissões necessárias ao grupo, e todos os usuários do IAM da equipe assumirão essas permissões . Novos usuários do IAM podem ser atribuídos diretamente ao grupo.

Esse é um processo de gerenciamento muito mais simples do que ter que revisar quais políticas um novo usuário do IAM para a equipe de operações deve receber e adicionar manualmente essas políticas ao usuário.

Há duas maneiras de associar uma política a um grupo do IAM:

- Política de Grupo - Essas políticas existem apenas no contexto do grupo ao qual estão anexadas. No AWS Management Console, uma política de grupo é inserida na interface do usuário na página Grupo do IAM.
- Políticas gerenciadas - Da mesma maneira que as políticas gerenciadas (discutidas na seção "Autorização") podem ser associadas aos usuários do IAM, elas também podem ser associadas aos grupos do IAM.

Uma boa primeira etapa é usar o usuário root para criar um novo grupo do IAM chamado "Administradores do IAM" e atribuir a política gerenciada, "IAMFullAccess".

Em seguida, crie um novo usuário do IAM chamado “Administrador”, atribua uma senha e adicione-a ao grupo Administradores do IAM. Nesse ponto, você pode fazer logoff como usuário raiz e realizar toda a administração adicional com a conta de usuário do IAM.

A maneira final de um ator poder ser associado a uma política é assumindo um papel. Nesse caso, o ator pode ser:

- Um usuário IAM autenticado (pessoa ou processo). Nesse caso, o usuário do IAM deve ter os direitos para assumir a função.
- Uma pessoa ou processo autenticado por um serviço confiável fora da AWS, como um diretório LDAP exclusivo ou um serviço de autenticação da web. Nessa situação, um serviço da AWS Cloud assumirá a função em nome do ator e devolverá um token ao ator.

Depois que um ator assume uma função, ele recebe um token de segurança temporário associado às políticas dessa função. O token contém todas as informações necessárias para autenticar as chamadas de API.

Essas informações incluem uma chave de acesso padrão e um token de sessão adicional necessário para autenticar chamadas em uma função assumida.

Outros recursos principais

Além dos conceitos críticos de entidades, autenticação e autorização, existem vários outros recursos do serviço IAM que são importantes para entender para obter todos os benefícios do IAM.

Autenticação Multifator (MFA)

A autenticação multifator (MFA) pode adicionar uma camada extra de segurança à sua infraestrutura, adicionando um segundo método de autenticação além de apenas uma senha ou chave de acesso.

Com o MFA, a autenticação também exige a inserção de uma Senha de uso único (OTP) a partir de um dispositivo pequeno.

O dispositivo MFA pode ser um pequeno dispositivo de hardware que você carrega com você ou um dispositivo virtual por meio de um aplicativo no seu smartphone (por exemplo, o aplicativo AWS Virtual MFA). O MFA exige que você verifique sua identidade com algo que você conhece e algo que possui.

O MFA pode ser atribuído a qualquer conta de usuário do IAM, independentemente de representar uma pessoa ou aplicativo.

Quando uma pessoa que usa um usuário do IAM configurado com o MFA tenta acessar o Console de Gerenciamento da AWS, depois de fornecer sua senha, será solicitado a inserir o código atual exibido no dispositivo MFA antes de receber acesso.

Um aplicativo que usa um usuário do IAM configurado com o MFA deve consultar o usuário do aplicativo para fornecer o código atual, que o aplicativo passará para a API.

É altamente recomendável que os clientes da AWS adicionem proteção MFA ao usuário root.

Chaves Rotativas

O risco de segurança de qualquer credencial aumenta com a idade da credencial. Para esse fim, é uma prática recomendada de segurança girar as chaves de acesso associadas aos usuários do IAM.

O IAM facilita esse processo, permitindo duas chaves de acesso ativas por vez. O processo para girar as chaves pode ser conduzido por meio do console, CLI ou SDKs:

1. Crie uma nova chave de acesso para o usuário.

2. Reconfigure todos os aplicativos para usar a nova chave de acesso.

3. Desabilite a chave de acesso original (a desativação em vez de excluir neste estágio é crítica, pois permite a reversão da chave original se houver problemas com a rotação).

4. Verifique o funcionamento de todos os aplicativos.

5. Exclua a chave de acesso original.

As teclas de acesso devem ser rotacionadas regularmente.

Resolvendo Várias Permissões

Ocasionalmente, várias permissões serão aplicáveis ao determinar se um principal tem o privilégio de executar alguma ação.

Essas permissões podem vir de várias políticas associadas a uma política principal ou de recursos anexada ao recurso da AWS em questão. É importante saber como os conflitos entre essas permissões são resolvidos:

1. Inicialmente, a solicitação é negada por padrão.

2. Todas as políticas apropriadas são avaliadas; se houver uma "negação" explícita encontrada em qualquer política, a solicitação será negada e a avaliação será interrompida.

3. Se nenhum "negar" explícito for encontrado e um "permitir" explícito for encontrado em qualquer política, a solicitação será permitida.

4. Se não houver permissões explícitas de "permitir" ou "negar" encontradas, o padrão "negar" será mantido e a solicitação será negada.

A única exceção a essa regra é que, se uma chamada AssumeRole incluir uma função e uma política, a política não poderá expandir os privilégios da função (por exemplo, a política não poderá substituir nenhuma permissão negada por padrão na função).

Bancos de Dados na AWS

Quase todos os aplicativos contam com um banco de dados para armazenar dados e registros importantes para seus usuários.

Um mecanismo de banco de dados permite que seu aplicativo acesse, gerencie e pesquise grandes volumes de registros de dados. Em um aplicativo bem arquitetado, o banco de dados precisará atender às demandas de desempenho, às necessidades de disponibilidade e às características de recuperação do sistema.

Os sistemas e mecanismos de banco de dados podem ser agrupados em duas grandes categorias: bancos de dados Relational Database Management Systems (RDBMS) e NoSQL (ou não relacional).

Não é incomum criar um aplicativo usando uma combinação de bancos de dados RDBMS e NoSQL.

Bancos de Dados Relacionais

O tipo mais comum de banco de dados em uso atualmente é o banco de dados relacional. O banco de dados relacional tem raízes que remontam à década de 1970, quando Edgar F. Codd, trabalhando para a IBM, desenvolveu os conceitos do modelo relacional.

Hoje, os bancos de dados relacionais fornecem todos os tipos de aplicativos, de aplicativos de mídia social, sites de comércio eletrônico e blogs a aplicativos corporativos complexos.

Os pacotes de software de banco de dados relacional comumente usados incluem MySQL, PostgreSQL, Microsoft SQL Server e Oracle.

Os bancos de dados relacionais fornecem uma interface comum que permite aos usuários ler e gravar no banco de dados usando comandos ou consultas gravadas usando o SQL (Structured Query Language).

Um banco de dados relacional consiste em uma ou mais tabelas e uma tabela consiste em colunas e linhas semelhantes a uma planilha.

Uma coluna do banco de dados contém um atributo específico do registro, como

como nome, endereço e número de telefone de uma pessoa.

Cada atributo recebe um tipo de dados como texto, número ou data, e o mecanismo do banco de dados rejeitará entradas inválidas.

Um registro em uma tabela pode estar relacionado a um registro em outra tabela fazendo referência à chave primária de um registro.

Esse ponteiro ou referência é chamado de chave estrangeira.

Por exemplo, a tabela Notas que registra notas para cada aluno teria sua própria chave primária e uma coluna adicional conhecida como chave estrangeira que se refere à chave primária do registro do aluno.

Ao fazer referência às chaves primárias de outras tabelas, os bancos de dados relacionais minimizam a duplicação de dados nas tabelas associadas.

Com bancos de dados relacionais, é importante observar que a estrutura da tabela (como o número de colunas e o tipo de dados de cada coluna) deve ser definida antes que os dados sejam adicionados à tabela.

Um banco de dados relacional pode ser categorizado como um sistema de banco de dados OLTP (Online Transaction Processing) ou OLAP (Online Analytical Processing), dependendo de como as tabelas são organizadas e de como o aplicativo usa o banco de dados relacional.

OLTP refere-se a aplicativos orientados a transações que frequentemente escrevem e alteram dados (por exemplo, entrada de dados e comércio eletrônico).

OLAP é normalmente o domínio de data warehouses e refere-se a relatórios ou análises de grandes conjuntos de dados.

Aplicativos grandes geralmente têm uma combinação de OLTP e OLAP.

O Amazon Relational Database Service (Amazon RDS) simplifica significativamente a configuração e a manutenção dos bancos de dados OLTP e OLAP.

O Amazon RDS fornece suporte para seis mecanismos populares de banco de dados relacional: MySQL, Oracle, PostgreSQL, Microsoft SQL Server, MariaDB, e Amazon Aurora. Você também pode optar por executar praticamente qualquer mecanismo de banco de dados usando instâncias do Amazon Elastic Compute Cloud (Amazon EC2) para Windows ou Linux e gerenciar você mesmo a instalação e administração.

Data Warehouses

Um data warehouse é um repositório central de dados que podem vir de uma ou mais fontes. Esse repositório de dados geralmente é um tipo especializado de banco de dados relacional que pode ser usado para geração de relatórios e análises via OLAP.

As organizações geralmente usam data warehouses para compilar relatórios e pesquisar no banco de dados usando consultas altamente complexas.

Os data warehouses também são normalmente atualizados em uma programação em lote várias vezes por dia ou por hora, em comparação com um banco de dados relacional OLTP que pode ser atualizado milhares de vezes por segundo.

Muitas organizações dividem seus bancos de dados relacionais em dois bancos de dados diferentes: um banco de dados como principal banco de dados de produção para transações OLTP e outro banco de dados como seu data warehouse para OLAP.

As transações OLTP ocorrem com frequência e são relativamente simples.

As transações OLAP ocorrem com muito menos frequência, mas são muito mais complexas.

O Amazon RDS é frequentemente usado para cargas de trabalho OLTP, mas também pode ser usado para OLAP.

O Amazon Redshift é um data warehouse de alto desempenho projetado especificamente para casos de uso OLAP.

Também é comum combinar o Amazon RDS com o Amazon Redshift no mesmo aplicativo e extrair periodicamente transações recentes e carregá-las em um banco de dados de relatórios.

Bancos de dados NoSQL

Os bancos de dados NoSQL ganharam popularidade significativa nos últimos anos porque geralmente são mais simples de usar, mais flexíveis e podem atingir níveis de desempenho difíceis ou impossíveis com os bancos de dados relacionais tradicionais.

Os bancos de dados relacionais tradicionais são difíceis de escalar além de um único servidor sem custos e engenharia significativos, mas uma arquitetura NoSQL permite escalabilidade horizontal em hardware comum.

Os bancos de dados NoSQL não são relacionais e não possuem a mesma semântica de tabela e coluna de um banco de dados relacional.

Os bancos de dados NoSQL geralmente são repositórios de chave / valor ou repositórios de documentos com esquemas flexíveis que podem evoluir ao longo do tempo ou variar. Compare isso com um banco de dados relacional, que requer um esquema muito rígido.

Muitos dos conceitos das arquiteturas NoSQL remontam a seus conceitos fundamentais de volta aos whitepapers publicados em 2006 e 2007 que descreviam sistemas distribuídos como o Dynamo na Amazon.

Hoje, muitas equipes de aplicativos usam Hbase, MongoDB, Cassandra, CouchDB, Riak e Amazon DynamoDB para armazenar grandes volumes de dados com altas taxas de transação.

Muitos desses mecanismos de banco de dados oferecem suporte ao cluster e escalam horizontalmente em muitas máquinas para desempenho e tolerância a falhas.

Um caso de uso comum para o NoSQL é gerenciar o estado da sessão do usuário, perfis de usuário, dados do carrinho de compras ou dados de séries temporais.

Você pode executar qualquer tipo de banco de dados NoSQL na AWS usando o Amazon EC2 ou pode escolher um serviço gerenciado como o Amazon DynamoDB para lidar com o trabalho pesado envolvido na criação de um cluster distribuído que abrange vários data centers.

Serviço de Banco de Dados Relacional da Amazon (Amazon RDS)

O Amazon RDS é um serviço que simplifica a configuração, operações e dimensionamento de um banco de dados relacional na AWS.

Com o Amazon RDS, você pode dedicar mais tempo ao aplicativo e ao esquema e permitir que o Amazon RDS descarregue tarefas comuns, como backups, aplicação de patches, dimensionamento e replicação.

O Amazon RDS ajuda a otimizar a instalação do software de banco de dados e também o provisionamento da capacidade da infraestrutura.

Em alguns minutos, o Amazon RDS pode iniciar um dos muitos mecanismos de banco de dados populares prontos para começar a realizar transações SQL.

Após o lançamento inicial, o Amazon RDS simplifica a manutenção contínua, automatizando tarefas administrativas comuns de forma recorrente.

Com o Amazon RDS, você pode acelerar os cronogramas de desenvolvimento e estabelecer um modelo operacional consistente para gerenciar bancos de dados relacionais.

Por exemplo, o Amazon RDS facilita a replicação de seus dados para aumentar a disponibilidade, melhorar a durabilidade ou ampliar ou além de uma instância de

banco de dados única para cargas de trabalho de banco de dados com leitura pesada.

O Amazon RDS expõe um terminal de banco de dados ao qual o software cliente pode conectar e executar o SQL.

O Amazon RDS não fornece acesso de shell às instâncias de banco de dados (DB) e restringe o acesso a determinados procedimentos e tabelas do sistema que exigem privilégios avançados.

Com o Amazon RDS, você normalmente pode usar as mesmas ferramentas para consultar, analisar, modificar e administrar o banco de dados.

Por exemplo, as ferramentas atuais Extract, Transform, Load (ETL) e as ferramentas de relatório podem se conectar aos bancos de dados do Amazon RDS da mesma maneira com os mesmos drivers e, geralmente, tudo o que é necessário para reconfigurar é alterar o nome do host na cadeia de conexão.

Instâncias de Banco de Dados (DB)

O próprio serviço Amazon RDS fornece uma API (Application Programming Interface) que permite criar e gerenciar uma ou mais instâncias de banco de dados.

Uma Instância de banco de dados é um ambiente de banco de dados isolado implantado em seus segmentos de rede privados na nuvem.

Cada instância de banco de dados executa e gerencia um mecanismo de banco de dados comercial ou de código aberto popular em seu nome.

Atualmente, o Amazon RDS suporta os seguintes mecanismos de banco de dados: MySQL, PostgreSQL, MariaDB, Oracle, SQL Server e Amazon Aurora.

Você pode iniciar uma nova instância de banco de dados chamando a API CreateDBInstance ou usando o AWS Management Console. As instâncias de banco de dados existentes podem ser alteradas ou redimensionadas usando a API ModifyDBInstance. Uma instância de banco de dados pode conter vários bancos de

dados diferentes, todos você cria e gerencia dentro da própria Instância de banco de dados executando comandos SQL com o terminal do Amazon RDS.

Os diferentes bancos de dados podem ser criados, acessados e gerenciados usando as mesmas ferramentas e aplicativos do cliente SQL usados hoje.

Os recursos de computação e memória de uma instância de banco de dados são determinados por sua classe de instância de banco de dados.

Você pode selecionar a classe Instância do banco de dados que melhor atenda às suas necessidades de computação e memória.

O intervalo de classes de Instância de banco de dados se estende de um db.t2.micro com 1 CPU virtual (vCPU) e 1 GB de memória, até um db.r3.8xlarge com 32 vCPUs e 244 GB de memória. À medida que suas necessidades mudam com o tempo, você pode alterar a classe da instância e o equilíbrio da computação da memória, e o Amazon RDS migrará seus dados para uma classe de instância maior ou menor.

Independente da classe Instância do banco de dados selecionada, também é possível controlar o tamanho e as características de desempenho do armazenamento usado.

O Amazon RDS suporta uma grande variedade de mecanismos, versões e combinações de recursos.

Muitos recursos e definições de configuração comuns são expostos e gerenciados usando grupos de parâmetros do DB e grupos de opções do BD.

Um grupo de parâmetros do BD atua como um contêiner para a configuração do mecanismo valores que podem ser aplicados a uma ou mais instâncias de banco de dados.

Você pode alterar o grupo de parâmetros do BD para uma instância existente, mas é necessária uma reinicialização.

Um grupo de opções de banco de dados atua como um contêiner para os recursos do mecanismo, que estão vazios por padrão. Para habilitar recursos específicos de um banco de dados (por exemplo, Oracle Statspack, Microsoft SQL Server

Mirroring), crie um novo grupo de opções de banco de dados e defina as configurações adequadamente.

Os bancos de dados existentes podem ser migrados para o Amazon RDS usando ferramentas e técnicas nativas que variam dependendo do mecanismo. Por exemplo, com o MySQL, você pode exportar um backup usando o mysqldump e importar o arquivo para o Amazon RDS MySQL.

Você também pode usar o Serviço de Migração de Banco de Dados da AWS, que fornece uma interface gráfica que simplifica a migração do esquema e dos dados entre os bancos de dados.

O AWS Database Migration Service também ajuda a converter bancos de dados de um mecanismo de banco de dados para outro.

Benefícios Operacionais

O Amazon RDS aumenta a confiabilidade operacional de seus bancos de dados aplicando um modelo operacional e de implantação muito consistente.

Esse nível de consistência é alcançado em parte pela limitação dos tipos de alterações que podem ser feitas na infraestrutura subjacente e pelo uso extensivo da automação.

Por exemplo, com o Amazon RDS, você não pode usar o Secure Shell (SSH) para efetuar login na instância do host e instalar um software personalizado. No entanto, você pode conectar-se usando as ferramentas de administrador do SQL ou usar grupos de opções de banco de dados e grupos de parâmetros de banco de dados para alterar o comportamento ou a configuração de recurso de uma instância de banco de dados.

Se você deseja controle total do sistema operacional (SO) ou precisa de permissões elevadas para executar, considere instalar seu banco de dados no Amazon EC2 em vez do Amazon RDS.

O Amazon RDS foi projetado para simplificar as tarefas comuns necessárias para operar um banco de dados relacional de maneira confiável.

Mecanismos de banco de dados

O Amazon RDS suporta seis mecanismos de banco de dados: MySQL, PostgreSQL, MariaDB, Oracle, SQL Server e Amazon Aurora. Recursos e capacidades variam um pouco, dependendo do mecanismo que você selecionar.

MySQL

O MySQL é um dos bancos de dados de código aberto mais populares do mundo e é usado para alimentar uma ampla gama de aplicativos, de pequenos blogs pessoais a alguns dos maiores sites do mundo.

Até o momento em que este artigo foi escrito, o Amazon RDS for MySQL atualmente suporta o MySQL 8.0.19.

O mecanismo está executando o Community Edition de código aberto com o InnoDB como o mecanismo de armazenamento de banco de dados padrão e recomendado. O Amazon RDS MySQL permite conectar-se usando ferramentas padrão do MySQL, como MySQL Workbench ou SQL Workbench / J.

O Amazon RDS MySQL suporta implantações Multi-AZ para alta disponibilidade e réplicas de leitura para dimensionamento horizontal.

PostgreSQL

O PostgreSQL é um mecanismo de banco de dados de código aberto amplamente utilizado, com um conjunto muito rico de recursos e funcionalidade avançada.

O Amazon RDS suporta instâncias de banco de dados executando várias versões do PostgreSQL. Até o momento da redação deste artigo, o Amazon RDS suporta várias versões do PostgreSQL, incluído a ultima, versão 12.

O Amazon RDS PostgreSQL pode ser gerenciado usando ferramentas padrão como pgAdmin e suporta drivers JDBC / ODBC padrão. O Amazon RDS PostgreSQL também oferece suporte à implantação Multi-AZ para alta disponibilidade e réplicas de leitura para dimensionamento horizontal.

MariaDB

O Amazon RDS recentemente adicionou suporte para instâncias de banco de dados executando o MariaDB. O MariaDB é um popular mecanismo de banco de dados de código aberto, criado pelos criadores do MySQL e aprimorado com ferramentas e funcionalidades empresariais.

O MariaDB adiciona recursos que aprimoram o desempenho, disponibilidade e escalabilidade do MySQL.

O Amazon RDS suporta totalmente o mecanismo de armazenamento XtraDB para as instâncias de banco de dados MariaDB e, como o Amazon RDS MySQL e o PostgreSQL, oferece suporte à implantação Multi-AZ e às réplicas de leitura.

Oracle

O Oracle é um dos bancos de dados relacionais mais populares usados na empresa e é totalmente suportado pelo Amazon RDS.

O Amazon RDS suporta Instâncias de banco de dados executando várias edições do Oracle 11g e Oracle 12c. O Amazon RDS oferece suporte ao acesso a esquemas em uma instância de banco de dados usando qualquer aplicativo cliente padrão do SQL, como o Oracle SQL Plus.

O Amazon RDS Oracle suporta três edições diferentes do popular mecanismo de banco de dados: Standard Edition One, Standard Edition e Enterprise Edition

Microsoft SQL Server

O Microsoft SQL Server é outro banco de dados relacional muito popular utilizado pela Amazon RDS. O Amazon RDS permite que os administradores de banco de

dados (DBAs) se conectem à instância de banco de dados do SQL Server na nuvem usando ferramentas nativas como o SQL Server Management Studio.

O Amazon RDS SQL Server também oferece suporte a quatro edições diferentes do SQL Server: Express Edition, Web Edition, Standard Edition e Enterprise Edition.

Licenciamento

O Amazon RDS Oracle e o Microsoft SQL Server são produtos de software comercial que exigem licenças apropriadas para operar na nuvem.

A AWS oferece dois modelos de licenciamento: Licença Incluída e Traga Sua Própria Licença (BYOL).

- Licença incluída: No modelo Licença incluída, a licença é mantida pela AWS e incluída no preço da instância do Amazon RDS.
    - Para Oracle, a Licença Incluída fornece licenciamento para o Standard Edition One.
    - Para o SQL Server, a Licença Incluída fornece licenciamento para o SQL Server Express Edition, Web Edition e Standard Edition.
- Traga sua própria licença (BYOL): No modelo BYOL, você fornece sua própria licença.
    - Para Oracle, você deve ter a licença apropriada do Oracle Database para a classe DB Instance e a edição do Oracle Database que deseja executar. Você pode adquirir as edições Standard Edition One, Standard Edition e Enterprise Edition.
    - Para o SQL Server, você fornece sua própria licença no programa Microsoft License Mobility. Você pode trazer o Microsoft SQL Standard Edition e também o Enterprise Edition.

Você é responsável por rastrear e gerenciar como as licenças são alocadas.

Amazon Aurora

O Amazon Aurora oferece tecnologia de banco de dados comercial de nível empresarial, oferecendo a simplicidade e a relação custo-benefício de um banco de dados de código aberto.

Isso é obtido através da reformulação dos componentes internos do MySQL para adotar uma abordagem mais orientada a serviços.

Como outros mecanismos do Amazon RDS, o Amazon Aurora é um serviço totalmente gerenciado, compatível com o MySQL e fornece maior confiabilidade e desempenho em implantações padrão do MySQL.

O Amazon Aurora pode fornecer até cinco vezes o desempenho do MySQL sem exigir alterações na maioria dos aplicativos da web existentes.

Você pode usar o mesmo código, ferramentas e aplicativos que usa com os bancos de dados MySQL existentes com o Amazon Aurora. Ao criar uma instância do Amazon Aurora, você cria um cluster de banco de dados.

Um cluster de banco de dados possui uma ou mais instâncias e inclui um volume de cluster que gerencia os dados para essas instâncias.

Um volume de cluster do Amazon Aurora é um volume de armazenamento de banco de dados virtual que abrange várias zonas de disponibilidade, com cada zona de disponibilidade tendo uma cópia dos dados do cluster.

Um cluster do Amazon Aurora DB consiste em dois tipos diferentes de instâncias:

- Instância primária: Esta é a instância principal, que suporta cargas de trabalho de leitura e gravação. Ao modificar seus dados, você está modificando a instância principal. Cada cluster do Amazon Aurora DB possui uma instância principal.
- Réplica do Amazon Aurora: Esta é uma instância secundária que suporta apenas operações de leitura. Cada cluster de banco de dados pode ter até 15 réplicas do Amazon Aurora, além da instância principal. Ao usar várias réplicas do Amazon Aurora, você pode distribuir a carga de trabalho de leitura entre várias instâncias, aumentando o desempenho. Você também

pode localizar as réplicas do Amazon Aurora em várias zonas de disponibilidade para aumentar a disponibilidade do banco de dados.

Opções de Armazenamento

O Amazon RDS é desenvolvido usando o Amazon Elastic Block Store (Amazon EBS) e permite selecionar a opção de armazenamento correta com base em seus requisitos de desempenho e custo.

Dependendo do mecanismo do banco de dados e da carga de trabalho, você pode escalar até 4 a 6 TB em armazenamento provisionado e até 30.000 IOPS. O Amazon RDS suporta três tipos de armazenamento: Magnetic, General Purpose (Solid State Drive [SSD]) e Provisioned IOPS (SSD).

- Magnético: O armazenamento magnético, também chamado de armazenamento padrão, oferece armazenamento econômico, ideal para aplicativos com requisitos leves de E / S.
- Armazenamento de uso geral (SSD), também chamado gp2, pode fornecer acesso mais rápido que o armazenamento magnético. Esse tipo de armazenamento pode fornecer desempenho intermitente para atender a picos e é excelente para bancos de dados pequenos e médios.
- IOPS provisionado (SSD): O armazenamento IOPS provisionado (SSD) foi projetado para atender às necessidades de cargas de trabalho intensivas em E / S, particularmente cargas de trabalho de banco de dados, sensíveis ao desempenho de armazenamento e consistência na taxa de transferência de E / S de acesso aleatório.

Para a maioria das aplicações, o General Purpose (SSD) é a melhor opção e fornece uma boa combinação de características de menor custo e maior desempenho.

Restaurar e Recuperar

O Amazon RDS fornece um modelo operacional consistente para procedimentos de backup e recuperação nos diferentes mecanismos de banco de dados.

O Amazon RDS fornece dois mecanismos para fazer backup do banco de dados: backups automatizados e snapshots manuais. Usando uma combinação das duas técnicas, você pode criar um modelo de recuperação de backup para proteger os dados do aplicativo.

Cada organização normalmente define um RPO (objetivo do ponto de recuperação) e um RTO (objetivo do tempo de recuperação) para aplicativos importantes com base na criticidade do aplicativo e nas expectativas dos usuários.

É comum que os sistemas corporativos tenham um RPO medido em minutos e um RTO medido em horas ou até dias, enquanto alguns aplicativos críticos podem ter tolerâncias muito mais baixas.

RPO é definido como o período máximo de perda de dados aceitável no caso de uma falha ou incidente.

Por exemplo, muitos sistemas fazem backup dos logs de transações a cada 15 minutos para minimizar a perda de dados no caso de uma exclusão acidental ou falha de hardware.

RTO é definido como a quantidade máxima de tempo de inatividade que é permitido recuperar do backup e retomar o processamento.

Para bancos de dados grandes em particular, pode levar horas para restaurar a partir de um backup completo. No caso de uma falha de hardware, você pode reduzir seu RTO para minutos fazendo failover para um nó secundário.

Você deve criar um plano de recuperação que, no mínimo, permita a recuperação de um backup recente.

Backups automatizados

Um backup automatizado é um recurso do Amazon RDS que rastreia continuamente as alterações e faz backup do seu banco de dados.

O Amazon RDS cria um snapshot de volume de armazenamento da sua Instância de banco de dados, fazendo backup de toda a instância do banco de dados e não apenas de bancos de dados individuais.

Você pode definir o período de retenção de backup ao criar uma instância de banco de dados.

Um dia de backups será retido por padrão, mas você pode modificar o período de retenção até um máximo de 35 dias.

Lembre-se de que, quando você exclui uma instância de banco de dados, todos os snapshots de backup automatizados são excluídos e não podem ser recuperados.

Snapshots manuais, no entanto, não são excluídos.

Os backups automatizados ocorrerão diariamente durante uma janela de manutenção configurável de 30 minutos chamada janela de backup.

Os backups automatizados são mantidos por um número configurável de dias, chamado período de retenção de backup.

Você pode restaurar sua instância de banco de dados para um horário específico durante esse período de retenção, criando uma nova instância de banco de dados.

Snapshots de Banco de Dados Manuais

Além dos backups automatizados, você pode executar snapshots de banco de dados manuais a qualquer momento. Um snapshot de banco de dados é iniciado por você e pode ser criado com a frequência que você deseja.

Você pode restaurar a Instância do banco de dados para o estado específico no snapshot do banco de dados a qualquer momento.

Os snapshots do banco de dados podem ser criados com o console do Amazon RDS ou a ação CreateDBSnapshot.

Diferentemente das capturas de snapshots automatizadas que são excluídas após o período de retenção, as capturas de snapshots de banco de dados manuais são mantidas até que você as exclua explicitamente no console do Amazon RDS ou na ação DeleteDBSnapshot.

Para bancos de dados ocupados, use o Multi-AZ para minimizar o impacto no desempenho de um snapshot.

Durante a janela de backup, a E / S de armazenamento pode ser suspensa durante o backup dos dados e você pode enfrentar uma latência elevada.

Essa suspensão de E / S normalmente dura a duração da captura de snapshot.

Esse período de suspensão de E / S é mais curto para implantações do Multi-AZ DB porque o backup é retirado do modo de espera, mas pode ocorrer latência durante o processo de backup.

Recuperação

O Amazon RDS permite recuperar seu banco de dados rapidamente, esteja você executando backups automatizados ou snapshots manuais do banco de dados.

Você não pode restaurar de um snapshot de banco de dados para uma Instância de banco de dados existente. Uma nova instância de banco de dados é criada quando você restaura.

Quando você restaura uma instância de banco de dados, apenas o parâmetro do banco de dados padrão e os grupos de segurança são associados à instância restaurada.

Assim que a restauração for concluída, você deverá associar qualquer parâmetro de banco de dados personalizado ou grupos de segurança usados pela instância a partir da qual você restaurou.

Ao usar backups automatizados, o Amazon RDS combina os backups diários realizados durante a janela de manutenção predefinida em conjunto com os logs de transações para permitir a restauração de instâncias de banco de dados para qualquer ponto durante o seu período de retenção, geralmente até os últimos cinco minutos.

Alta disponibilidade com Multi-AZ

Um dos recursos mais poderosos do Amazon RDS são as implantações Multi-AZ, que permitem criar um cluster de banco de dados em várias zonas de disponibilidade.

A configuração de um banco de dados relacional para execução de maneira altamente disponível e tolerante a falhas é uma tarefa desafiadora.

Com o Amazon RDS Multi-AZ, você pode reduzir a complexidade envolvida nessa tarefa administrativa comum. Com uma única opção, o Amazon RDS pode aumentar a disponibilidade do seu banco de dados usando a replicação.

O Multi-AZ permite que você atenda às metas mais exigentes de RPO e RTO usando replicação síncrona para minimizar o RPO e failover rápido para minimizar o RTO para minutos.

O Multi-AZ permite colocar uma cópia secundária do seu banco de dados em outra zona de disponibilidade para fins de recuperação de desastres. As implantações Multi-AZ estão disponíveis para todos os tipos de mecanismos de banco de dados Amazon RDS.

Quando você cria uma Instância de banco de dados Multi-AZ, uma instância primária é criada em uma zona de disponibilidade e uma instância secundária é criada em outra zona de disponibilidade.

Você recebe um terminal da instância do banco de dados, como o seguinte:

```
my_app_db.ch6fe7ykq1zd.us-west-2.rds.amazonaws.com
```

Esse endpoint é um nome de sistema de nome de domínio (DNS) que a AWS assume a responsabilidade de resolver para um endereço IP específico.

Você usa esse nome DNS ao criar a conexão com seu banco de dados.

O Amazon RDS replica automaticamente os dados do banco de dados mestre ou instância primária para o banco de dados escravo ou instância secundária usando replicação síncrona.

Cada zona de disponibilidade é executada em sua própria infraestrutura fisicamente distinta e independente e é projetada para ser altamente confiável.

O Amazon RDS detecta e se recupera automaticamente dos cenários de falha mais comuns para implantações Multi-AZ, para que você possa retomar as operações do banco de dados o mais rápido possível, sem intervenção administrativa.

O Amazon RDS executa automaticamente um failover no caso de qualquer um dos seguintes:

- Perda de disponibilidade na zona de disponibilidade primária
- Perda de conectividade de rede ao banco de dados primário
- Falha na unidade de computação no banco de dados primário
- Falha de armazenamento no banco de dados primário

O Amazon RDS efetuará failover automaticamente para a instância em espera sem intervenção do usuário. O nome DNS permanece o mesmo, mas o serviço Amazon RDS altera o CNAME para apontar para o modo de espera.

A Instância de banco de dados principal alterna automaticamente para a réplica em espera se houver uma interrupção do serviço da Zona de Disponibilidade, se a Instância de banco de dados principal falhar ou se o tipo de instância for alterado.

Você também pode executar um failover manual da instância do banco de dados.

O failover entre a instância primária e a secundária é rápido, e o tempo necessário para concluir o failover automático é geralmente de um a dois minutos.

É importante lembrar que as implantações Multi-AZ são apenas para recuperação de desastres, eles não foram feitos para melhorar o desempenho do banco de dados.

A Instância de banco de dados em espera não está disponível para consultas offline da Instância de banco de dados principal.

Para melhorar o desempenho do banco de dados usando várias instâncias de banco de dados, use réplicas de leitura ou outras tecnologias de cache de banco de dados, como o Amazon ElastiCache.

Dimensionamento para Cima e para Fora

À medida que o número de transações aumenta para um banco de dados relacional, a expansão ou verticalização de uma máquina maior permite processar mais leituras e gravações.

A expansão horizontal ou vertical também é possível, mas geralmente é mais difícil.

O Amazon RDS permite dimensionar a computação e o armazenamento verticalmente e, para alguns mecanismos de banco de dados, você pode dimensionar horizontalmente.

Escalabilidade Vertical

A adição de recursos adicionais de computação, memória ou armazenamento ao seu banco de dados permite processar mais transações, executar mais consultas e armazenar mais dados.

O Amazon RDS facilita o aumento ou a redução da camada do banco de dados para atender às demandas do seu aplicativo.

As alterações podem ser agendadas para ocorrer durante a próxima janela de manutenção ou começar imediatamente usando a ação ModifyDBInstance.

Para alterar a quantidade de computação e memória, você pode selecionar uma classe de Instância de banco de dados diferente do banco de dados.

Depois de selecionar uma classe de instância de banco de dados maior ou menor, o Amazon RDS automatiza o processo de migração para uma nova classe com apenas uma pequena interrupção e um esforço mínimo.

Você também pode aumentar a quantidade de armazenamento, a classe de armazenamento e o desempenho de armazenamento para uma instância do Amazon RDS.

Cada instância do banco de dados pode escalar de 5 GB a 6 TB no armazenamento provisionado, dependendo do tipo e do mecanismo de armazenamento.

O armazenamento para Amazon RDS pode ser aumentado com o tempo, à medida que as necessidades aumentam com um impacto mínimo no banco de dados em execução.

A expansão de armazenamento é suportada para todos os mecanismos de banco de dados, exceto para o SQL Server.

Escalabilidade horizontal com particionamento

Um banco de dados relacional pode ser dimensionado verticalmente apenas muito antes de você atingir o tamanho máximo da instância.

Particionar um grande banco de dados relacional em várias instâncias ou shards é uma técnica comum para lidar com mais solicitações além dos recursos de uma única instância.

O particionamento, ou sharding, permite escalar horizontalmente para lidar com mais usuários e solicitações, mas requer lógica adicional na camada do aplicativo.

O aplicativo precisa decidir como rotear as solicitações do banco de dados para o shard correto e fica limitado nos tipos de consultas que pode ser executado através dos limites do servidor.

Os bancos de dados NoSQL, como o Amazon DynamoDB ou o Cassandra, foram projetados para serem dimensionados horizontalmente.

Escalabilidade Horizontal com Réplicas de Leitura

Outra técnica de dimensionamento importante é usar réplicas de leitura para descarregar transações de leitura do banco de dados primário e aumentar o número geral de transações.

O Amazon RDS suporta réplicas de leitura que permitem expandir de maneira elástica além das restrições de capacidade de uma única instância de banco de dados para cargas de trabalho de banco de dados com muita leitura.

Há vários casos de uso em que a implantação de uma ou mais instâncias de banco de dados de réplica de leitura é útil. Alguns cenários comuns incluem:

- Escale além da capacidade de uma única instância de banco de dados para cargas de trabalho com muita leitura.
- Manipule o tráfego de leitura enquanto a Instância do banco de dados de origem não estiver disponível. Por exemplo, devido à suspensão de E / S para

backups ou manutenção agendada, você pode direcionar o tráfego de leitura para uma réplica.

- Descarregue cenários de relatório ou data warehousing em uma réplica em vez da Instância de banco de dados principal. Por exemplo, um site de blog pode ter muito pouca atividade de gravação, exceto os comentários ocasionais, e a grande maioria das atividades do banco de dados será somente leitura. Ao descarregar parte ou toda a atividade de leitura para uma ou mais réplicas de leitura, a instância principal do banco de dados pode se concentrar em manipular as gravações e replicar os dados nas réplicas.

Atualmente, as réplicas de leitura são suportadas no Amazon RDS para MySQL, PostgreSQL, MariaDB e Amazon Aurora.

O Amazon RDS usa a funcionalidade de replicação interna dos mecanismos MySQL, MariaDB e PostgreSQL DB para criar um tipo especial de Instância de banco de dados, chamada réplica de leitura, de uma Instância de banco de dados de origem.

As atualizações feitas na Instância do banco de dados de origem são copiadas de forma assíncrona na réplica de leitura. Você pode reduzir a carga na sua Instância de banco de dados de origem, roteando consultas de leitura de seus aplicativos para a réplica de leitura.

Você pode criar uma ou mais réplicas de um banco de dados em uma única região da AWS ou em várias regiões da AWS.

Para aprimorar seus recursos de recuperação de desastres ou reduzir latências globais, você pode usar réplicas de leitura entre regiões para servir o tráfego de leitura de uma região mais próxima dos usuários globais ou migrar seus bancos de dados pelas regiões da AWS

Segurança

Proteger suas instâncias de banco de dados Amazon RDS DB e bancos de dados relacionais requer um plano abrangente que aborde as muitas camadas comumente encontradas em sistemas orientados a bancos de dados. Isso inclui os recursos de infraestrutura, o banco de dados e a rede.

Proteja o acesso aos seus recursos de infraestrutura usando as políticas do AWS Identity and Access Management (IAM) que limitam quais ações os administradores da AWS podem executar.

Por exemplo, algumas ações principais do administrador que podem ser controladas no IAM incluem CreateDBInstance e DeleteDBInstance.

Outra prática recomendada de segurança é implantar suas instâncias do Amazon RDS DB em uma sub-rede privada dentro de uma Amazon Virtual Private Cloud (Amazon VPC) que limita o acesso da rede à instância do DB.

Antes de poder implantar em um Amazon VPC, você deve primeiro criar um grupo de sub-redes de bancos de dados que predefine quais sub-redes estão disponíveis para implantações do Amazon RDS.

Além disso, restrinja o acesso à rede usando listas de controle de acesso (ACLs) e grupos de segurança para limitar o tráfego de entrada a uma pequena lista de endereços IP de origem.

No nível do banco de dados, você também precisará criar usuários e conceder a eles permissões para ler e gravar em seus bancos de dados.

O acesso ao banco de dados é controlado usando os mecanismos específicos de controle de acesso e gerenciamento de usuários do mecanismo de banco de dados.

Crie usuários no nível do banco de dados com senhas fortes que você alterna frequentemente.

Por fim, proteja a confidencialidade de seus dados em trânsito e em repouso com vários recursos de criptografia fornecidos com o Amazon RDS.

Os recursos de segurança variam um pouco de um mecanismo para outro, mas todos os mecanismos oferecem suporte a alguma forma de criptografia em trânsito e também em criptografia em repouso.

Você pode conectar com segurança um cliente a uma Instância de banco de dados em execução usando o Secure Sockets Layer (SSL) para proteger os dados em trânsito. A criptografia em repouso é possível para todos os mecanismos que usam o Amazon Key Management Service (KMS) ou o Transparent Data Encryption (TDE).

Todos os logs, backups e capturas instantâneas são criptografados para uma instância criptografada do Amazon RDS

Amazon Redshift

O Amazon Redshift é um serviço de data warehouse rápido, poderoso, totalmente gerenciado e em escala de petabytes na nuvem. O Amazon Redshift é um banco de dados relacional projetado para cenários OLAP e otimizado para análises e relatórios de alto desempenho de conjuntos de dados muito grandes.

Os data warehouses tradicionais são difíceis e caros de gerenciar, especialmente para grandes conjuntos de dados. O Amazon Redshift não apenas reduz significativamente o custo de um data warehouse, mas também facilita a análise rápida de grandes quantidades de dados.

O Amazon Redshift oferece recursos de consulta rápida sobre dados estruturados usando comandos SQL padrão para oferecer suporte a consultas interativas em grandes conjuntos de dados.

Com conectividade via ODBC ou JDBC, o Amazon Redshift se integra bem a vários carregamentos, relatórios, mineração de dados e ferramentas de análise.

O Amazon Redshift é baseado no PostgreSQL padrão do setor, portanto, a maioria dos aplicativos clientes SQL existentes funcionará com apenas alterações mínimas.

O Amazon Redshift gerencia o trabalho necessário para configurar, operar e dimensionar um data warehouse, desde o provisionamento da capacidade da infraestrutura até a automação de tarefas administrativas em andamento, como backups e correções.

O Amazon Redshift monitora automaticamente seus nós e unidades para ajudá-lo a se recuperar de falhas.

Clusters e Nós

O principal componente de um data warehouse do Amazon Redshift é um cluster. Um cluster é composto por um nó líder e um ou mais nós de computação. O aplicativo cliente interage diretamente apenas com o nó líder e os nós de computação são transparentes para aplicativos externos.

Atualmente, o Amazon Redshift tem suporte para seis tipos diferentes de nós e cada um tem uma mistura diferente de CPU, memória e armazenamento.

Os seis tipos de nós estão agrupados em duas categorias: Computação Densa e Armazenamento Denso.

Os tipos de nós do Computação Densa suportam clusters de até 326 TB usando SSDs rápidos, enquanto os nós de armazenamento denso suportam clusters de até 2PB usando grandes discos magnéticos.

Cada cluster consiste em um nó líder e um ou mais nós de computação. Cada cluster contem um ou mais bancos de dados. Os dados do usuário para cada tabela são distribuídos pelos nós de computação.

Seu aplicativo ou cliente SQL se comunica com o Amazon Redshift usando conexões JDBC ou ODBC padrão com o nó líder, que por sua vez coordena a execução da consulta com os nós de computação.

Seu aplicativo não interage diretamente com os nós de computação.

O armazenamento em disco para um nó de computação é dividido em várias fatias. O número de fatias por nó depende do tamanho do nó do cluster e normalmente varia entre 2 e 16.

Todos os nós participam da execução da consulta paralela, trabalhando em dados que são distribuídos o mais uniformemente possível pelas fatias.

Você pode aumentar o desempenho da consulta adicionando vários nós a um cluster. Quando você envia uma consulta, o Amazon Redshift distribui e executa a consulta em paralelo em todos os nós de computação de um cluster. O Amazon

Redshift também espalha seus dados de tabela em toda a computação em nós em um cluster com base em uma estratégia de distribuição que você especificar.

Esse particionamento de dados entre vários recursos de computação permite atingir altos níveis de desempenho.

O Amazon Redshift permite redimensionar um cluster para adicionar armazenamento e capacidade de computação ao longo do tempo à medida que suas necessidades evoluem.

Você também pode alterar o tipo de nó de um cluster e manter o tamanho geral igual.

Sempre que você executa uma operação de redimensionamento, o Amazon Redshift cria um novo cluster e migra os dados do cluster antigo para o novo.

Durante uma operação de redimensionamento, o banco de dados se tornará somente leitura até que a operação seja concluída.

Design

Cada cluster do Amazon Redshift pode suportar um ou mais bancos de dados e cada banco de dados pode conter muitas tabelas.

Como a maioria dos bancos de dados baseados em SQL, você pode criar uma tabela usando o comando CREATE TABLE. Este comando especifica o nome da tabela, as colunas e seus tipos de dados.

Além de colunas e tipos de dados, o comando Amazon Redshift CREATE TABLE também suporta a especificação de codificações de compactação, estratégia de distribuição e chaves de classificação.

Tipos de dados

As colunas do Amazon Redshift oferecem suporte a uma ampla variedade de tipos de dados. Isso inclui tipos de dados numéricos comuns como INTEGER, DECIMAL e

DOUBLE, tipos de dados de texto como CHAR e VARCHAR e tipos de dados de data como DATE e TIMESTAMP.

Colunas adicionais podem ser adicionadas a uma tabela usando o comando ALTER TABLE. No entanto, as colunas existentes não podem ser modificadas.

Codificação de Compressão

Uma das principais otimizações de desempenho usadas pelo Amazon Redshift é a compactação de dados. Ao carregar dados pela primeira vez em uma tabela vazia, o Amazon Redshift irá amostrar automaticamente seus dados e selecionar o melhor esquema de compactação para cada coluna.

Como alternativa, você pode especificar a codificação de compactação por coluna como parte do comando CREATE TABLE.

Estratégia de distribuição

Uma das principais decisões ao criar uma tabela no Amazon Redshift é como distribuir os registros pelos nós e fatias em um cluster.

Você pode configurar o estilo de distribuição de uma tabela para fornecer dicas do Amazon Redshift sobre como os dados devem ser particionados para melhor atender seus padrões de consulta.

Quando você executa uma consulta, o otimizador muda as linhas para os nós conforme necessário para executar quaisquer junções e agregados.

O objetivo na seleção de um estilo de distribuição de tabela é minimizar o impacto da etapa de redistribuição, colocando os dados onde eles precisam estar antes da execução da consulta.

O estilo de distribuição de dados que você seleciona para o banco de dados tem um grande impacto no desempenho da consulta, nos requisitos de armazenamento, no carregamento e na manutenção de dados.

Ao escolher a melhor estratégia de distribuição para cada tabela, você pode equilibrar sua distribuição de dados e melhorar significativamente o desempenho geral do sistema.

Ao criar uma tabela, você pode escolher entre um dos três estilos de distribuição: MESMO, CHAVE ou TODOS.

Distribuição uniforme

Essa é a opção padrão e resulta na distribuição uniforme dos dados nas fatias, independentemente dos dados.

Distribuição de chaves

Com a distribuição KEY, as linhas são distribuídas de acordo com os valores em uma coluna. O nó líder armazenará valores correspondentes próximos e aumentará o desempenho da consulta para junções.

ALL distribuição

Com ALL, uma cópia completa da tabela inteira é distribuída para cada nó. Isso é útil para tabelas de pesquisa e outras tabelas grandes que não são atualizadas frequentemente.

Chaves de classificação

Outra decisão importante a ser tomada durante a criação de uma tabela é especificar uma ou mais colunas como chaves de classificação.

A classificação permite o manuseio eficiente de predicados com restrição de intervalo. Se uma consulta usar um predicado com intervalo restrito, o processador de consultas poderá ignorar rapidamente muitos blocos durante as varreduras de tabela.

As chaves de classificação para uma tabela podem ser compostas ou intercaladas. Uma chave de classificação composta é mais eficiente quando os predicados da consulta usam um prefixo, que é um subconjunto das colunas da chave de classificação em ordem.

Uma chave de classificação intercalada fornece peso igual a cada coluna na chave de classificação, para que os predicados de consulta possam usar qualquer subconjunto das colunas que compõem a chave de classificação, em qualquer ordem.

Carregando dados

O Amazon Redshift suporta comandos SQL padrão, como INSERT e UPDATE, para criar e modificar registros em uma tabela.

Para operações em massa, no entanto, o Amazon Redshift fornece o comando COPY como uma alternativa muito mais eficiente do que chamar repetidamente INSERT.

Um comando COPY pode carregar dados em uma tabela da maneira mais eficiente e suporta vários tipos de fontes de dados de entrada.

A maneira mais rápida de carregar dados no Amazon Redshift é fazer carregamentos de dados em massa de arquivos simples armazenados em um Amazon S3 ou de uma tabela do Amazon DynamoDB.

Ao carregar dados do Amazon S3, o comando COPY pode ler de vários arquivos ao mesmo tempo.

O Amazon Redshift pode distribuir a carga de trabalho para os nós e executar o processo de carregamento em paralelo. Em vez de ter um único arquivo grande com seus dados, você pode ativar o processamento paralelo com um cluster com vários nós e vários arquivos de entrada.

Após cada carregamento de dados em massa que modifica uma quantidade significativa de dados, você precisará executar um comando VACUUM para reorganizar seus dados e recuperar espaço após exclusões.

Também é recomendável executar um comando ANALYZE para atualizar as estatísticas da tabela.

Os dados também podem ser exportados para fora do Amazon Redshift usando o comando UNLOAD. Este comando pode ser usado para gerar arquivos de texto delimitados e armazená-los no Amazon S3.

Consultando Dados

O Amazon Redshift permite escrever comandos SQL padrão para consultar suas tabelas. Ao oferecer suporte a comandos como o SELECT para consultar e ingressar em tabelas, os analistas podem se tornar produtivos rapidamente usando o Amazon Redshift ou integrá-lo facilmente.

Para consultas complexas, você pode analisar o plano de consulta para otimizar melhor seu padrão de acesso. Você pode monitorar o desempenho do cluster e consultas específicas usando o Amazon CloudWatch e o console da web do Amazon Redshift.

Para grandes clusters do Amazon Redshift que oferecem suporte a muitos usuários, você pode configurar o Workload Management (WLM) para enfileirar e priorizar consultas. O WLM permite definir várias filas e definir o nível de simultaneidade para cada fila.

Por exemplo, convém configurar uma fila para consultas de longa execução e limitar a simultaneidade e outra fila para consultas de execução curta e permitem níveis mais altos de simultaneidade.

Snapshots

Semelhante ao Amazon RDS, você pode criar snapshots point-in-time do seu cluster Amazon Redshift. Um snapshot pode ser usado para restaurar uma cópia ou criar um clone do seu cluster Amazon Redshift original.

Os snapshots são armazenados de forma durável internamente no Amazon S3 pelo Amazon Redshift.

O Amazon Redshift suporta snapshots automatizados e manuais. Com os snapshots automatizados, o Amazon Redshift tira periodicamente snapshots do seu cluster e mantém uma cópia por um período de retenção configurável.

Você também pode executar snapshots manuais e compartilhá-los entre regiões ou mesmo com outras contas da AWS. As capturas instantâneas manuais são mantidas até que você as exclua explicitamente.

Segurança

Proteger o cluster do Amazon Redshift é semelhante a proteger outros bancos de dados em execução na nuvem. Seu plano de segurança deve incluir controles para proteger os recursos da infraestrutura, o esquema do banco de dados, os registros na tabela e o acesso à rede.

Ao abordar a segurança em todos os níveis, você pode operar com segurança um data warehouse do Amazon Redshift na nuvem.

A primeira camada de segurança vem no nível da infraestrutura usando políticas do IAM que limitam as ações que os administradores da AWS podem executar. Com o IAM, você pode criar políticas que concedam a outros usuários da AWS a permissão para criar e gerenciar o ciclo de vida de um cluster, incluindo operações de dimensionamento, backup e recuperação.

No nível da rede, os clusters do Amazon Redshift podem ser implantados no espaço de endereço IP privado do Amazon VPC para restringir a conectividade geral da rede.

O acesso à rede refinada pode ser ainda mais restrito usando grupos de segurança e ACLs de rede no nível da sub-rede.

Além de controlar o acesso à infraestrutura no nível da infraestrutura, você deve proteger o acesso no nível do banco de dados. Ao criar inicialmente um cluster do Amazon Redshift, você criará uma conta de usuário mestre e uma senha.

A conta principal pode ser usada para fazer login no banco de dados Amazon Redshift e criar mais usuários e grupos. Cada usuário do banco de dados pode receber permissão para esquemas, tabelas e outros objetos de banco de dados.

Essas permissões são independentes das políticas do IAM usadas para controlar o acesso aos recursos de infraestrutura e à configuração de cluster do Amazon Redshift.

Proteger os dados armazenados no Amazon Redshift é outro aspecto importante do seu design de segurança. O Amazon Redshift oferece suporte à criptografia de dados em trânsito usando conexões criptografadas por SSL e também à criptografia de dados em repouso usando várias técnicas.

Para criptografar dados em repouso, o Amazon Redshift se integra ao KMS e ao AWS CloudHSM para serviços de gerenciamento de chaves de criptografia.

A criptografia em repouso e em trânsito ajuda a atender aos requisitos de conformidade, como a Lei de Portabilidade e Responsabilidade do Seguro de Saúde (HIPAA) ou o Padrão de Segurança de Dados da Indústria de Cartões de Pagamento (PCI DSS) e fornece proteções adicionais para seus dados.

Amazon DynamoDB

O Amazon DynamoDB é um serviço de banco de dados NoSQL totalmente gerenciado que fornece desempenho rápido e de baixa latência, dimensionado com facilidade.

O Amazon DynamoDB permite descarregar os encargos administrativos da operação de um banco de dados NoSQL distribuído e focar no aplicativo. O Amazon DynamoDB simplifica significativamente o provisionamento, instalação e configuração de hardware, replicação, aplicação de patches de software e dimensionamento de cluster dos bancos de dados NoSQL.

O Amazon DynamoDB foi projetado para simplificar o gerenciamento de bancos de dados e cluster, fornecer níveis consistentemente altos de desempenho, simplificar tarefas de escalabilidade e melhorar a confiabilidade com replicação automática.

Os desenvolvedores podem criar uma tabela no Amazon DynamoDB e escrever um número ilimitado de itens com latência consistente.

O Amazon DynamoDB pode fornecer níveis de desempenho consistentes distribuindo automaticamente os dados e o tráfego para uma tabela em várias partições.

Depois de configurar uma certa capacidade de leitura ou gravação, o Amazon DynamoDB adicionará automaticamente capacidade de infraestrutura suficiente para apoiar os níveis de taxa de transferência solicitados.

À medida que sua demanda muda com o tempo, você pode ajustar a capacidade de leitura ou gravação após a criação de uma tabela, e o Amazon DynamoDB adicionará ou removerá a infraestrutura e ajustará o particionamento interno de acordo.

Para ajudar a manter níveis de desempenho rápidos e consistentes, todos os dados da tabela são armazenados em unidades de disco SSD de alto desempenho.

Métricas de desempenho, incluindo taxas de transações, podem ser monitoradas usando o Amazon CloudWatch. Além de fornecer níveis de alto desempenho, o Amazon DynamoDB também fornece proteções automáticas de alta disponibilidade e durabilidade, replicando dados em várias zonas de disponibilidade em uma região da AWS.

Modelo de Dados

Os componentes básicos do modelo de dados do Amazon DynamoDB incluem tabelas, itens e atributos.

Em um banco de dados relacional, uma tabela possui um esquema predefinido, como o nome da tabela, a chave primária, a lista dos nomes de suas colunas e seus tipos de dados.

Todos os registros armazenados na tabela devem ter o mesmo conjunto de colunas. Por outro lado, o Amazon DynamoDB exige apenas que uma tabela tenha uma chave primária, mas não requer que você defina todos os nomes de atributos e tipos de dados com antecedência.

Itens individuais em uma tabela do Amazon DynamoDB podem ter qualquer número de atributos, embora exista um limite de 400 KB no tamanho do item.

Cada atributo em um item é um par de nome / valor. Um atributo pode ser um conjunto de valor único ou de valores múltiplos.

Por exemplo, um item de livro pode ter atributos de título e autores. Cada livro tem um título, mas pode ter muitos autores.

Os aplicativos podem se conectar ao terminal do serviço Amazon DynamoDB e enviar solicitações por HTTP / S para ler e gravar itens em uma tabela ou até mesmo para criar e excluir tabelas.

O DynamoDB fornece uma API de serviço da web que aceita solicitações no formato JSON.

Embora você possa programar diretamente nos endpoint da API de serviço web, a maioria dos desenvolvedores optam por usar o AWS Software Development Kit (SDK) para interagir com seus itens e tabelas.

O AWS SDK está disponível em muitos idiomas diferentes e fornece uma interface de programação simplificada e de alto nível.

Tipos de Dados

O Amazon DynamoDB oferece muita flexibilidade com seu esquema de banco de dados. Ao contrário de um banco de dados relacional tradicional que requer que

você defina seus tipos de coluna com antecedência, o DynamoDB requer apenas um atributo de chave primária.

Cada item adicionado à tabela pode adicionar atributos adicionais. Isso oferece flexibilidade ao longo do tempo para expandir seu esquema, sem precisar reconstruir a tabela inteira e lidar com as diferenças de versão de registro com a lógica do aplicativo.

Ao criar uma tabela ou um índice secundário, você deve especificar os nomes e os tipos de dados de cada atributo de chave primária (chave de partição e chave de classificação).

O Amazon DynamoDB suporta uma ampla variedade de tipos de dados para atributos. Os tipos de dados se enquadram em três categorias principais: Escalar, Conjunto ou Documento.

Tipos de dados escalares

Um tipo escalar representa exatamente um valor. O Amazon DynamoDB suporta os seguintes cinco tipos escalares:

String

Texto e caracteres de comprimento variável de até 400 KB. Suporta Unicode com codificação UTF8

Número

Número positivo ou negativo com até 38 dígitos de precisão

Binário

Dados binários, imagens, objetos compactados com tamanho de até 400 KB

Boleano

Sinalizador binário que representa um valor verdadeiro ou falso

Nulo

Representa um estado em branco, vazio ou desconhecido. String, Number, Binary, Boolean não podem estar vazios.

Definir tipos de dados

Os conjuntos são úteis para representar uma lista exclusiva de um ou mais valores escalares. Cada valor em um conjunto precisa ser exclusivo e deve ter o mesmo tipo de dados. Os conjuntos não garantem a ordem. O Amazon DynamoDB suporta três tipos de conjuntos: Conjunto de cadeias, Conjunto de números e Conjunto binário.

String

Definir lista exclusiva de atributos String

Número

Definir lista exclusiva de atributos numéricos

Binário

Definir lista exclusiva de atributos binários

Chave primária

Ao criar uma tabela, você deve especificar a chave primária da tabela, além do nome da tabela. Como um banco de dados relacional, a chave primária identifica exclusivamente cada item da tabela.

Uma chave primária apontará para exatamente um item. O Amazon DynamoDB oferece suporte a dois tipos de chaves primárias e essa configuração não pode ser alterada após a criação de uma tabela:

Chave de Partição

A chave primária é composta de um atributo, uma chave de partição (ou hash). O Amazon DynamoDB cria um índice de hash não ordenado nesse atributo de chave primária.

Chave de Partição e Classificação

A chave primária é composta por dois atributos. O primeiro atributo é a chave de partição e o segundo é a chave de classificação (ou intervalo).

Cada item da tabela é identificado exclusivamente pela combinação de sua partição e valores da chave de classificação. É possível que dois itens tenham o mesmo valor da chave de partição, mas esses dois itens devem ter valores diferentes da chave de classificação.

Além disso, cada atributo de chave primária deve ser definido como tipo string, número ou binário.

O Amazon DynamoDB usa a chave de partição para distribuir a solicitação para a partição correta.

Se você estiver executando muitas leituras ou gravações por segundo na mesma chave primária, não poderá usar totalmente a capacidade de computação do cluster do Amazon DynamoDB.

Uma prática recomendada é maximizar sua taxa de transferência distribuindo solicitações em toda a gama de chaves de partição.

Capacidade provisionada

Ao criar uma tabela do Amazon DynamoDB, você deve fornecer uma certa quantidade de capacidade de leitura e gravação para lidar com as cargas de trabalho esperadas.

Com base nas definições de configuração, o DynamoDB fornecerá a quantidade certa de capacidade de infraestrutura para atender aos seus requisitos com tempos de resposta sustentados e de baixa latência. A capacidade geral é medida em unidades de capacidade de leitura e gravação.

Posteriormente, esses valores podem ser ampliados ou reduzidos usando uma ação UpdateTable.

Cada operação em uma tabela do Amazon DynamoDB consumirá algumas das unidades de capacidade provisionadas.

A quantidade específica de unidades de capacidade consumidas depende em grande parte do tamanho do item, mas também de outros fatores.

Para operações de leitura, a quantidade de capacidade consumida também depende da consistência de leitura selecionada na solicitação.

Por exemplo, dada uma tabela sem um índice secundário local, você consumirá 1 unidade de capacidade se ler um item com 4KB ou menor.

Da mesma forma, para operações de gravação, você consumirá 1 unidade de capacidade se gravar um item com 1 KB ou menos. Isso significa que, se você ler um item com 110 KB, consumirá 28 unidades de capacidade ou 110/4 = 27,5 arredondado para 28.

Para operações de leitura altamente consistentes, elas usarão o dobro do número de unidades de capacidade, ou 56 neste exemplo.

Você pode usar o Amazon CloudWatch para monitorar sua capacidade do Amazon DynamoDB e tomar decisões de dimensionamento. Há um rico conjunto de métricas, incluindo ConsumedReadCapacityUnits e ConsumedWriteCapacityUnits.

Se você exceder sua capacidade provisionada por um período de tempo, as solicitações serão limitadas e poderão ser tentadas novamente mais tarde. Você

pode monitorar e alertar na métrica ThrottledRequests usando o Amazon CloudWatch para notificá-lo sobre alterações nos padrões de uso.

Índices Secundários

Ao criar uma tabela com uma chave de partição e de classificação (anteriormente conhecida como chave de hash e intervalo), é possível definir opcionalmente um ou mais índices secundários nessa tabela.

Um índice secundário permite consultar os dados na tabela usando uma chave alternativa, além de consultas na chave primária. O Amazon DynamoDB suporta dois tipos diferentes de índices:

Índice Secundário Global

O índice secundário global é um índice com uma chave de partição e classificação que pode ser diferente da tabela.

Você pode criar ou excluir um índice secundário global em uma tabela a qualquer momento.

Índice Secundário Local

O índice secundário local é um índice que possui o mesmo atributo de chave de partição que a chave primária da tabela, mas uma chave de classificação diferente.

Você só pode criar um índice secundário local quando criar uma tabela.

Os índices secundários permitem procurar uma tabela grande com eficiência e evitar uma operação cara de digitalização para encontrar itens com atributos específicos.

Esses índices permitem suportar diferentes padrões de acesso à consulta e casos de uso além do possível com apenas uma chave primária. Enquanto uma tabela pode ter apenas um índice secundário local, você pode ter vários índices secundários globais.

O Amazon DynamoDB atualiza cada índice secundário quando um item é modificado.

Essas atualizações consomem unidades de capacidade de gravação. Para um índice secundário local, as atualizações de itens consumirão unidades de capacidade de gravação da tabela principal, enquanto os índices secundários globais mantêm suas próprias configurações de taxa de transferência provisionadas separadas da tabela.

Escrevendo e lendo dados

Depois de criar uma tabela com uma chave primária e índices, você pode começar a escrever e ler itens na tabela.

O Amazon DynamoDB fornece várias operações que permitem criar, atualizar e excluir itens individuais. O Amazon DynamoDB também fornece várias opções de consulta que permitem pesquisar uma tabela ou um índice ou recuperar um item específico ou um lote de itens.

Escrevendo itens

O Amazon DynamoDB fornece três ações principais da API para criar, atualizar e excluir itens: PutItem, UpdateItem e DeleteItem.

Usando a ação PutItem, você pode criar um novo item com um ou mais atributos. As chamadas para PutItem atualizarão um item existente se a chave primária já existir.

PutItem requer apenas um nome de tabela e uma chave primária; qualquer atributos adicionais são opcionais. A ação UpdateItem encontrará itens existentes com base na chave primária e substituirá os atributos.

Esta operação pode ser útil para atualizar apenas um único atributo e deixar os outros atributos inalterados. O UpdateItem também pode ser usado para criar itens se eles ainda não existirem.

Por fim, você pode remover um item de uma tabela usando DeleteItem e especificando uma chave primária específica.

A ação UpdateItem também fornece suporte para contadores atômicos. Os contadores atômicos permitem aumentar e diminuir um valor e garantem a consistência entre várias solicitações simultâneas.

Por exemplo, um atributo de contador usado para rastrear a pontuação geral de um jogo para celular pode ser atualizado por muitos clientes ao mesmo tempo.

Essas três ações também suportam expressões condicionais que permitem executar a validação antes que uma ação seja aplicada.

Por exemplo, você pode aplicar uma expressão condicional em PutItem que verifica se certas condições são atendidas antes da criação do item.

Isso pode ser útil para evitar substituições acidentais ou para impor algum tipo de verificação da lógica de negócios.

Lendo itens

Após a criação de um item, ele pode ser recuperado através de uma pesquisa direta chamando a ação GetItem ou através de uma pesquisa usando a ação Consulta ou Verificação.

GetItem permite recuperar um item com base em sua chave primária. Todos os atributos do item são retornados por padrão, e você tem a opção de selecionar atributos individuais para filtrar os resultados.

Se uma chave primária for composta por uma chave de partição, toda a chave de partição precisará ser especificada para recuperar o item. Se a chave primária for composta de uma chave de partição e uma chave de classificação, GetItem também exigirá a chave da partição e da chave de classificação.

Cada chamada para GetItem consome unidades de capacidade de leitura com base no tamanho do item e na opção de consistência selecionada.

Por padrão, uma operação GetItem executa uma leitura eventualmente consistente. Opcionalmente, você pode solicitar uma leitura fortemente consistente; isso consumirá unidades de capacidade de leitura adicionais, mas retornará a versão mais atualizada do item.

Consistência Eventual

Ao ler itens do Amazon DynamoDB, a operação pode eventualmente ser consistente ou fortemente consistente.

O Amazon DynamoDB é um sistema distribuído que armazena várias cópias de um item em uma região da AWS para fornecer alta disponibilidade e maior durabilidade.

Quando um item é atualizado no Amazon DynamoDB, ele começa a replicar em vários servidores. Como o Amazon DynamoDB é um sistema distribuído, a replicação pode levar algum tempo para ser concluída.

Por esse motivo, nos referimos aos dados como eventualmente consistentes, o que significa que uma solicitação de leitura imediatamente após uma operação de gravação pode não mostrar a alteração mais recente.

Em alguns casos, o aplicativo precisa garantir que os dados sejam os mais recentes e o Amazon DynamoDB oferece uma opção para leituras fortemente consistentes.

Leituras eventualmente consistentes

Quando você lê dados, a resposta pode não refletir os resultados de uma operação de gravação concluída recentemente. A resposta pode incluir alguns dados antigos.

A consistência em todas as cópias dos dados geralmente é alcançada em um segundo; se você repetir sua solicitação de leitura após um curto período de tempo, a resposta retornará os dados mais recentes.

Leituras fortemente consistentes

Ao emitir uma solicitação de leitura fortemente consistente, o Amazon DynamoDB retorna uma resposta com os dados mais atualizados que refletem atualizações de

todas as operações de gravação relacionadas anteriores às quais o Amazon DynamoDB retornou uma resposta bem-sucedida.

Uma leitura fortemente consistente pode estar menos disponível em caso de atraso ou interrupção da rede. Você pode solicitar um resultado de leitura fortemente consistente especificando parâmetros opcionais em sua solicitação.

Operações em lote

O Amazon DynamoDB também fornece várias operações projetadas para trabalhar com grandes lotes de itens, incluindo BatchGetItem e BatchWriteItem. Usando a ação BatchWriteItem, você pode executar até 25 itens criados ou atualizados com uma única operação.

Isso permite minimizar a sobrecarga de cada chamada individual ao processar um grande número de itens.

Pesquisando itens

O Amazon DynamoDB também oferece duas operações, Consulta e Verificação, que podem ser usadas para pesquisar uma tabela ou um índice.

Uma operação de consulta é a operação de pesquisa principal que você pode usar para localizar itens em uma tabela ou em um índice secundário usando apenas valores de atributo da chave primária. Cada consulta requer um nome de atributo de chave de partição e um valor distinto para pesquisar.

Opcionalmente, você pode fornecer um valor de chave de classificação e usar um operador de comparação para refinar os resultados da pesquisa. Os resultados são classificados automaticamente pela chave primária e são limitados a 1 MB.

Ao contrário de uma consulta, uma operação de digitalização lê todos os itens de uma tabela ou de um índice secundário.

Por padrão, uma operação de Varredura retorna todos os atributos de dados para cada item da tabela ou índice.

Cada solicitação pode retornar até 1 MB de dados. Os itens podem ser filtrados usando expressões, mas isso pode ser uma operação que consome muitos recursos. Se o conjunto de resultados de uma Consulta ou Varredura exceder 1 MB, você poderá paginar os resultados em incrementos de 1 MB.

Para a maioria das operações, executar uma operação de consulta em vez de uma operação de digitalização será a opção mais eficiente. A execução de uma operação de Verificação resultará em uma verificação completa de toda a tabela ou índice secundário e, em seguida, filtra os valores para fornecer o resultado desejado.

Use uma operação de consulta quando possível e evite uma digitalização em uma tabela ou índice grande para apenas um pequeno número de itens.

Segurança

O Amazon DynamoDB oferece controle granular sobre os direitos e permissões de acesso para usuários e administradores. O Amazon DynamoDB integra-se ao serviço IAM para fornecer forte controle sobre permissões usando políticas.

Você pode criar uma ou mais políticas que permitem ou negam operações específicas em tabelas específicas. Você também pode usar condições para restringir o acesso a itens ou atributos individuais.

Todas as operações devem primeiro ser autenticadas como um usuário ou sessão de usuário válido. Os aplicativos que precisam ler e gravar no Amazon DynamoDB precisam obter um conjunto de chaves de controle de acesso temporário ou permanente.

Embora essas chaves possam ser armazenadas em um arquivo de configuração, uma prática recomendada é que os aplicativos em execução na AWS usem perfis de instância do IAM Amazon EC2 para gerenciar credenciais.

Os perfis ou funções da instância do Amazon EC2 do IAM permitem evitar o armazenamento de chaves sensíveis nos arquivos de configuração que devem ser protegidos.

Para aplicativos móveis, a melhor prática é usar uma combinação de federação de identidade da web com o AWS Security Token Service (AWS STS) para emitir chaves temporárias que expiram após um curto período.

O Amazon DynamoDB também oferece suporte ao controle de acesso refinado que pode restringir o acesso a itens específicos em uma tabela ou mesmo atributos específicos em um item. Por exemplo, você pode limitar um usuário a acessar apenas seus itens em uma tabela e impedir o acesso a itens associados a um usuário diferente.

O uso de condições em uma política do IAM permite restringir quais ações um usuário pode executar, em quais tabelas e para quais atributos um usuário pode ler ou gravar.

Fluxos do Amazon DynamoDB

Um requisito comum para muitos aplicativos é acompanhar as alterações recentes e executar algum tipo de processamento nos registros alterados.

O Amazon DynamoDB Streams facilita a obtenção de uma lista de modificações de itens nas últimas 24 horas. Por exemplo, pode ser necessário calcular métricas de forma contínua e atualizar um painel, ou talvez sincronizar duas tabelas ou registrar atividades e alterações em uma trilha de auditoria.

Com o Amazon DynamoDB Streams, esses tipos de aplicativos ficam mais fáceis de criar.

O Amazon DynamoDB Streams permite estender a funcionalidade do aplicativo sem modificar o aplicativo original. Ao ler o log de alterações de atividades no fluxo, você pode criar novas integrações ou dar suporte a novos requisitos de relatório que não faziam parte do design original.

Cada alteração de item é armazenada em buffer em uma sequência ou fluxo ordenado por tempo que pode ser lido por outros aplicativos. As alterações são registradas no fluxo quase em tempo real e permitem que você responda

rapidamente ou encadeie uma sequência de eventos com base em uma modificação.

Os fluxos podem ser ativados ou desativados para uma tabela do Amazon DynamoDB usando o AWS Management Console, a Command Line Interface (CLI) ou o SDK.

Um fluxo consiste em registros de fluxo. Cada registro de fluxo representa uma única modificação de dados na tabela do Amazon DynamoDB à qual o fluxo pertence. Cada registro de fluxo recebe um número de sequência, refletindo a ordem em que o registro foi publicado no fluxo.

Os registros de fluxo são organizados em grupos, também chamados de shards. Cada fragmento atua como um contêiner para vários registros de fluxo e contém informações sobre como acessar e iterar através dos registros. Os shards vivem por no máximo 24 horas e, com níveis de carga flutuantes, podem ser divididos uma ou mais vezes antes de serem fechados.

SQS, SWF e SNS

Serviço de Fila Simples da Amazon (Amazon SQS)

O Amazon SQS é um serviço de enfileiramento de mensagens rápido, confiável, escalável e totalmente gerenciado. O Amazon SQS torna mais simples e econômico desacoplar os componentes de um aplicativo em nuvem.

Você pode usar o Amazon SQS para transmitir qualquer volume de dados, em qualquer nível de taxa de transferência, sem perder mensagens ou exigir que outros serviços estejam disponíveis continuamente.

Com o Amazon SQS, você pode descarregar a carga administrativa de operar e dimensionar um cluster de mensagens altamente disponível, pagando um preço baixo apenas pelo que usar.

Usando o Amazon SQS, você pode armazenar mensagens de aplicativos em infraestrutura confiável e escalável, permitindo mover dados entre componentes distribuídos para executar tarefas diferentes, conforme necessário.

Uma fila do Amazon SQS é basicamente um buffer entre os componentes do aplicativo que recebem dados e os componentes que processam os dados no seu sistema. Se os servidores de processamento não puderem processar o trabalho com rapidez suficiente (talvez devido a um aumento no tráfego), o trabalho será colocado na fila para que os servidores de processamento possam acessá-lo quando estiverem prontos.

Isso significa que o trabalho não é perdido devido a recursos insuficientes.

O Amazon SQS garante a entrega de cada mensagem pelo menos uma vez e oferece suporte a vários leitores e gravadores interagindo com a mesma fila. Uma única fila

pode ser usada simultaneamente por muitos componentes de aplicativos distribuídos, sem a necessidade de esses componentes se coordenarem entre si para compartilhar a fila.

Embora na maioria das vezes cada mensagem seja entregue ao seu aplicativo exatamente uma vez, você deve projetar seu sistema para ser idempotente (ou seja, ele não deve ser afetado adversamente se processar a mesma mensagem mais de uma vez).

O Amazon SQS foi projetado para estar altamente disponível e entregar mensagens de maneira confiável e eficiente; no entanto, o serviço não garante a entrega de mensagens First In, First Out (FIFO).

Para muitos aplicativos distribuídos, cada mensagem pode ser independente e, se todas as mensagens forem entregues, o pedido não será importante. Se o seu sistema exigir que a ordem seja preservada, você poderá colocar as informações de seqüência em cada mensagem para poder reordenar as mensagens quando elas forem recuperadas da fila.

Filas de atraso e Limites de Tempo de Visibilidade

As filas de atraso permitem adiar a entrega de novas mensagens em uma fila por um número específico de segundos. Se você criar uma fila de atraso, qualquer mensagem que você enviar para essa fila ficará invisível para os consumidores durante o período de atraso.

Para criar uma fila de atraso, use CreateQueue e defina o atributo DelaySeconds para qualquer valor entre 0 e 900 (15 minutos).

Você também pode transformar uma fila existente em uma fila de atraso usando SetQueueAttributes para definir o atributo DelaySeconds da fila. O valor padrão para DelaySeconds é 0.

As filas de atraso são semelhantes aos tempos limite de visibilidade, pois os dois recursos tornam as mensagens indisponíveis para os consumidores por um período específico. A diferença é que uma fila de atraso oculta uma mensagem quando é

adicionada pela primeira vez à fila, enquanto um tempo limite de visibilidade oculta uma mensagem somente depois que a mensagem for recuperada da fila.

Quando uma mensagem está na fila, mas não está atrasada nem com o tempo limite de visibilidade, ela é considerada "em voo". Você pode receber até 120.000 mensagens em voo a qualquer momento. O Amazon SQS suporta o tempo limite máximo de visibilidade de até 12 horas.

Taxa de transferência separada da latência Como muitos outros serviços da AWS Cloud, o Amazon SQS é acessado através da resposta HTTP request, e uma resposta de solicitação típica do Amazon SQS leva um pouco menos de 20ms do Amazon Elastic Compute Cloud (Amazon EC2).

Isso significa que, a partir de um único encadeamento, é possível emitir, em média, mais de 50 solicitações de API (Application Programming Interface) por segundo (um pouco menos para solicitações de API em lote, mas essas funcionam mais).

A taxa de transferência é horizontal, portanto, quanto mais threads e hosts você adicionar, maior será a taxa de transferência. Usando esse modelo de escala, alguns clientes da AWS têm filas que processam milhares de mensagens a cada segundo.

Operações de fila, IDs Exclusivos e Metadados

As operações definidas para as filas do Amazon SQS são CreateQueue, ListQueues, DeleteQueue, SendMessage, SendMessageBatch, ReceiveMessage, DeleteMessage, DeleteMessageBatch, PurgeQueue, ChangeMessageVisibility, ChangeMessageVisibilityBatch, SetQueueAttributes, GetQueueAttributes, GetQueueUrl, ListDeadLetterSourceQueues, AddPermission, e RemovePermission. Somente o proprietário da conta da AWS ou uma identidade da AWS que recebeu as permissões apropriadas podem executar operações.
Suas mensagens são identificadas por meio de um ID exclusivo globalmente que o Amazon SQS retorna quando a mensagem é entregue na fila. O ID não é necessário para executar outras ações na mensagem, mas é útil para rastrear se uma mensagem específica na fila foi recebida.

Quando você recebe uma mensagem da fila, a resposta inclui um identificador de recebimento, que você deve fornecer ao excluir a mensagem.

Identificadores de fila e mensagem O Amazon SQS usa três identificadores com os quais você precisa se familiarizar: URLs da fila, IDs de mensagens e identificadores de recebimento.

Ao criar uma nova fila, você deve fornecer um nome de fila exclusivo no escopo de todas as suas filas. O Amazon SQS atribui a cada fila um identificador chamado URL da fila, que inclui o nome da fila e outros componentes que o Amazon SQS determina.

Sempre que desejar executar uma ação em uma fila, você deve fornecer o URL da fila. O Amazon SQS atribui a cada mensagem um ID exclusivo que ele retorna para você na resposta SendMessage. Esse identificador é útil para identificar mensagens, mas observe que, para excluir uma mensagem, você precisa do identificador de recebimento da mensagem em vez do ID da mensagem.

O comprimento máximo de um ID de mensagem é de 100 caracteres.

Cada vez que você recebe uma mensagem de uma fila, recebe um identificador de recibo para essa mensagem.

O identificador está associado ao ato de receber a mensagem, não à própria mensagem. Como afirmado anteriormente, para excluir a mensagem ou alterar a visibilidade da mensagem, você deve fornecer o identificador de recebimento e não o ID da mensagem.

Isso significa que você deve sempre receber uma mensagem antes de poder excluí-la (ou seja, não é possível colocar uma mensagem na fila e depois rechamar). O comprimento máximo de um identificador de recebimento é 1.024 caracteres.

Atributos da Mensagem

O Amazon SQS fornece suporte para atributos de mensagem. Os atributos da mensagem permitem fornecer itens de metadados estruturados (como registros

de data e hora, dados geoespaciais, assinaturas e identificadores) sobre a mensagem.

Os atributos da mensagem são opcionais e separados, mas enviados junto com o corpo da mensagem. O destinatário da mensagem pode usar essas informações para ajudar a decidir como lidar com a mensagem sem precisar processar o corpo da mensagem primeiro.

Cada mensagem pode ter até 10 atributos. Para especificar atributos de mensagem, você pode usar o AWS Management Console, o AWS Software Development Kits (SDKs) ou uma API de consulta.

Long Polling

Quando seu aplicativo consulta a fila do Amazon SQS por mensagens, ele chama a função ReceiveMessage. ReceiveMessage verificará a existência de uma mensagem na fila e retornará imediatamente, com ou sem uma mensagem. Se o seu código fizer chamadas periódicas para a fila, esse padrão será suficiente.

Se o seu cliente SQS é apenas um loop que verifica repetidamente novas mensagens, no entanto, esse padrão se torna problemático, pois as chamadas constantes para ReceiveMessage queimam ciclos da CPU e amarram um encadeamento.

Nessa situação, você desejará usar uma pesquisa longa. Com uma pesquisa longa, você envia um argumento WaitTimeSeconds para ReceiveMessage de até 20 segundos. Se não houver mensagem na fila, a chamada aguardará até WaitTimeSeconds para que uma mensagem apareça antes de retornar.

Se uma mensagem aparecer antes do tempo expirar, a chamada retornará a mensagem imediatamente. A pesquisa longa reduz drasticamente a quantidade de carga no seu cliente.

Filas de Mensagens não Entregues

O Amazon SQS fornece suporte para filas de mensagens não entregues. Uma fila de mensagens não entregues é uma fila que outras filas (de origem) podem

direcionar para enviar mensagens que, por algum motivo, não puderam ser processadas com êxito.

Um benefício principal do uso de uma fila de devoluções é a capacidade de marginalizar e isolar as mensagens processadas sem êxito.

Você pode analisar todas as mensagens enviadas para a fila de devoluções para tentar determinar a causa da falha. As mensagens podem ser enviadas e recebidas de uma fila de devoluções, como qualquer outra fila do Amazon SQS. Você pode criar uma fila de devoluções a partir da API do Amazon SQS e do console do Amazon SQS.

Controle de Acesso

Embora o IAM possa ser usado para controlar as interações de diferentes identidades da AWS com filas, geralmente há momentos em que você deseja expor filas para outras contas. Essas situações podem incluir:

- Você deseja conceder a outra conta da AWS um tipo específico de acesso à sua fila (por exemplo, SendMessage).
- Você deseja conceder acesso a outra conta da AWS em sua fila por um período específico.
- Você deseja conceder acesso a outra conta da AWS em sua fila somente se as solicitações vierem de suas instâncias do Amazon EC2.
- Você deseja negar o acesso de outra conta da AWS à sua fila.

Embora a coordenação estreita entre contas possa permitir esses tipos de ações por meio do uso de funções do IAM, esse nível de coordenação é frequentemente inviável.

O Amazon SQS Access Control permite atribuir políticas a filas que concedem interações específicas a outras contas sem que essa conta precise assumir funções do IAM da sua conta. Essas políticas são escritas no mesmo idioma JSON do IAM.

Durabilidade e Latência

O Amazon SQS não retorna com êxito uma chamada da API SendMessage até que a mensagem seja armazenada de forma durável no Amazon SQS. Isso torna o modelo de programação muito simples, sem dúvida sobre a segurança das mensagens, diferente da situação de um modelo de mensagens assíncronas.

Se você não precisar de um sistema de mensagens durável, poderá criar um lote assíncrono do lado do cliente sobre as bibliotecas do Amazon SQS que atrasa o enfileiramento de mensagens no Amazon SQS e transmite um conjunto de mensagens em um lote. Esteja ciente de que, com uma abordagem de lote do lado do cliente, você poderá perder mensagens quando o processo ou o host do cliente morrer por qualquer motivo.

Serviço de Fluxo de Trabalho Simples da Amazon (Amazon SWF)

O Amazon SWF facilita a criação de aplicativos que coordenam o trabalho entre componentes distribuídos. No Amazon SWF, uma tarefa representa uma unidade lógica de trabalho que é executada por um componente do seu aplicativo.

A coordenação de tarefas no aplicativo envolve o gerenciamento de dependências, agendamento e simultaneidade entre tarefas, de acordo com o fluxo lógico do aplicativo.

O Amazon SWF oferece controle total sobre a implementação e coordenação de tarefas sem se preocupar com complexidades subjacentes, como acompanhar o progresso e manter o estado.

Ao usar o Amazon SWF, você implementa trabalhadores para executar tarefas.

Esses funcionários podem executar na infraestrutura de nuvem, como o Amazon EC2, ou em suas próprias instalações. Você pode criar tarefas de longa execução que podem falhar, atingir o tempo limite ou exigir reinicializações ou tarefas que podem ser concluídas com taxa de transferência e latência variáveis.

O Amazon SWF armazena tarefas, as atribui aos trabalhadores quando estão prontos, monitora seu progresso e mantém seu estado, incluindo detalhes sobre sua conclusão.

Para coordenar tarefas, você escreve um programa que obtém o estado mais recente de cada tarefa do Amazon SWF e o utiliza para iniciar tarefas subseqüentes. O Amazon SWF mantém o estado de execução de um aplicativo de forma durável, para que ele seja resistente a falhas em componentes individuais.

Com o Amazon SWF, você pode implementar, implantar, dimensionar e modificar esses componentes do aplicativo de forma independente.

Fluxos de Trabalho

Usando o Amazon SWF, você pode implementar aplicativos assíncronos distribuídos como fluxos de trabalho. Os fluxos de trabalho coordenam e gerenciam a execução de atividades que podem ser executadas de forma assíncrona em vários dispositivos de computação e que podem apresentar processamento sequencial e paralelo.

Ao projetar um fluxo de trabalho, analise seu aplicativo para identificar suas tarefas componentes, representadas no Amazon SWF como atividades. A lógica de coordenação do fluxo de trabalho determina a ordem em que as atividades são executadas.

Domínios de Fluxo de Trabalho

Os domínios fornecem uma maneira de definir o escopo dos recursos do Amazon SWF na sua conta da AWS. Você deve especificar um domínio para todos os componentes de um fluxo de trabalho, como o tipo de fluxo de trabalho e os tipos de atividade. É possível ter mais de um fluxo de trabalho em um domínio; no entanto, os fluxos de trabalho em domínios diferentes não podem interagir entre si.

Histórico do Fluxo de Trabalho

O histórico do fluxo de trabalho é um registro detalhado, completo e consistente de todos os eventos que ocorreram desde o início da execução do fluxo de trabalho. Um evento representa uma alteração discreta no estado de execução do seu fluxo de trabalho, como atividades agendadas e concluídas, tempos limite de tarefas e sinais.

Atores

O Amazon SWF consiste em vários tipos diferentes de recursos programáticos conhecidos como atores. Os atores podem ser iniciantes no fluxo de trabalho, decisores ou trabalhadores de atividades.

Esses atores se comunicam com o Amazon SWF por meio de sua API. Você pode desenvolver atores em qualquer linguagem de programação.

Um iniciador de fluxo de trabalho é qualquer aplicativo que possa iniciar execuções de fluxo de trabalho. Por exemplo, um iniciador de fluxo de trabalho pode ser um site de comércio eletrônico em que um cliente faz um pedido.

Outro iniciador de fluxo de trabalho pode ser um aplicativo móvel, em que um cliente solicita comida para viagem ou solicita um táxi.

As atividades em um fluxo de trabalho podem ser executadas sequencialmente, em paralelo, de forma síncrona ou assíncrona. A lógica que coordena as tarefas em um fluxo de trabalho é chamada de decisão.

O decisor agenda as tarefas da atividade e fornece dados de entrada para os trabalhadores da atividade. O decisor também processa eventos que chegam enquanto o fluxo de trabalho está em andamento e fecha o fluxo de trabalho quando o objetivo foi concluído.

Um trabalhador da atividade é um processo (ou thread) de um único computador que executa as tarefas da atividade no seu fluxo de trabalho. Diferentes tipos de trabalhadores da atividade processam tarefas de diferentes tipos de atividade, e vários trabalhadores da atividade podem processar o mesmo tipo de tarefa.

Quando um trabalhador de atividade está pronto para processar uma nova tarefa de atividade, ele pesquisa o Amazon SWF em busca de tarefas apropriadas para esse trabalhador de atividade. Após receber uma tarefa, o responsável pela atividade processa a tarefa até a conclusão e, em seguida, retorna o status e o resultado ao Amazon SWF. O trabalhador da atividade pesquisa uma nova tarefa.

Tarefas

O Amazon SWF fornece atribuições de trabalho aos trabalhadores e decisores, dados como um dos três tipos de tarefas: tarefas de atividade, tarefas do AWS Lambda e tarefas de decisão.

Uma tarefa de atividade instrui um trabalhador de atividade a desempenhar sua função, como verificar inventário ou cobrar um cartão de crédito. A tarefa de atividade contém todas as informações que o responsável pela atividade precisa para desempenhar sua função.

Uma tarefa do AWS Lambda é semelhante a uma tarefa de atividade, mas executa uma função do AWS Lambda em vez de uma atividade tradicional do Amazon SWF. Para obter mais informações sobre como definir uma tarefa do AWS Lambda, consulte a documentação da AWS sobre as tarefas do AWS Lambda.

Uma tarefa de decisão informa ao decisor que o estado da execução do fluxo de trabalho foi alterado para que o decisor possa determinar a próxima atividade que precisa ser executada. A tarefa de decisão contém o histórico atual do fluxo de trabalho.

O Amazon SWF agenda uma tarefa de decisão quando o fluxo de trabalho é iniciado e sempre que o estado do fluxo de trabalho muda, como quando uma tarefa de atividade é concluída.

Cada tarefa de decisão contém uma exibição paginada de todo o histórico de execução do fluxo de trabalho. O decisor analisa o histórico de execução do fluxo de trabalho e responde ao Amazon SWF com um conjunto de decisões que especificam o que deve ocorrer em seguida na execução do fluxo de trabalho.

Basicamente, todas as tarefas de decisão dão ao decisor a oportunidade de avaliar o fluxo de trabalho e fornecer orientações ao Amazon SWF.

Listas de Tarefas

As listas de tarefas fornecem uma maneira de organizar as várias tarefas associadas a um fluxo de trabalho. Você pode pensar nas listas de tarefas como semelhantes às filas dinâmicas.

Quando uma tarefa é agendada no Amazon SWF, você pode especificar uma fila (lista de tarefas) para inseri-la. Da mesma forma, ao pesquisar no Amazon SWF por uma tarefa, você determina de qual fila (lista de tarefas) obter a tarefa.

As listas de tarefas fornecem um mecanismo flexível para rotear tarefas para os trabalhadores, conforme seu caso de uso. As listas de tarefas são dinâmicas, pois você não precisa registrar uma lista de tarefas ou criá-la explicitamente por meio de uma ação. Basta agendar uma tarefa para criar a lista de tarefas, caso ela ainda não exista.

Long Polling

Pessoas que decidem e desenvolvem atividades se comunicam com o Amazon SWF usando pesquisas longas. O responsável pela atividade ou responsável pela atividade inicia periodicamente a comunicação com o Amazon SWF, notificando o Amazon SWF de sua disponibilidade para aceitar uma tarefa e, em seguida, especifica uma lista de tarefas para obter as tarefas.

A pesquisa longa funciona bem para o processamento de tarefas de alto volume. Pessoas que decidem e desenvolvem atividades podem gerenciar sua própria capacidade.

Identificadores de Objeto

Os objetos Amazon SWF são identificados exclusivamente por tipo de fluxo de trabalho, tipo de atividade, tarefas de decisão e atividade e execução de fluxo de trabalho:

Um tipo de fluxo de trabalho registrado é identificado por seu domínio, nome e versão. Os tipos de fluxo de trabalho são especificados na chamada para RegisterWorkflowType.

Um tipo de atividade registrado é identificado por seu domínio, nome e versão. Os tipos de atividades são especificados na chamada para RegisterActivityType.

Cada tarefa de decisão e atividade é identificada por um token de tarefa exclusivo. O token de tarefa é gerado pelo Amazon SWF e é retornado com outras informações sobre a tarefa na resposta de PollForDecisionTask ou PollForActivityTask.

Embora o token seja mais comumente usado pelo processo que recebeu a tarefa, esse processo pode passar o token para outro processo, que pode então relatar a conclusão ou falha da tarefa.

Uma única execução de um fluxo de trabalho é identificada pelo domínio, ID do fluxo de trabalho e ID da execução.

Os dois primeiros são parâmetros que são passados para StartWorkflowExecution. O ID da execução é retornado pelo StartWorkflowExecution.

Encerramento da Execução do Fluxo de Trabalho

Depois de iniciar uma execução do fluxo de trabalho, ela é aberta. Uma execução de fluxo de trabalho aberto pode ser fechada como concluída, cancelada, com falha ou com tempo limite esgotado. Também pode ser continuado como uma nova execução ou pode ser finalizado. O decisor, a pessoa que administra o fluxo de trabalho ou o Amazon SWF pode fechar uma execução do fluxo de trabalho.

Ciclo de Vida de uma Execução de Fluxo de Trabalho

Desde o início da execução de um fluxo de trabalho até sua conclusão, o Amazon SWF interage com os atores, atribuindo-lhes tarefas apropriadas: tarefas de atividade ou de decisão.

Serviço de Notificação Simples da Amazon (Amazon SNS)

O Amazon SNS é um serviço da Web para mensagens móveis e corporativas que permite configurar, operar e enviar notificações. Ele foi desenvolvido para facilitar a computação em escala da web para os desenvolvedores.

O Amazon SNS segue o paradigma de mensagens de publicação-assinatura (pub-sub), com notificações sendo entregues aos clientes usando um mecanismo de envio que elimina a necessidade de verificar periodicamente (ou pesquisa) por novas informações e atualizações.

Por exemplo, você pode enviar notificações para dispositivos Apple, Android, Fire OS e Windows. Na China, você pode enviar mensagens para dispositivos Android com o Baidu Cloud Push.

Você pode usar o Amazon SNS para enviar mensagens SMS (Short Message Service) para usuários de dispositivos móveis nos Estados Unidos ou para destinatários de e-mail em todo o mundo.

O Amazon SNS consiste em dois tipos de clientes: editores e assinantes (às vezes conhecidos como produtores e consumidores). Os editores se comunicam de maneira assíncrona com os assinantes enviando uma mensagem para um tópico.

Um tópico é simplesmente um ponto de acesso lógico / canal de comunicação que contém uma lista de assinantes e os métodos usados para se comunicar com eles.

Quando você envia uma mensagem para um tópico, ela é encaminhada automaticamente para cada assinante desse tópico usando o método de comunicação configurado para esse assinante.

Ao usar o Amazon SNS, você (como proprietário) cria um tópico e controla o acesso a ele, definindo políticas que determinam quais editores e assinantes podem se comunicar com o tópico e por quais tecnologias.

Os editores enviam mensagens para tópicos que eles criaram ou que eles têm permissão para publicar em. Em vez de incluir um endereço de destino específico em cada mensagem, um editor envia uma mensagem para o tópico, e o Amazon SNS entrega a mensagem para cada assinante desse tópico.

Cada tópico possui um nome exclusivo que identifica o endpoint Amazon SNS em que os editores postam mensagens e os inscritos se registram para receber notificações.

Os assinantes recebem todas as mensagens publicadas nos tópicos em que assinam e todos os assinantes de um tópico recebem as mesmas mensagens.

Cenários Comuns do Amazon SNS

O Amazon SNS pode atender a uma ampla variedade de necessidades, incluindo aplicativos de monitoramento, sistemas de fluxo de trabalho, atualizações de informações sensíveis ao tempo, aplicativos móveis e qualquer outro aplicativo que gere ou consome notificações.

Por exemplo, você pode usar o Amazon SNS para retransmitir eventos em sistemas de fluxo de trabalho entre aplicativos de computador distribuídos, mover dados entre repositórios de dados ou atualizar registros em sistemas comerciais.

Atualizações e notificações de eventos relacionadas à validação, aprovação, alterações de inventário e status da remessa são entregues imediatamente aos componentes e usuários finais relevantes do sistema.

Outro exemplo de uso do Amazon SNS é retransmitir eventos críticos para aplicativos e dispositivos móveis.

Como o Amazon SNS é altamente confiável e escalável, fornece vantagens significativas aos desenvolvedores que constroem aplicativos que dependem de eventos em tempo real.

Para ajudar a ilustrar, as seções a seguir descrevem alguns cenários comuns do Amazon SNS, incluindo cenários de fanout, alertas de aplicativos e sistema, email push e mensagens de texto e notificações móveis.

Fanout

Um cenário de fanout é quando uma mensagem do Amazon SNS é enviada para um tópico e, em seguida, replicada e enviada para várias filas do Amazon SQS, endpoint HTTP ou endereços de email.

Isso permite o processamento assíncrono paralelo. Por exemplo, você pode desenvolver um aplicativo que envie uma mensagem do Amazon SNS para um tópico sempre que um pedido for feito para um produto.

Em seguida, as filas do Amazon SQS inscritas nesse tópico receberão notificações idênticas para o novo pedido. Uma instância do Amazon EC2 conectada a uma das filas lida com o processamento ou o atendimento do pedido, enquanto uma instância do Amazon EC2 conectada a uma fila paralela envia dados do pedido para um aplicativo / serviço de armazém de dados para análise.

Outra maneira de usar o fanout é replicar os dados enviados ao seu ambiente de produção e integrá-los ao seu ambiente de desenvolvimento.

Expandindo o exemplo anterior, é possível inscrever mais uma fila no mesmo tópico para novos pedidos recebidos. Em seguida, anexando essa nova fila ao seu ambiente de desenvolvimento, você pode continuar aprimorando e testando seu aplicativo usando os dados recebidos do seu ambiente de produção.

Alertas de Aplicativos e Sistemas

Os alertas de aplicativos e do sistema são notificações por SMS e / ou email que são acionadas por limites predefinidos. Por exemplo, como muitos serviços da AWS Cloud usam o Amazon SNS, você pode receber uma notificação imediata quando ocorrer um evento, como uma alteração específica no seu grupo de Auto Scaling na AWS.

Enviar email e Mensagens de Texto

O envio por e-mail e as mensagens de texto são duas maneiras de transmitir mensagens a indivíduos ou grupos via e-mail e / ou SMS. Por exemplo, você pode usar o Amazon SNS para enviar manchetes de notícias direcionadas aos assinantes por email ou SMS.

Ao receber o email ou o texto SMS, os leitores interessados podem optar por aprender mais visitando um site ou iniciando um aplicativo.

Notificações Push Móveis

As notificações push móveis permitem enviar mensagens diretamente para aplicativos móveis. Por exemplo, você pode usar o Amazon SNS para enviar notificações para um aplicativo, indicando que uma atualização está disponível. A mensagem de notificação pode incluir um link para baixar e instalar a atualização.

DNS e Route 53

Sistema de Nomes de Domínio (DNS)

O sistema de nomes de domínio (DNS) às vezes é um conceito difícil de entender porque é usado de maneira ubíqua para fazer a Internet funcionar.

Antes de entrarmos em detalhes, vamos começar com uma analogia simples. O endereço IP do seu site é como o seu número de telefone - pode mudar se você mudar para uma nova área (pelo menos sua linha terrestre pode mudar).

DNS é como a lista telefônica. Se alguém quiser ligar para você em sua nova casa ou local, poderá procurar por nome na lista telefônica. Se a agenda telefônica deles não tiver sido atualizada desde que você se mudou, eles podem ligar para sua casa antiga.

Quando um visitante deseja acessar seu site, seu computador pega o nome de domínio digitado (www.amazon .com, por exemplo) e consulta o endereço IP desse domínio usando DNS.

Mais especificamente, o DNS é um serviço distribuído globalmente, que é fundamental para a maneira como as pessoas usam a Internet. O DNS usa uma estrutura hierárquica de nomes e diferentes níveis na hierarquia são separados por um ponto (.).

Considere os nomes de domínio www.amazon .com e aws.amazon.com. Nos dois exemplos, com é o domínio de nível superior (TLD) e a amazon é o domínio de segundo nível (SLD).

Pode haver qualquer número de níveis mais baixos (por exemplo, www e aws) abaixo do SLD.

Os computadores usam a hierarquia DNS para converter nomes legíveis por humanos (por exemplo, www.amazon.com) nos endereços IP (por exemplo, 192.0.2.1) que os computadores usam para se conectar.

Sempre que você usa um nome de domínio, um serviço DNS deve converter o nome no endereço IP correspondente. Em resumo, se você usou a Internet, usou o DNS.

O Amazon Route 53 é um sistema DNS autoritativo. Um sistema DNS autoritativo fornece um mecanismo de atualização que os desenvolvedores usam para gerenciar seus nomes DNS públicos.

Em seguida, ele responde às consultas DNS, convertendo nomes de domínio em endereços IP, para que os computadores possam se comunicar.

Conceitos de sistema de nome de domínio (DNS)

Esta seção define os termos do DNS, descreve como o DNS funciona e explica os tipos de registro mais usados.

Domínios de nível superior (TLDs)

Um domínio de nível superior (TLD) é a parte mais geral do domínio. O TLD é a parte mais distante à direita (separada por um ponto).

Os TLDs comuns são .com, .net, .org, .gov, .edu e .io.

Os TLDs estão no topo da hierarquia em termos de nomes de domínio.

Algumas partes recebem controle de gerenciamento sobre TLDs pela Internet Corporation para nomes e números atribuídos (ICANN).

Essas partes podem distribuir nomes de domínio no TLD, geralmente por meio de um registrador de domínio. Esses domínios são registrados no Network Information Center (InterNIC), um serviço da ICANN, que reforça a exclusividade dos nomes de domínio na Internet.

Cada nome de domínio é registrado em um banco de dados central, conhecido como banco de dados WhoIS.

Nomes de Domínio

Um nome de domínio é o nome amigável ao ser humano que estamos acostumados a associar a um recurso da Internet. Por exemplo, amazon.com é um nome de domínio. Algumas pessoas dirão que a parte amazon é o domínio, mas geralmente podemos nos referir à forma combinada como o nome de domínio.

O URL aws.amazon.com está associado aos servidores de propriedade da AWS. O DNS permite que os usuários acessem os servidores da AWS quando digitam aws.amazon.com em seus navegadores.

Endereços IP

Um endereço IP é um local endereçável da rede. Cada endereço IP deve ser exclusivo em sua rede. Para sites públicos, essa rede é a Internet inteira.

Os endereços IPv4, a forma mais comum de endereços, consistem em quatro conjuntos de números separados por um ponto, com cada conjunto com até três dígitos. Por exemplo, 111.222.111.222 pode ser um endereço IP IPv4 válido.

Com o DNS, mapeamos um nome para esse endereço, para que você não precise se lembrar de um conjunto complicado de números para cada local que deseja visitar em uma rede.

Devido ao tremendo crescimento da Internet e ao número de dispositivos conectados a ela, o intervalo de endereços IPv4 foi rapidamente esgotado.

O IPv6 foi criado para resolver esse problema de esgotamento e possui um espaço de endereço de 128 bits, o que permite 340.282.366.920.938.463, 463.374.607.431.768.211.456, ou 340 undecilhões, endereços exclusivos.

Para seres humanos, esse número é difícil de imaginar, então considere o seguinte: se cada endereço IPv4 fosse um grão de areia, você teria endereços suficientes para encher aproximadamente um caminhão basculante com areia.

Se cada endereço IPv6 fosse um grão de areia, você teria areia suficiente para igualar o tamanho aproximado do sol.

Hoje, a maioria dos dispositivos e redes ainda se comunica usando o IPv4, mas a migração para o IPv6 continua gradualmente ao longo do tempo.

Hosts

Dentro de um domínio, o proprietário do domínio pode definir hosts individuais, que se referem a computadores ou serviços separados acessíveis por um domínio.

Por exemplo, a maioria dos proprietários de domínio torna seus servidores da Web acessíveis por meio do domínio base (exemplo.com.br) e também pela definição de host www (como em www.examplo.com.br).

Você pode ter outras definições de host no domínio geral, como acesso à API (Application Program Interface) por meio de um host da API (api.example.com) ou acesso ao File Transfer Protocol (FTP) com uma definição de host de FTP ou arquivos (ftp. exemplo.com ou arquivos.exemplo.com).

Os nomes de host podem ser arbitrários se forem exclusivos para o domínio.

Subdomínios

O DNS funciona de maneira hierárquica e permite que um domínio grande seja particionado ou estendido em vários subdomínios.

Os TLDs podem ter muitos subdomínios. Por exemplo, zappos.com e audible.com são ambos subdomínios do TLD .com (embora normalmente sejam chamados apenas de domínios).

O zappos ou parte audível pode ser chamado de SLD.

Da mesma forma, cada SLD pode ter subdomínios localizados abaixo dele. Por exemplo, o URL do departamento de história de uma escola pode ser www.history.school.edu.

A diferença entre um nome de host e um subdomínio é que um host define um computador ou recurso, enquanto um subdomínio estende o domínio pai.

Subdomínios são um método de subdividir o próprio domínio.

Seja falando sobre subdomínios ou hosts, você pode ver que as partes mais à esquerda de um domínio são as mais específicas. É assim que o DNS funciona: do mais para o menos específico, enquanto você lê da esquerda para a direita.

Nome de domínio totalmente qualificado (FQDN)

Os locais de domínio em um DNS podem ser relativos um ao outro e, como tal, podem ser um pouco ambíguos. Um nome de domínio totalmente qualificado (FQDN), também conhecido como nome de domínio absoluto, especifica a localização de um domínio em relação à raiz absoluta do DNS.

Isso significa que o FQDN especifica cada domínio pai, incluindo o TLD. Um FQDN adequado termina com um ponto, indicando a raiz da hierarquia DNS. Por exemplo, o email .amazon.com é um FQDN. Às vezes, o software que solicita um FQDN não exige o ponto final, mas é necessário estar em conformidade com os padrões da ICANN.

Servidores de Nomes

Um servidor de nomes é um computador designado para converter nomes de domínio em endereços IP. Esses servidores fazem a maior parte do trabalho no DNS.

Como o número total de traduções de domínio é muito alto para qualquer servidor, cada servidor pode redirecionar solicitações para outros servidores de nomes ou

delegar responsabilidade pelo subconjunto de subdomínios pelos quais são responsáveis.

Os servidores de nomes podem ter autoridade, o que significa que eles respondem a consultas sobre domínios sob seu controle. Caso contrário, eles podem apontar para outros servidores ou servir cópias em cache dos dados de outros servidores de nomes.

Arquivos de zona

Um arquivo de zona é um arquivo de texto simples que contém os mapeamentos entre nomes de domínio e endereços IP. É assim que um servidor DNS finalmente identifica qual endereço IP deve ser contatado quando um usuário solicita um determinado nome de domínio.

Os arquivos de zona residem em servidores de nomes e geralmente definem os recursos disponíveis em um domínio específico ou o local onde é possível obter essas informações.

Registradores de nomes de domínio de nível superior (TLD)

Como todos os nomes em um determinado domínio devem ser exclusivos, é necessário que haja uma maneira de organizá-los para que os nomes de domínio não sejam duplicados.

É aqui que entram os registradores de nomes de domínio. Um registrador de nomes de domínio é uma organização ou entidade comercial que gerencia a reserva de nomes de domínio da Internet.

Um registrador de nome de domínio deve ser credenciado por um registro genérico de TLD (gTLD) e / ou um registro de TLD com código de país (ccTLD). O gerenciamento é feito de acordo com as diretrizes dos registros de nomes de domínio designados.

Etapas envolvidas na resolução do sistema de nomes de domínio (DNS)

Quando você digita um nome de domínio no seu navegador, o computador primeiro verifica seu arquivo host para ver se ele possui esse nome de domínio armazenado localmente. Caso contrário, ele verificará seu cache DNS para ver se você já visitou o site anteriormente. Se ainda não tiver um registro desse nome de domínio, ele entrará em contato com um servidor DNS para resolver o nome de domínio.

O DNS é, em sua essência, um sistema hierárquico. No topo deste sistema estão os servidores raiz. A ICANN delega o controle desses servidores para várias organizações.

No momento da redação deste artigo, existem 13 servidores raiz em operação. Servidores raiz processam solicitações de informações sobre TLDs. Quando chega uma solicitação para um domínio que um servidor de nomes de nível inferior não pode resolver, é feita uma consulta ao servidor raiz do domínio.

Para lidar com o incrível volume de resoluções que acontecem todos os dias, esses servidores raiz são espelhados e replicados. Quando solicitações são feitas para um determinado servidor raiz, a solicitação será roteada para o espelho mais próximo desse servidor raiz.

Na verdade, os servidores raiz não sabem onde o domínio está hospedado.

No entanto, eles poderão direcionar o solicitante para os servidores de nomes que lidam com o TLD solicitado especificamente.

Por exemplo, se uma solicitação para www.wikipedia.org for feita ao servidor raiz, ele verificará seus arquivos de zona em busca de uma listagem que corresponda a esse nome de domínio, mas não encontrará uma em seus registros.

Em vez disso, ele encontrará um registro para o TLD .org e fornecerá à entidade solicitante o endereço do servidor de nomes responsável pelos endereços .org.

Servidores de domínio de nível superior (TLD)

Depois que um servidor raiz retorna o endereço IP do servidor apropriado responsável pelo TLD de uma solicitação, o solicitante envia uma nova solicitação para esse endereço.

Para continuar o exemplo da seção anterior, a entidade solicitante enviaria uma solicitação ao servidor de nomes responsável por conhecer os domínios .org para verificar se ele pode localizar www.wikipedia.org.

Mais uma vez, quando o servidor de nomes pesquisar seus arquivos de zona em busca de uma listagem www.wikipedia.org, ele não encontrará um em seus registros. No entanto, ele encontrará uma lista para o endereço IP do servidor de nomes responsável pelo wikipedia.org. Isso está ficando muito mais próximo do endereço IP correto.

Servidores de nomes em nível de domínio

Nesse momento, o solicitante possui o endereço IP do servidor de nomes responsável por saber o endereço IP real do recurso.

Ele envia uma nova solicitação ao servidor de nomes perguntando, mais uma vez, se ele pode resolver o site www.wikipedia.org.

O servidor de nomes verifica seus arquivos de zona e encontra um arquivo de zona associado ao wikipedia.org.

Dentro deste arquivo, há um registro que contém o endereço IP do host .www. O servidor de nomes retorna o endereço final ao solicitante.

Resolvendo servidores de nomes

No cenário anterior, nos referimos a um solicitante. Qual é o solicitante nessa situação?

Em quase todos os casos, o solicitante será chamado de servidor de nomes de resolução, que é um servidor configurado para fazer perguntas a outros servidores.

Sua função principal é atuar como intermediário para um usuário, armazenando em cache os resultados de consultas anteriores para melhorar a velocidade e fornecendo os endereços dos servidores raiz apropriados para resolver novas solicitações.

Um usuário geralmente terá alguns servidores de nomes resolvidos configurados no sistema do computador.

Os servidores de nomes de resolução geralmente são fornecidos por um provedor de serviços de Internet (ISP) ou outra organização.

Existem vários servidores DNS de resolução pública que você pode consultar.

Eles podem ser configurados no seu computador automática ou manualmente.

Quando você digita um URL na barra de endereços do seu navegador, o computador primeiro verifica se ele encontra a localização do recurso localmente. Ele verifica o arquivo host no computador e qualquer cache armazenado localmente.

Em seguida, ele envia a solicitação ao servidor de nomes que está resolvendo e aguarda o recebimento do endereço IP do recurso.

O servidor de nomes de resolução verifica o cache para obter a resposta. Se não encontrar, seguirá as etapas descritas nas seções anteriores.

A resolução de servidores de nomes comprime o processo de solicitação para o usuário final. Os clientes simplesmente precisam saber para perguntar aos servidores de nomes de resolução onde um recurso está localizado, e os servidores de nomes de resolução farão o trabalho para investigar e retornar a resposta final.

Mais sobre arquivos de zona

Os arquivos de zona são a maneira como os servidores de nomes armazenam informações sobre os domínios que eles conhecem.

Quanto mais arquivos de zona tiver um servidor de nomes, mais solicitações ele poderá responder com autoridade.

A maioria das solicitações para o servidor de nomes médio, no entanto, é para domínios que não estão no arquivo de zona local.

Se o servidor estiver configurado para lidar com consultas recursivas, como um servidor de nomes resolvido, ele encontrará a resposta e a retornará.

Caso contrário, ele informará a entidade solicitante para onde procurar em seguida. Um arquivo de zona descreve uma zona DNS, que é um subconjunto de todo o DNS.

Os arquivos de zona geralmente são usados para configurar um único domínio e podem conter vários registros que definem onde estão os recursos para o domínio em questão.

A diretiva $ ORIGIN do arquivo de zona é um parâmetro igual ao nível mais alto de autoridade da região por padrão.

Se um arquivo de zona for usado para configurar o domínio example.com, o $ ORIGIN será definido como example.com.

Esse parâmetro é configurado na parte superior do arquivo de zona ou definido no arquivo de configuração do servidor DNS que faz referência ao arquivo de zona. De qualquer forma, esse parâmetro define quais registros oficiais a região governa.

Da mesma forma, a diretiva $ TTL configura o valor padrão de tempo de vida (TTL) para registros de recursos na zona.

Este valor define o período em que os resultados consultados anteriormente estão disponíveis para um servidor de nomes em cache antes de expirarem.

Tipos de registro

Cada arquivo de zona contém registros. Na sua forma mais simples, um registro é um único mapeamento entre um recurso e um nome.

Eles podem mapear um nome de domínio para um endereço IP ou definir recursos para o domínio, como servidores de nomes ou servidores de email. Esta seção descreve cada tipo de registro em detalhes.

Registro de início de autoridade (SOA)

Um registro de início de autoridade (SOA) é obrigatório em todos os arquivos de zona e identifica as informações DNS básicas sobre o domínio.

Cada zona contém um único registro SOA.

O registro SOA armazena informações sobre o seguinte:

- O nome do servidor DNS para essa zona
- O administrador da zona
- A versão atual do arquivo de dados
- O número de segundos que um servidor de nome secundário deve esperar antes de verificar se há atualizações
- O número de segundos que um servidor de nome secundário deve esperar antes de tentar novamente uma transferência de zona com falha
- O número máximo de segundos que um servidor de nome secundário pode usar dados antes de ser atualizado ou expirar
- O valor TTL padrão (em segundos) para registros de recursos na zona

A e AAAA

Os dois tipos de registros de endereço mapeiam um host para um endereço IP. O registro A é usado para mapear um host para um endereço IP IPv4, enquanto os registros AAAA são usados para mapear um host para um endereço IPv6.

Nome canônico (CNAME)

Um registro de nome canônico (CNAME) é um tipo de registro de recurso no DNS que define um alias para o CNAME do seu servidor (o nome de domínio definido em um registro A ou AAAA).

Mail Exchange (MX)

Os registros MX (Mail Exchange) são usados para definir os servidores de email usados para um domínio e garantir que as mensagens de email sejam roteadas corretamente. O registro MX deve apontar para um host definido por um registro A ou AAAA e não um definido por um CNAME.

Servidor de nomes (NS)

Os registros do servidor de nomes (NS) são usados pelos servidores de TLD para direcionar o tráfego para o servidor DNS que contém os registros DNS autoritativos.

Ponteiro (PTR)

Um registro de ponteiro (PTR) é essencialmente o inverso de um registro. Os registros PTR mapeiam um endereço IP para um nome DNS e são usados principalmente para verificar se o nome do servidor está associado ao endereço IP de onde a conexão foi iniciada.

Estrutura de Política do Remetente (SPF)

Os registros do Sender Policy Framework (SPF) são usados pelos servidores de email para combater o spam. Um registro SPF informa ao servidor de email quais endereços IP estão autorizados a enviar um email do seu nome de domínio.

Por exemplo, se você quiser garantir que apenas o seu servidor de email envie emails do domínio da sua empresa, como exemplo.com, crie um registro SPF com o endereço IP do seu servidor de email.

Dessa forma, um email enviado do seu domínio, como marketing@example.com, precisaria ter um endereço IP de origem do servidor de correio da empresa para ser aceito.

Isso impede que as pessoas falsifiquem emails do seu nome de domínio.

Texto (TXT)

Os registros de texto (TXT) são usados para armazenar informações de texto. Esse registro fornece a capacidade de associar algum texto arbitrário e não formatado a um host ou outro nome, como informações legíveis por humanos sobre um servidor, rede, data center e outras informações contábeis.

Serviço (SRV)

Um registro de serviço (SRV) é uma especificação de dados no DNS que define o local (o nome do host e o número da porta) dos servidores para serviços especificados.

A idéia por trás do SRV é que, dado um nome de domínio (por exemplo, exemplo.com) e um nome de serviço (por exemplo, web [HTTP], que é executado em um protocolo [TCP]), uma consulta DNS pode ser emitida para encontrar o nome do host que fornece esse serviço para o domínio, que pode ou não estar dentro do domínio.

Visão geral do Amazon Route 53

Agora que você tem um entendimento básico do DNS e dos diferentes tipos de registros DNS, pode explorar o Amazon Route 53.

O Amazon Route 53 é um serviço da Web DNS em nuvem altamente disponível e escalonável, projetado para oferecer aos desenvolvedores e empresas uma

maneira extremamente confiável e econômica de direcionar os usuários finais para aplicativos da Internet.

O Amazon Route 53 executa três funções principais:

Registro de domínio - o Amazon Route 53 permite registrar nomes de domínio, como exemplo.com.

Serviço DNS - o Amazon Route 53 traduz nomes de domínio amigáveis como www.example.com em endereços IP como 192.0.2.1. O Amazon Route 53 responde a consultas DNS usando uma rede global de servidores DNS autorizados, o que reduz a latência. Para estar em conformidade com os padrões DNS, as respostas enviadas pelo UDP (User Datagram Protocol) estão limitadas a 512 bytes. Respostas com mais de 512 bytes são truncadas e o resolvedor deve reemitir a solicitação pelo TCP.

Verificação de integridade

Amazon Route 53 envia solicitações automatizadas pela Internet ao seu aplicativo para verificar se é acessível, disponível e funcional.

Você pode usar qualquer combinação dessas funções. Por exemplo, você pode usar o Amazon Route 53 como seu registrador e seu serviço DNS, ou você pode usar o Amazon Route 53 como o serviço DNS para um domínio que você registrou em outro registrador de domínio.

Registro do Domínio

Se você deseja criar um site, primeiro você precisa registrar o nome de domínio. Se você já registrou um nome de domínio em outro registrador, tem a opção de transferir o registro de domínio para o Amazon Route 53.

Não é necessário usar o Amazon Route 53 como seu serviço DNS ou configurar a verificação de integridade de seus recursos. O Amazon Route 53 oferece suporte ao registro de domínio para uma ampla variedade de TLDs genéricos (por exemplo,

.com e .org) e TLDs geográficos (por exemplo, .be e .us). Para obter uma lista completa dos TLDs suportados, consulte o Amazon Route 53 Developer Guide em https://docs.aws.amazon.com/Route53/latest/DeveloperGuide/.

Serviço DNS (Domain Name System)

Como afirmado anteriormente, o Amazon Route 53 é um serviço DNS autorizado que roteia o tráfego da Internet para o seu site, convertendo nomes de domínio amigáveis em endereços IP.

Quando alguém digita seu nome de domínio em um navegador ou envia um e-mail, uma solicitação de DNS é encaminhada para o servidor DNS do Amazon Route 53 mais próximo em uma rede global de servidores DNS autorizados.

O Amazon Route 53 responde com o endereço IP que você especificou. Se você registrar um novo nome de domínio no Amazon Route 53, o Amazon Route 53 será automaticamente configurado como o serviço DNS do domínio e uma zona hospedada será criada para o seu domínio.

Você adiciona conjuntos de registros de recursos à zona hospedada, que define como você deseja que o Amazon Route 53 responda às consultas DNS do seu domínio (por exemplo, com o endereço IP de um servidor da Web, o endereço IP do local de borda mais próximo do Amazon CloudFront, ou o endereço IP de um balanceador de carga do Elastic Load Balancing).

Se você registrou seu domínio em outro registrador de domínio, esse registrador provavelmente está fornecendo o serviço DNS para o seu domínio. Você pode transferir o serviço DNS para o Amazon Route 53, com ou sem a transferência do registro para o domínio.

Se você estiver usando o Amazon CloudFront, o Amazon Simple Storage Service (Amazon S3) ou o Elastic Load Balancing, poderá configurar o Amazon Route 53 para rotear o tráfego da Internet para esses recursos.

Zonas hospedadas

Uma zona hospedada é uma coleção de conjuntos de registros de recursos hospedados pelo Amazon Route 53. Como um arquivo de zona DNS tradicional, uma zona hospedada representa conjuntos de registros de recursos que são gerenciados juntos sob um único nome de domínio.

Cada zona hospedada possui seus próprios metadados e informações de configuração.

Existem dois tipos de zonas hospedadas: privada e pública. Uma zona hospedada privada é um contêiner que contém informações sobre como você deseja rotear o tráfego para um domínio e seus subdomínios em um ou mais Amazon Virtual Private Clouds (Amazon VPCs).

Uma zona hospedada pública é um contêiner que contém informações sobre como você deseja rotear o tráfego na Internet para um domínio (por exemplo, example.com) e seus subdomínios (por exemplo, apex.example.com e acme.example.com) .

Os conjuntos de registros de recursos contidos em uma zona hospedada devem compartilhar o mesmo sufixo. Por exemplo, a zona hospedada example.com pode conter conjuntos de registros de recursos para os subdomínios www.example.com e www.aws.example.com, mas não pode conter conjuntos de registros de recursos para um subdomínio www.example.ca.

Você pode usar o Amazon S3 para hospedar seu site estático na zona hospedada (por exemplo, domain.com) e redirecionar todas as solicitações para um subdomínio (por exemplo, www.domain.com).

Em seguida, no Amazon Route 53, você pode criar um registro de recurso alternativo que envie solicitações para o domínio raiz ao bucket do Amazon S3.

Use um registro de alias, não um CNAME, para sua zona hospedada. CNAMEs não são permitidos para zonas hospedadas no Amazon Route 53.

Não use registros A para subdomínios (por exemplo, www.domain.com), pois eles se referem a endereços IP codificados. Em vez disso, use os registros de alias do Amazon Route 53 ou os registros CNAME tradicionais para sempre apontar para o recurso certo, onde quer que seu site esteja hospedado, mesmo quando o servidor físico tiver alterado seu endereço IP.

Tipos de registro suportados

O Amazon Route 53 suporta os seguintes tipos de registros de recursos DNS. Ao acessar o Amazon Route 53 usando a API, você verá exemplos de como formatar o elemento Value para cada tipo de registro. Os tipos de registro suportados incluem:

- A
- AAAA
- CNAME
- MX
- NS
- PTR
- SOA
- SPF
- SRV
- TXT
- Políticas de roteamento

Ao criar um conjunto de registros de recursos, você escolhe uma política de roteamento, que determina como o Amazon Route 53 responde às consultas. As opções de política de roteamento são simples, ponderadas, baseadas em latência, failover e geolocalização.

Quando especificado, o Amazon Route 53 avalia o peso relativo de um recurso, a latência da rede do cliente em relação ao recurso ou a localização geográfica do cliente ao decidir qual recurso enviar de volta em uma resposta DNS.

As políticas de roteamento podem ser associadas a verificações de integridade, portanto, o status de integridade do recurso é considerado antes mesmo de se tornar candidato em uma árvore de decisão condicional. Uma descrição de possíveis políticas de roteamento e mais sobre verificação de integridade é abordada nesta seção.

Simples

Essa é a política de roteamento padrão quando você cria um novo recurso. Use uma política de roteamento simples quando tiver um único recurso que execute uma determinada função para o seu domínio (por exemplo, um servidor da web que serve conteúdo para o site example.com).

Nesse caso, o Amazon Route 53 responde a consultas DNS com base apenas nos valores no conjunto de registros de recursos (por exemplo, o endereço IP em um registro A).

Weighted

Com o DNS ponderado, você pode associar vários recursos (como instâncias do Amazon Elastic Compute Cloud [Amazon EC2] ou balanceadores de carga do Elastic Load Balancing) a um único nome DNS.

Use a política de roteamento ponderado quando tiver vários recursos que executam a mesma função (como servidores Web que atendem ao mesmo site) e desejar que o Amazon Route 53 direcione o tráfego para esses recursos nas proporções especificadas.

Por exemplo, você pode usar isso para balanceamento de carga entre diferentes regiões da AWS ou para testar novas versões do seu site (você pode enviar 10% do tráfego para o ambiente de teste e 90% do tráfego para a versão mais antiga do seu site).

Para criar um grupo de conjuntos de registros de recursos ponderados, é necessário criar dois ou mais conjuntos de registros de recursos que tenham o mesmo nome e tipo DNS.

Em seguida, você atribui a cada registro de recurso um identificador exclusivo e um peso relativo.

Ao processar uma consulta DNS, o Amazon Route 53 procura um conjunto de registros de recursos ou um grupo de conjuntos de registros de recursos que tenham o mesmo nome e tipo de registro DNS (como um registro A). O Amazon Route 53 seleciona um registro do grupo.

Baseado em latência

O roteamento baseado em latência permite rotear seu tráfego com base na latência de rede mais baixa para o usuário final (por exemplo, usando a região da AWS que fornecerá a eles o tempo de resposta mais rápido).

Use a política de roteamento de latência quando tiver recursos que executam a mesma função em várias zonas ou regiões de disponibilidade da AWS e desejar que o Amazon Route 53 responda às consultas DNS usando os recursos que fornecem a melhor latência.

Por exemplo, suponha que você tenha balanceadores de carga do Elastic Load Balancing na região Oeste dos EUA (Oregon) e na região Ásia-Pacífico (Cingapura) e tenha criado um registro de recurso de latência definido no Amazon Route 53 para cada balanceador de carga. Um usuário em Londres digita o nome do seu domínio em um navegador e o DNS roteia a solicitação para um servidor de nomes do Amazon Route 53. O Amazon Route 53 refere-se aos dados de latência entre Londres e a região de Cingapura e entre Londres e a região de Oregon.

Se a latência for menor entre Londres e a região do Oregon, o Amazon Route 53 responderá à solicitação do usuário com o endereço IP do seu balanceador de carga no Oregon. Se a latência for menor entre Londres e a região de Cingapura, o Amazon Route 53 responderá com o endereço IP do seu balanceador de carga em Cingapura.

Failover

Use uma política de roteamento de failover para configurar o failover ativo-passivo, no qual um recurso recebe todo o tráfego quando está disponível e o outro recurso recebe todo o tráfego quando o primeiro recurso não está disponível.

Observe que você não pode criar conjuntos de registros de recursos de failover para zonas hospedadas privadas. Por exemplo, você pode desejar que seu conjunto de registros de recursos primário esteja no oeste dos EUA (norte da Califórnia) e seu (s) recurso (s) secundário (s) de recuperação de desastres (DR) no leste dos EUA (norte da Virgínia).

O Amazon Route 53 monitorará a integridade dos endpoint de recursos principais usando uma verificação de integridade.

Uma verificação de integridade informa ao Amazon Route 53 como enviar solicitações ao terminal cuja integridade você deseja verificar: qual protocolo usar (HTTP, HTTPS ou TCP), qual endereço IP e porta usar e, para verificações de integridade HTTP / HTTPS , um nome de domínio e caminho.

Depois de configurar uma verificação de integridade, a Amazon monitorará a integridade do terminal DNS selecionado. Se sua verificação de integridade falhar, as políticas de roteamento de failover serão aplicadas e o DNS efetuará failover no site de recuperação de desastres.

Geolocalização

O roteamento de geolocalização permite escolher para onde o Amazon Route 53 enviará seu tráfego com base na localização geográfica de seus usuários (o local de origem das consultas DNS). Por exemplo, convém que todas as consultas da Europa sejam roteadas para uma frota de instâncias do Amazon EC2 configuradas especificamente para seus clientes europeus, com idiomas locais e preços em euros.

Você também pode usar o roteamento de geolocalização para restringir a distribuição de conteúdo apenas aos locais em que você possui direitos de distribuição. Outro uso possível é balancear a carga entre os terminais de maneira

previsível e fácil de gerenciar, para que cada local do usuário seja consistentemente roteado para o mesmo ponto final.

Você pode especificar localizações geográficas por continente, país ou mesmo estado nos Estados Unidos. Você também pode criar conjuntos de registros de recursos separados para regiões geográficas sobrepostas e a prioridade vai para a menor região geográfica.

Por exemplo, você pode ter um registro de recurso definido para a Europa e um para o Reino Unido. Isso permite rotear algumas consultas para países selecionados (neste exemplo, Reino Unido) para um recurso e encaminhar consultas para o resto do continente (neste exemplo, Europa) para um recurso diferente.

A geolocalização funciona mapeando endereços IP para locais. No entanto, você deve ter cuidado, pois alguns endereços IP não são mapeados para localizações geográficas.

Mesmo se você criar conjuntos de registros de recursos de localização geográfica que abranjam todos os sete continentes, o Amazon Route 53 receberá algumas consultas DNS de locais que não podem ser identificados.

Nesse caso, você pode criar um conjunto de registros de recursos padrão que lide com as consultas de endereços IP que não são mapeados para qualquer local e as consultas provenientes de locais para os quais você não criou conjuntos de registros de recursos de localização geográfica.

Se você não criar um conjunto de registros de recursos padrão, o Amazon Route 53 retornará uma resposta "sem resposta" para consultas desses locais.

Você não pode criar dois conjuntos de registros de recursos de localização geográfica que especificam o mesmo local geográfico. Você também não pode criar conjuntos de registros de recursos de localização geográfica que tenham os mesmos valores para "Nome" e "Tipo" que os "Nome" e "Tipo" de conjuntos de registros de recursos que não sejam de localização geográfica.

Mais sobre verificação de saúde

As verificações de integridade do Amazon Route 53 monitoram a integridade de seus recursos, como servidores Web e servidores de email. Você pode configurar os alarmes do Amazon CloudWatch para suas verificações de saúde para receber uma notificação quando um recurso ficar indisponível.

Você também pode configurar o Amazon Route 53 para rotear o tráfego da Internet para longe de recursos indisponíveis.

As verificações de integridade e o failover de DNS são as principais ferramentas do conjunto de recursos do Amazon Route 53 que ajudam a tornar seu aplicativo altamente disponível e resiliente a falhas.

Se você implantar um aplicativo em várias zonas de disponibilidade e várias regiões da AWS, com as verificações de integridade do Amazon Route 53 anexadas a cada terminal, o Amazon Route 53 poderá enviar apenas uma lista de terminais íntegros.

As verificações de integridade podem alternar automaticamente para um endpoint íntegro, com interrupção mínima para seus clientes e sem nenhuma alteração na configuração.

Você pode usar esse cenário de recuperação automática em configurações ativo-ativo ou ativo-passivo, dependendo se seus endpoint adicionais são sempre atingidos pelo tráfego ativo ou somente após a falha de todos os endpoint principais.

Usando verificações de integridade e failovers automáticos, o Amazon Route 53 aprimora o tempo de atividade do serviço, especialmente quando comparado à abordagem tradicional de monitor-alerta-reinício de solucionar falhas.

As verificações de integridade do Amazon Route 53 não são acionadas por consultas de DNS; eles são executados periodicamente pela AWS e os resultados são publicados em todos os servidores DNS.

Dessa forma, os servidores de nomes podem estar cientes de um terminal não íntegro e rotear de maneira diferente em aproximadamente 30 segundos após um problema (após três testes falhados consecutivos), e novos resultados de DNS

serão conhecidos pelos clientes um minuto depois (supondo que seu TTL seja 60 segundos), aumentando o tempo de recuperação completo para um minuto e meio no total nesse cenário.

A sessão SDD408 da AWS re: Invent de 2014, “Mergulho profundo do Amazon Route 53: fornecendo resiliência, minimizando a latência”, introduziu um conjunto de práticas recomendadas para o Amazon Route 53. Explore essas práticas recomendadas para ajudar você a começar a usar o Amazon Route 53 como um edifício bloco para fornecer aplicativos altamente disponíveis e resilientes na AWS.

Amazon Route 53 permite resiliência

Ao reunir esses conceitos para criar um aplicativo altamente disponível e resiliente a falhas, considere estes blocos de construção:

- Em todas as regiões da AWS, um balanceador de carga do Elastic Load Balancing é configurado com balanceamento de carga entre zonas e drenagem de conexão. Isso distribui a carga uniformemente em todas as instâncias em todas as zonas de disponibilidade e garante que as solicitações em voo sejam totalmente atendidas antes uma instância do Amazon EC2 é desconectada de um balanceador de carga do Elastic Load Balancing por qualquer motivo.
- Cada balanceador de carga do Elastic Load Balancing delega solicitações para instâncias do Amazon EC2 em execução em várias zonas de disponibilidade em um grupo de dimensionamento automático. Isso protege o aplicativo contra interrupções na Zona de Disponibilidade, garante que uma quantidade mínima de instâncias esteja sempre em execução e responde a alterações na carga, dimensionando adequadamente as instâncias do Amazon EC2 de cada grupo.
- Cada balanceador de carga do Elastic Load Balancing possui verificações de integridade definidas para garantir que ele delegue solicitações apenas para instâncias íntegras.

- Cada balanceador de carga do Elastic Load Balancing também possui uma verificação de integridade do Amazon Route 53 para garantir que as solicitações sejam roteadas apenas para balanceadores de carga com instâncias saudáveis do Amazon EC2.
- O ambiente de produção do aplicativo (por exemplo, prod.domain.com) possui registros de alias do Amazon Route 53 que apontam para os balanceadores de carga do Elastic Load Balancing. O ambiente de produção também usa uma política de roteamento baseada em latência associada às verificações de integridade do Elastic Load Balancing. Isso garante que as solicitações sejam roteadas para um balanceador de carga íntegro, fornecendo latência mínima para um cliente.
- O ambiente de failover do aplicativo (por exemplo, fail.domain.com) possui um registro alternativo do Amazon Route 53 que aponta para uma distribuição do Amazon CloudFront de um bucket do Amazon S3 que hospeda uma versão estática do aplicativo.
- O subdomínio do aplicativo (por exemplo, www.domain.com) possui um registro de alias do Amazon Route 53 que aponta para prod.domain.com (como destino primário) e fail.domain .com (como destino secundário) usando uma política de roteamento de failover. Isso garante www.domain.com rotas para os balanceadores de carga de produção, se pelo menos um deles estiver saudável ou a "baleia falhada", se todos parecerem não saudáveis.
- A zona hospedada do aplicativo (por exemplo, domain.com) possui um registro de alias do Amazon Route 53 que redireciona solicitações para www.domain.com usando um bucket do Amazon S3 com o mesmo nome.
- O conteúdo do aplicativo (estático e dinâmico) pode ser exibido usando o Amazon CloudFront. Isso garante que o conteúdo seja entregue aos clientes dos pontos de presença do Amazon CloudFront espalhados por todo o mundo para fornecer latência mínima. A veiculação de conteúdo dinâmico de uma CDN (Rede de Entrega de Conteúdo), onde é armazenada em cache por curtos períodos de tempo (isto é, alguns segundos), retira a carga do aplicativo e melhora ainda mais sua latência e capacidade de resposta.

- O aplicativo é implantado em várias regiões da AWS, protegendo-o de uma interrupção regional.

Amazon ElastiCache

Armazenamento em cache na memória

Uma das características comuns de um aplicativo bem-sucedido é uma experiência do usuário rápida e responsiva. A pesquisa mostrou que os usuários ficam frustrados e deixam um site ou aplicativo quando a resposta é lenta.

Em 2007, os testes do site de varejo da Amazon.com mostraram que, a cada 100ms de aumento no tempo de carregamento, as vendas diminuíam 1%. Viagens de ida e volta para um banco de dados e seu armazenamento subjacente podem adicionar atrasos significativos e geralmente são os principais contribuintes para a latência do aplicativo.

O armazenamento em cache de dados usados com frequência é uma das otimizações de desempenho mais importantes que você pode fazer em seus aplicativos.

Comparado à recuperação de dados de um cache na memória, consultar um banco de dados é uma operação cara. Armazenando ou movendo dados frequentemente acessados na memória, os desenvolvedores de aplicativos podem melhorar significativamente o desempenho e a capacidade de resposta de aplicativos com muita leitura.

Por exemplo, o estado da sessão do aplicativo para um site grande pode ser armazenado em um mecanismo de armazenamento em cache na memória, em vez de armazenar a sessão de dados no banco de dados.

Por muitos anos, os desenvolvedores criam aplicativos que usam mecanismos de cache como o Memcached ou o Redis para armazenar dados na memória e obter um desempenho incrivelmente rápido dos aplicativos.

Memcached é um armazenamento de chave / valor na memória simples de usar que pode ser usado para armazenar tipos arbitrários de dados.

É um dos mecanismos de cache mais populares. O Redis é um armazenamento flexível da estrutura de dados na memória que pode ser usado como cache, banco de dados ou mesmo como intermediário de mensagens. O Amazon ElastiCache permite que os desenvolvedores implantem e gerenciem facilmente ambientes de cache executando o Memcached ou o Redis.

Amazon ElastiCache

O Amazon ElastiCache é um serviço da Web que simplifica a configuração e o gerenciamento de ambientes de cache distribuído na memória. Esse serviço facilita e economicamente o fornecimento de uma solução de armazenamento em cache escalável e de alto desempenho para seus aplicativos em nuvem.

Você pode usar o Amazon ElastiCache em seus aplicativos para acelerar a implantação de clusters de cache e reduzir a administração necessária para um ambiente de cache distribuído.

Com o Amazon ElastiCache, você pode escolher entre um mecanismo de cache compatível com o protocolo Memcached ou Redis e iniciar rapidamente um cluster em questão de minutos.

Como o Amazon ElastiCache é um serviço gerenciado, você pode começar a usar o serviço hoje com muito poucas ou nenhuma modificação para aplicativos existentes que usam Memcached ou Redis. Como o Amazon ElastiCache é compatível com o protocolo com esses dois mecanismos, você só precisa alterar o terminal nos arquivos de configuração.

Usando o Amazon ElastiCache, você pode implementar qualquer número de padrões de cache.

Embora certamente seja possível criar e gerenciar um cluster de cache no Amazon Elastic Compute Cloud (Amazon EC2), o Amazon ElastiCache permite descarregar o trabalho pesado de instalação, gerenciamento de patches e monitoramento na AWS, para que você possa se concentrar no seu aplicativo.

O Amazon ElastiCache também fornece vários recursos para aprimorar a confiabilidade de implantações críticas. Embora seja raro, as instâncias subjacentes do Amazon EC2 podem ficar prejudicadas.

O Amazon ElastiCache pode detectar e recuperar automaticamente da falha de um nó de cache. Com o mecanismo Redis, o Amazon ElastiCache facilita a configuração de réplicas de leitura e o failover do primário para uma réplica no caso de um problema.

Padrões de acesso a dados

A recuperação de uma chave simples de um cache na memória sempre será mais rápida que a consulta ao banco de dados mais otimizada.

Você deve avaliar o padrão de acesso dos dados antes de decidir armazená-los no cache.

Um bom exemplo de algo para armazenar em cache é a lista de produtos em um catálogo. Para um site ocupado, a lista de itens pode ser recuperada milhares de vezes por segundo. Embora faça sentido armazenar em cache os itens mais solicitados, você também pode se beneficiar do armazenamento em itens de cache que não são solicitados com freqüência.

Existem também alguns itens de dados que não devem ser armazenados em cache. Por exemplo, se você gerar uma página única a cada solicitação, provavelmente não deve armazenar em cache os resultados da página. No entanto, mesmo que a página mude sempre, faz sentido armazenar em cache os componentes da página que não são alterados.

Mecanismos de cache

O Amazon ElastiCache permite implantar rapidamente clusters de dois tipos diferentes de mecanismos de cache populares: Memcached e Redis. Em um nível

alto, Memcached e Redis podem parecer semelhantes, mas oferecem suporte a vários casos de uso diferentes e fornecem funcionalidade diferente.

Memcached Memcached fornece uma interface muito simples que permite gravar e ler objetos nos armazenamentos de dados de valores / chaves na memória.

Com o Amazon ElastiCache, você pode aumentar e diminuir elasticamente um cluster de nós do Memcached para atender às suas demandas. Você pode particionar seu cluster em shards e oferecer suporte a operações paralelas para obter uma taxa de transferência de desempenho muito alto.

O Memcached lida com objetos como blobs que podem ser recuperados usando uma chave exclusiva. O que você coloca no objeto é com você, e geralmente são os resultados serializados de uma consulta ao banco de dados. Isso pode ser simples valores de cadeia ou dados binários.

O Amazon ElastiCache oferece suporte a várias versões recentes do Memcached. Desde o início de 2016, o serviço suporta a versão 1.4.24 do Memcached e também versões anteriores à 1.4.5.

Quando uma nova versão do Memcached é lançada, o Amazon ElastiCache simplifica o processo de atualização, permitindo que você gire um novo cluster com a versão mais recente.

Redis No final de 2013, o Amazon ElastiCache adicionou suporte para implantar clusters Redis. No momento da redação deste artigo, o serviço suporta a implantação do Redis versão 2.8.24 e também várias versões mais antigas.

Além do suporte a objetos fornecido no Memcached, o Redis suporta um rico conjunto de tipos de dados gosta de strings, listas e conjuntos.

Ao contrário do Memcached, o Redis suporta a capacidade de persistir os dados na memória no disco. Isso permite criar snapshots que fazem backup de seus dados e depois recuperar ou replicar a partir dos backups.

Os clusters Redis também podem suportar até cinco réplicas de leitura para descarregar solicitações de leitura. No caso de falha do nó primário, uma réplica de

leitura pode ser promovida e se tornar o novo mestre usando grupos de replicação Multi-AZ.

O Redis também possui recursos avançados que facilitam a classificação e classificação de dados. Alguns casos de uso comuns incluem a construção de um cabeçalho para um aplicativo móvel ou o serviço de intermediário de mensagens de alta velocidade em um sistema distribuído.

Com um cluster Redis, você pode aproveitar uma abstração de mensagens de publicação e assinatura que permite dissociar os componentes de seus aplicativos. Uma arquitetura de mensagens de publicação e assinatura oferece a flexibilidade de alterar como você consome as mensagens no futuro sem afetar o componente que está produzindo as mensagens.

Nodes e Clusters

Cada implantação do Amazon ElastiCache consiste em um ou mais nós em um cluster. Existem muitos tipos diferentes de nós disponíveis para você escolher, com base no seu caso de uso e nos recursos necessários.

Um único cluster do Memcached pode conter até 20 nós. Os clusters Redis são sempre compostos de um único nó; no entanto, vários clusters podem ser agrupados em um grupo de replicação Redis.

Os tipos de nós individuais são derivados de um subconjunto das famílias de tipos de instância do Amazon EC2, como t2, m3 e r3. Os tipos de nós específicos podem mudar com o tempo, mas hoje eles variam de um tipo de nó t2.micro com 555MB de memória até uma extensão r3.8x com 237GB de memória, com muitas opções entre elas.

A família de nós do cache t2 é ideal para aplicativos de desenvolvimento e de baixo volume com rajadas ocasionais, mas alguns recursos podem não estar disponíveis.

A família m3 é uma boa combinação de computação e memória, enquanto a família r3 é otimizada para cargas de trabalho com muita memória.

Dependendo de suas necessidades, você pode optar por ter alguns nós grandes ou muitos nós menores em seu cluster ou grupo de replicação. À medida que a demanda por seu aplicativo for alterada, você também poderá adicionar ou remover nós de tempos em tempos.

Cada tipo de nó vem com uma quantidade pré-configurada de memória, com uma pequena quantidade de memória alocada para o mecanismo de armazenamento em cache e com o próprio sistema operacional.

Design para falha

Embora seja improvável, você deve planejar a possível falha de um nó de cache individual.

Para clusters Memcached, você pode diminuir o impacto da falha de um nó de cache usando um número maior de nós com uma capacidade menor, em vez de alguns nós grandes.

Caso o Amazon ElastiCache detecte a falha de um nó, ele provisionará uma substituição e a adicionará novamente ao cluster. Durante esse período, seu banco de dados sofrerá um aumento de carga, porque agora quaisquer solicitações que seriam armazenadas em cache precisarão ser lidas do banco de dados.

Para clusters Redis, o Amazon ElastiCache detectará falhas e substituirá o nó principal. Se um grupo de replicação Multi-AZ estiver ativado, uma réplica de leitura poderá ser promovida automaticamente para primária.

Descoberta Automática do Memcached

Para clusters Memcached particionados em vários nós, o Amazon ElastiCache oferece suporte à detecção automática com a biblioteca-cliente fornecida. A descoberta automática simplifica o código do aplicativo, não precisando mais do conhecimento da topologia de infraestrutura do cluster de cache na sua camada de aplicação.

Usando a descoberta automática

O cliente de detecção automática oferece aos seus aplicativos a capacidade de identificar automaticamente todos os nós em um cluster de cache e de iniciar e manter conexões com todos esses nós.

O cliente de detecção automática está disponível para plataformas .NET, Java e PHP. Escalonamento O Amazon ElastiCache permite ajustar o tamanho do seu ambiente para atender às necessidades de cargas de trabalho à medida que evoluem com o tempo.

A adição de nós de cache adicionais permite expandir facilmente horizontalmente e atingir níveis mais altos de desempenho de leitura ou gravação. Você também pode selecionar diferentes classes de nós de cache para dimensionar verticalmente.

Escala horizontal O Amazon ElastiCache também adiciona funcionalidade adicional que permite escalar horizontalmente o tamanho do seu ambiente de cache.

Essa funcionalidade difere dependendo do mecanismo de cache que você selecionou. Com o Memcached, você pode particionar seus dados e escalar horizontalmente para 20 nós ou mais.

Com a Descoberta Automática, seu aplicativo pode descobrir nós do Memcached adicionados ou removidos de um cluster.

Um cluster Redis consiste em um único nó de cache que está manipulando transações de leitura e gravação. Clusters adicionais podem ser criados e agrupados em um grupo de replicação Redis. Embora você possa ter apenas um nó manipulando comandos de gravação, pode ter até cinco réplicas de leitura manipulando solicitações somente leitura.

Escalonamento vertical O suporte para escalonamento vertical é mais limitado com o Amazon ElastiCache. Se você deseja alterar o tipo de nó de cache e dimensionar os recursos de computação verticalmente, o serviço não permite diretamente redimensionar seu cluster dessa maneira.

No entanto, você pode ativar rapidamente um novo cluster com os tipos de nó de cache desejados e começar a redirecionar o tráfego para o novo cluster. É importante entender que um novo cluster do Memcached sempre começa vazio, enquanto um cluster do Redis pode ser inicializado a partir de um backup.

Replicação e Multi-AZ

A replicação é uma técnica útil para fornecer recuperação rápida em caso de falha do nó e também para fornecer volumes muito altos de consultas de leitura além dos recursos de um único nó.

Os clusters do Amazon ElastiCache executando o Redis suportam esses dois requisitos de design. Ao contrário do Redis, os clusters de cache executando o Memcached são serviços autônomos na memória, sem quaisquer serviços de proteção de dados redundantes.

Clusters de cache executando o Redis suportam o conceito de grupos de replicação. Um grupo de replicação consiste em até seis clusters, com cinco deles designados como réplicas de leitura. Isso permite que você dimensione horizontalmente escrevendo código no seu aplicativo para descarregar leituras para um dos cinco clones

Grupos de Replicação Multi-AZ

Você também pode criar um grupo de replicação Multi-AZ que permita aumentar a disponibilidade e minimizar a perda de dados. O Multi-AZ simplifica o processo de lidar com uma falha automatizando a substituição e o failover a partir do nó primário.

Caso o nó primário falhe ou não possa ser alcançado, o Multi-AZ selecionará e promoverá uma réplica de leitura para se tornar o novo primário e um novo nó será provisionado para substituir o que falhou.

O Amazon ElastiCache atualizará a entrada DNS do novo nó primário para permitir que seu aplicativo continue processando sem nenhuma alteração na configuração e com apenas uma pequena interrupção.

Entenda que a Replicação é Assíncrona

É importante ter em mente que a replicação entre os clusters é realizada de forma assíncrona e haverá um pequeno atraso antes que os dados estejam disponíveis em todos os nós do cluster.

Restaurar e Recuperar

Os clusters do Amazon ElastiCache executando o Redis permitem que você persista seus dados da memória no disco e crie uma captura instantânea. Cada snapshot é um clone completo dos dados que podem ser usados para recuperar um ponto específico no tempo ou criar uma cópia para outros fins.

Snapshots não podem ser criados para clusters usando o mecanismo Memcached porque é um armazenamento de chave / valor puramente na memória e sempre começa vazio. O Amazon ElastiCache usa os recursos de backup nativos do Redis e gera um arquivo de backup de banco de dados Redis padrão que é armazenado no Amazon Simple Storage Service (Amazon S3).

Os snapshots exigem recursos de computação e memória para serem executados e podem potencialmente ter um impacto no desempenho de clusters usados com muita força. O Amazon ElastiCache tentará diferentes técnicas de backup, dependendo da quantidade de memória disponível atualmente.

Uma prática recomendada é configurar um grupo de replicação e executar uma captura instantânea em uma das réplicas de leitura em vez do nó principal.

Além dos snapshots iniciados manualmente, os snapshots podem ser criados automaticamente com base em uma programação.

Você também pode configurar uma janela para que a operação de captura instantânea seja concluída e especificar quantos dias de backups você deseja

armazenar. Os snapshots manuais são armazenados indefinidamente até que você os exclua.

Clusters de Redis de Backup

Use uma combinação de snapshots automáticas e manuais para atender aos seus objetivos de recuperação para o cluster Redis. Memcached é puramente na memória e não possui recursos de backup nativos.

Se o snapshot foi criado automaticamente ou manualmente, o snapshot pode ser usado para criar um novo cluster a qualquer momento. Por padrão, o novo cluster terá a mesma configuração que o cluster de origem, mas você pode substituir essas configurações.

Você também pode restaurar a partir de um arquivo RDB gerado a partir de qualquer outro cluster Redis compatível.

Controle de Acesso

O acesso ao cluster do Amazon ElastiCache é controlado principalmente pela restrição do acesso da rede de entrada ao cluster. O tráfego da rede de entrada é restrito pelo uso de grupos de segurança.

Cada grupo de segurança define uma ou mais regras de entrada que restringem o tráfego de origem. Quando implantado dentro de uma Virtual Private Cloud (VPC), cada nó receberá um endereço IP privado dentro de uma ou mais sub-redes que você selecionar.

Nós individuais nunca podem ser acessados da Internet ou de instâncias do Amazon EC2 fora da VPC.

Você pode restringir ainda mais a entrada de rede no nível da sub-rede modificando as ACLs (listas de controle de acesso à rede).

O acesso para gerenciar a configuração e a infraestrutura do cluster é controlado separadamente do acesso ao terminal real do serviço Memcached ou Redis.

Usando o serviço AWS Identity and Access Management (IAM), é possível definir políticas que controlam quais usuários da AWS podem gerenciar a infraestrutura do Amazon ElastiCache.

Algumas das principais ações que um administrador pode executar incluem CreateCacheCluster, ModifyCacheCluster ou DeleteCacheCluster. Os clusters Redis também oferecem suporte às ações CreateReplicationGroup e CreateSnapshot, entre outras.

Serviços Adicionais

Entrega de Armazenamento e Conteúdo

Esta seção abrange dois serviços adicionais de armazenamento e entrega de conteúdo que são importantes para um arquiteto de soluções entender: Amazon CloudFront e AWS Storage Gateway.

Amazon CloudFront

O Amazon CloudFront é um serviço global CDN (Content Delivery Network). Ele se integra a outros produtos da AWS para oferecer aos desenvolvedores e empresas uma maneira fácil de distribuir conteúdo para usuários finais com baixa latência, altas velocidades de transferência de dados e sem compromissos mínimos de uso.

Uma rede de entrega de conteúdo (CDN) é uma rede distribuída globalmente de servidores de cache que acelera o download de páginas da web e outros conteúdos.

As CDNs usam a localização geográfica do DNS (Sistema de Nomes de Domínio) para determinar a localização geográfica de cada solicitação de uma página da web ou outro conteúdo e, em seguida, veiculam esse conteúdo nos servidores de cache de borda mais próximos a esse local, em vez do servidor da web original.

Uma CDN permite aumentar a escalabilidade de um site ou aplicação móvel facilmente em resposta a picos de tráfego de pico. Na maioria dos casos, o uso de uma CDN é totalmente transparente para os usuários finais simplesmente experimentam um melhor desempenho do site, enquanto a carga no site original é reduzida.

O Amazon CloudFront é o CDN da AWS. Ele pode ser usado para entregar seu conteúdo da web usando a rede global de pontos de presença da Amazon. Quando

um usuário solicita o conteúdo que você está exibindo com o Amazon CloudFront, ele é roteado para o local da borda que fornece a menor latência (atraso de tempo), para que o conteúdo seja entregue com o melhor desempenho possível.

Se o conteúdo já estiver no local de borda com a menor latência, o Amazon CloudFront o entregará imediatamente. Se o conteúdo não estiver atualmente naquele local de borda, o Amazon CloudFront o recuperará do servidor de origem, como um bucket do Amazon Simple Storage Service (Amazon S3) ou um servidor da Web, que armazena as versões definitivas originais dos seus arquivos.

O Amazon CloudFront é otimizado para trabalhar com outros serviços de nuvem da AWS como servidor de origem, incluindo buckets do Amazon S3, sites estáticos do Amazon S3, Amazon Elastic Compute Cloud (Amazon EC2) e Elastic Load Balancing.

O Amazon CloudFront também funciona perfeitamente com qualquer servidor de origem que não seja da AWS, como um servidor da Web local existente. O Amazon CloudFront também se integra ao Amazon Route 53.

O Amazon CloudFront suporta todo o conteúdo que pode ser exibido por HTTP ou HTTPS. Isso inclui todos os arquivos estáticos populares que fazem parte do seu aplicativo Web, como arquivos HTML, imagens, JavaScript e arquivos CSS, além de áudio, vídeo, arquivos de mídia ou downloads de software.

O Amazon CloudFront também oferece suporte à exibição de páginas da web dinâmicas, para que possa realmente ser usado para entregar todo o site. Por fim, o Amazon CloudFront oferece suporte ao streaming de mídia, usando HTTP e RTMP.

Noções Básicas do Amazon CloudFront

Há três conceitos principais que você precisa entender para começar a usar o CloudFront: distribuições, origens e controle de cache. Com esses conceitos, você pode facilmente usar o CloudFront para acelerar a entrega de conteúdo estático de seus sites.

Distribuições

Para usar o Amazon CloudFront, comece criando uma distribuição, identificada por um nome de domínio DNS, como d111111abcdef8.cloudfront.net.

Para veicular arquivos do Amazon CloudFront, basta usar o nome de domínio de distribuição no lugar do nome de domínio do seu site; o restante dos caminhos do arquivo permanece inalterado. Você pode usar o nome de domínio de distribuição do Amazon CloudFront como está ou criar um nome DNS fácil de usar em seu próprio domínio, criando um registro CNAME no Amazon Route 53 ou outro serviço DNS.

O CNAME é redirecionado automaticamente para o seu nome de domínio de distribuição do Amazon CloudFront.

Origens

Ao criar uma distribuição, você deve especificar o nome de domínio DNS da origem - o bucket do Amazon S3 ou o servidor HTTP - do qual deseja que o Amazon CloudFront obtenha a versão definitiva dos seus objetos (arquivos da web). Por exemplo:

- Bucket do Amazon S3: meubucket.s3.amazonaws.com
- Instância do Amazon EC2: ec2–203–0–113–25.compute 1.amazonaws.com
- Balanceador de carga elástico: my-load-balancer-1234567890.us-west-2.elb.amazonaws.com
- URL do site: meuservidorweb.meudominio.com

Controle de Cache

Uma vez solicitados e veiculados a partir de um local de borda, os objetos permanecem no cache até expirarem ou serem despejados para liberar espaço para o conteúdo solicitado com mais frequência.

Por padrão, os objetos expiram no cache após 24 horas. Quando um objeto expira, a próxima solicitação resulta no Amazon CloudFront encaminhando a solicitação

para a origem para verificar se o objeto está inalterado ou buscar uma nova versão se ela tiver sido alterada.

Opcionalmente, você pode controlar por quanto tempo os objetos ficam no cache do Amazon CloudFront antes de expirar. Para fazer isso, você pode optar por usar os cabeçalhos de controle de cache definidos pelo servidor de origem ou pode definir o tempo de vida (TTL) mínimo, máximo e padrão para objetos na distribuição do Amazon CloudFront.

Você também pode remover cópias de um objeto de todos os locais de borda do Amazon CloudFront a qualquer momento chamando a API (Application Program Interface) de invalidação.

Esse recurso remove o objeto de todos os locais de borda do Amazon CloudFront, independentemente do período de expiração definido para esse objeto no servidor de origem.

O recurso de invalidação foi projetado para ser usado em circunstâncias inesperadas, como corrigir um erro ou fazer uma atualização imprevista em um site, não como parte do seu fluxo de trabalho diário.

Em vez de invalidar objetos manual ou programaticamente, é uma prática recomendada usar um identificador de versão como parte do nome do caminho do objeto (arquivo). Por exemplo:

Arquivo antigo: assets / v1 / css / narrow.css

Novo arquivo: assets / v2 / css / narrow.css

Ao usar o controle de versão, os usuários sempre veem o conteúdo mais recente no Amazon CloudFront quando você atualiza o site sem usar a invalidação. As versões antigas expiram do cache automaticamente.

Recursos avançados do Amazon CloudFront

O CloudFront pode fazer muito mais do que simplesmente servir arquivos da Web estáticos.

Para começar a usar os recursos avançados do CloudFront, você precisará entender como usar comportamentos de cache e como restringir o acesso a conteúdo confidencial.

Conteúdo dinâmico, várias origens e comportamentos de servir cache em ativos estáticos, como descrito anteriormente, é uma maneira comum de usar uma CDN. Uma distribuição do Amazon CloudFront, no entanto, pode ser facilmente configurada para veicular conteúdo dinâmico, além de conteúdo estático e usar mais de um servidor de origem.

Você controla quais solicitações são atendidas por qual origem e como as solicitações são armazenadas em cache usando um recurso chamado comportamentos de cache.

Um comportamento de cache permite configurar uma variedade de funcionalidades do Amazon CloudFront para um determinado padrão de caminho da URL para arquivos em seu site.

A funcionalidade que você pode configurar para cada comportamento de cache inclui o seguinte:

- O padrão do caminho
- Qual origem encaminhar suas solicitações para
- Se encaminhar as strings de consulta para sua origem
- Se o acesso aos arquivos especificados requer URLs assinados
- Exigir acesso HTTPS

A quantidade de tempo que esses arquivos permanecem no cache do Amazon CloudFront (independentemente do valor de qualquer cabeçalho de controle de cache que sua origem adiciona aos arquivos)

Os comportamentos de cache são aplicados em ordem; se uma solicitação não corresponder ao primeiro padrão de caminho, ela desce para o próximo padrão de caminho. Normalmente, o último padrão de caminho especificado é * para corresponder a todos os arquivos.

Site inteiro

Usando comportamentos de cache e várias origens, você pode usar facilmente o Amazon CloudFront para servir todo o site e oferecer suporte a comportamentos diferentes para diferentes dispositivos clientes.

Conteúdo Privado

Em muitos casos, convém restringir o acesso ao conteúdo no Amazon CloudFront apenas para solicitantes selecionados, como assinantes pagos ou aplicativos ou usuários na rede da sua empresa.

O Amazon CloudFront fornece vários mecanismos para permitir que você possa pservir conteúdo privado. Esses incluem:

URLs Assinados

Use URLs válidos apenas entre determinados horários e, opcionalmente, de determinados endereços IP.

Cookies Assinados

Requer autenticação através de pares de chaves públicas e privadas.

Identidades de acesso de origem (OAI)

Restrinja o acesso a um bucket do Amazon S3 apenas a um usuário especial do Amazon CloudFront associado à sua distribuição. Essa é a maneira mais fácil de garantir que o conteúdo de um bucket seja acessado apenas pelo Amazon CloudFront.

Casos de Uso

Existem vários casos de uso em que o Amazon CloudFront é uma excelente opção, incluindo, entre outros:

Atendendo os ativos estáticos de sites populares

Os ativos estáticos, como imagens, CSS e JavaScript, compõem tradicionalmente a maior parte das solicitações para sites típicos. O uso do Amazon CloudFront acelerará a experiência do usuário e reduzirá a carga no próprio site.

Como veicular um site inteiro ou aplicativo da Web

O Amazon CloudFront pode veicular um site inteiro contendo conteúdo dinâmico e estático usando várias origens, comportamentos de cache e TTLs curtos para conteúdo dinâmico.

Exibição de conteúdo a usuários geograficamente amplamente distribuídos

O Amazon CloudFront aprimora o desempenho do site, especialmente para usuários distantes, e reduz a carga no servidor de origem.

Distribuição de software ou outros arquivos grandes

O Amazon CloudFront ajudará a acelerar o download desses arquivos para os usuários finais.

Exibição de mídia de streaming

O Amazon CloudFront ajuda a exibir mídia de streaming, como áudio e vídeo.

Também há casos de uso em que o CloudFront não é apropriado, incluindo:

Todas ou a maioria das solicitações vêm de um único local

Se todas ou a maioria de suas solicitações vierem de um único local geográfico, como um grande campus corporativo, você não aproveitará os vários locais de borda.

Todas ou a maioria das solicitações são feitas por meio de uma VPN corporativa

Da mesma forma, se os usuários se conectarem por meio de uma VPN (Rede virtual privada) corporativa, mesmo que sejam distribuídos, as solicitações de usuários aparecerão no CloudFront como originárias de um ou alguns locais.

Esses casos de uso geralmente não são beneficiados pelo uso do Amazon CloudFront.

Gateway de armazenamento da AWS

O AWS Storage Gateway é um serviço que conecta um dispositivo de software local com armazenamento baseado em nuvem para fornecer integração perfeita e segura entre o ambiente de TI de uma organização e a infraestrutura de armazenamento da AWS.

O serviço permite que você armazene dados com segurança na nuvem da AWS de maneira escalável e econômica.

O AWS Storage Gateway suporta protocolos de armazenamento padrão do setor que funcionam com seus aplicativos existentes. Ele fornece desempenho de baixa latência, armazenando em cache os dados acessados com frequência em criptografia e armazenando todos os seus dados no Amazon S3 ou Amazon Glacier.

O dispositivo de software do AWS Storage Gateway está disponível para download como uma imagem de máquina virtual (VM) que você instala em um host no seu datacenter e depois se registra na sua conta da AWS através do AWS Management Console.

O armazenamento associado ao dispositivo é exposto como um dispositivo iSCSI que pode ser montado por seus aplicativos locais.

Existem três configurações para o AWS Storage Gateway: volumes armazenados em cache do gateway, volumes armazenados no gateway e bibliotecas de fitas virtuais do gateway (VTL).

Volumes em cache do gateway

Os volumes em cache do gateway permitem expandir sua capacidade de armazenamento local no Amazon S3. Todos os dados armazenados em um volume em cache do gateway são movidos para o Amazon S3, enquanto os dados de leitura recentes são retidos no armazenamento local para fornecer acesso de baixa latência.

Embora cada volume esteja limitado a um tamanho máximo de 32 TB, um único gateway pode suportar até 32 volumes para um armazenamento máximo de 1 PB.

É possível tirar snapshots point-in-time para fazer backup do seu AWS Storage Gateway. Esses snapshots são executados de forma incremental e apenas os dados que foram alterados desde o último snapshot são armazenados.

Todos os dados de volume e snapshot em cache do Gateway são transferidos para o Amazon S3 através de conexões criptografadas de Secure Sockets Layer (SSL). Ele é criptografado em repouso no Amazon S3 usando a Criptografia no lado do servidor (SSE).

No entanto, você não pode acessar diretamente esses dados com a API do Amazon S3 ou outras ferramentas, como o console do Amazon S3, em vez disso, você deve acessá-lo através do serviço AWS Storage Gateway.

Volumes armazenados no gateway

Os volumes armazenados no gateway permitem armazenar seus dados em seu armazenamento local e faça backup assíncrono desses dados no Amazon S3.

Isso fornece acesso de baixa latência a todos os dados, além de fornecer backups externos tirando proveito da durabilidade do Amazon S3. É feito backup dos dados na forma de snapshots do Amazon Elastic Block Store (Amazon EBS).

Embora cada volume seja limitado a um tamanho máximo de 16 TB, um único gateway pode suportar até 32 volumes para um armazenamento máximo de 512 TB.

Semelhante aos volumes em cache do gateway, você pode tirar snapshots de seus volumes armazenados no gateway. O gateway armazena esses snapshots no Amazon S3 como snapshots do Amazon EBS.

Quando você tira uma nova captura instantânea, apenas os dados que foram alterados desde o último snapshot são armazenados. Você pode iniciar os snapshots de maneira programada ou única.

Como esses snapshots são armazenados como snapshots do Amazon EBS, você pode criar um novo volume do Amazon EBS a partir de um volume armazenado no gateway.

Todos os dados de volume e snapshot armazenados em gateway são transferidos para o Amazon S3 através de conexões SSL criptografadas. Ele é criptografado em repouso no Amazon S3 usando o SSE. No entanto, você não pode acessar esses dados com a API do Amazon S3 ou outras ferramentas, como o console do Amazon S3.

Se seu dispositivo local ou até mesmo um datacenter inteiro ficar indisponível, os dados no AWS Storage Gateway ainda poderão ser recuperados. Se for apenas o dispositivo indisponível, um novo dispositivo pode ser iniciado no datacenter e anexado ao AWS Storage existente.

Um novo dispositivo também pode ser iniciado em outro data center ou mesmo em uma instância do Amazon EC2 na nuvem.

Bibliotecas de fitas virtuais de gateway (VTL)

O Gateway-VTL oferece uma solução durável e econômica para arquivar seus dados na nuvem da AWS. A interface VTL permite que você aproveite sua infraestrutura

de aplicativo de backup baseado em fita existente para armazenar dados em cartuchos de fita virtuais criados no Gateway-VTL.

Uma fita virtual é análoga a um cartucho de fita físico, exceto que os dados são armazenados na nuvem da AWS. As fitas são criadas em branco no console ou programaticamente e, em seguida, preenchidas com dados de backup. Um gateway pode conter até 1.500 fitas (1 PB) do total de dados da fita.

As fitas virtuais aparecem na VTL do seu gateway, uma versão virtualizada de uma biblioteca de fitas física. As fitas virtuais são descobertas pelo seu aplicativo de backup usando seu procedimento padrão de inventário de mídia.

Quando o software da fita ejeta uma fita, ela é arquivada em uma prateleira de fita virtual (VTS) e armazenada no Amazon Glacier. Você permitiu 1 VTS por região da AWS, mas vários gateways na mesma região podem compartilhar um VTS.

Casos de Uso

Existem vários casos de uso em que o AWS Storage Gateway é uma excelente opção, incluindo, entre outros:

- Os volumes em cache do gateway permitem expandir o hardware de armazenamento local para o Amazon S3, permitindo armazenar muito mais dados sem aumentar drasticamente o hardware de armazenamento ou alterar os processos de armazenamento.
- Os volumes armazenados em gateway fornecem backup contínuo, assíncrono e seguro do seu armazenamento on-prememises sem novos processos ou hardware.
- As Gateway-VTLs permitem manter o software e os processos atuais de backup em fita, enquanto armazena seus dados de maneira mais econômica e simples na nuvem.

Segurança

A segurança da nuvem na AWS é a maior prioridade. Os clientes da AWS se beneficiam dos datacenters e arquiteturas de rede criados para atender aos requisitos das organizações mais sensíveis à segurança.

Uma vantagem da nuvem da AWS é que ela permite aos clientes escalar e inovar, mantendo um ambiente seguro. A segurança na nuvem é muito parecida com a segurança nos datacenters locais, apenas sem os custos de manutenção de instalações e hardware.

Na nuvem, você não precisa gerenciar servidores físicos ou dispositivos de armazenamento. Em vez disso, você usa ferramentas de segurança baseadas em software para monitorar e proteger o fluxo de informações dentro e fora da nuvem

Recursos.

Esta seção se concentrará em quatro serviços da AWS diretamente relacionados aos objetivos de segurança específicos: Serviço de diretório da AWS para gerenciamento de identidades, Serviço de gerenciamento de chaves da AWS (KMS), AWS CloudHSM para gerenciamento de chaves e AWS CloudTrail para auditoria.

Serviço de diretório da AWS

O AWS Directory Service é uma oferta de serviço gerenciado que fornece diretórios que contêm informações sobre sua organização, incluindo usuários, grupos, computadores e outros recursos.

Você pode escolher entre três tipos de diretório:

- Serviço de diretório da AWS para Microsoft Active Directory (Enterprise Edition), também conhecido como Microsoft AD
- AD simples
- Conector AD

Como uma oferta gerenciada, o AWS Directory Service foi projetado para reduzir as tarefas de gerenciamento de identidades, permitindo que você concentre mais seu tempo e recursos nos seus negócios.

Não há necessidade de criar sua própria topologia de diretório complexa e altamente disponível, porque cada diretório é implantado em várias zonas de disponibilidade, e o monitoramento detecta e substitui automaticamente os controladores de domínio que falham.

Além disso, a replicação de dados e os snapshots diários automáticos estão configurados para você. Não há software para instalar e a AWS lida com todas as atualizações de patches e software.

Serviço de Diretório da AWS para Microsoft Active Directory (Enterprise Edition)

O Serviço de Diretório da AWS para Microsoft Active Directory (Enterprise Edition) é um Microsoft Active Directory gerenciado hospedado na nuvem da AWS.

Ele fornece grande parte da funcionalidade oferecida pelo Microsoft Active Directory, além de integração com aplicativos da AWS. Com a funcionalidade adicional do Active Directory, você pode, por exemplo, configurar facilmente relações de confiança com os domínios existentes do Active Directory para estender esses diretórios aos serviços de nuvem da AWS.

AD simples

O Simple AD é um diretório compatível com o Microsoft Active Directory do AWS Directory Service, desenvolvido com o Samba 4. O Simple AD suporta recursos do Active Directory comumente usados, como contas de usuário, associações de grupos, instâncias do Amazon EC2 que ingressam no domínio executando Linux e Microsoft Windows, Kerberos logon único (SSO) e políticas de grupo.

Isso facilita ainda mais o gerenciamento de instâncias do Amazon EC2 executando Linux e Windows e a implantação de aplicativos Windows na nuvem da AWS.

Muitos dos aplicativos e ferramentas que você usa hoje em dia e que requerem suporte ao Microsoft Active Directory podem ser usados com o Simple AD.

As contas de usuário no Simple AD também podem acessar os aplicativos da AWS, como Amazon WorkSpaces, Amazon WorkDocs ou Amazon WorkMail.

Eles também podem usar as funções do AWS IAM para acessar o AWS Management Console e gerenciar os recursos da AWS. Por fim, o Simple AD fornece snapshots automáticos diários para permitir a recuperação point-in-time.

Observe que você não pode configurar relações de confiança entre o Simple AD e outros domínios do Active Directory. Outros recursos não suportados no momento da redação deste documento pelo Simple AD incluem atualização dinâmica de DNS, extensões de esquema, autenticação multifator (MFA), comunicação através de LDAP (Lightweight Directory Access Protocol), cmdlets do PowerShell AD e a transferência de Funções de Master Operations (FSMO).

Conector AD

O AD Connector é um serviço de proxy para conectar o Microsoft Active Directory local à nuvem da AWS sem exigir sincronização de diretório complexa ou o custo e a complexidade de hospedar uma infraestrutura de federação.

O AD Connector encaminha solicitações de entrada para os controladores de domínio do Active Directory para autenticação e fornece aos aplicativos a capacidade de consultar dados no diretório. Após a instalação, seus usuários podem usar suas credenciais corporativas existentes para fazer logon nos aplicativos da AWS, como Amazon WorkSpaces, Amazon WorkDocs ou Amazon WorkMail.

Com as permissões adequadas do IAM, eles também podem acessar o AWS Management Console e gerenciar recursos da AWS, como instâncias do Amazon EC2 ou buckets do Amazon S3.

Você também pode usar o AD Connector para habilitar o MFA, integrando-o à sua infraestrutura MFA baseada no Serviço de Discagem por Autenticação Remota

(RADIUS) existente para fornecer uma camada adicional de segurança quando os usuários acessam os aplicativos da AWS.

Com o AD Connector, você continua a gerenciar seu Active Directory como de costume. Por exemplo, a adição de novos usuários, a adição de novos grupos ou a atualização de senhas são realizadas usando ferramentas de administração de diretório padrão com o diretório local.

Portanto, além de fornecer uma experiência otimizada para seus usuários, o AD Connector permite a aplicação consistente de suas políticas de segurança existentes, como expiração de senha, histórico de senhas e bloqueios de conta, estejam os usuários acessando recursos no local ou na nuvem da AWS.

Casos de Uso

O AWS Directory Service fornece várias maneiras de usar o Microsoft Active Directory com outros serviços em nuvem da AWS.

Você pode escolher o serviço de diretório com os recursos necessários a um custo adequado ao seu orçamento.

Serviço de diretório da AWS para Microsoft Active Directory (Enterprise Edition) Este serviço de diretório é a melhor opção se você tiver mais de 5.000 usuários e precisar de uma relação de confiança configurada entre um diretório hospedado pela AWS e seus diretórios locais.

AD simples Na maioria dos casos, o AD simples é a opção mais barata e a melhor opção, se você tiver 5.000 ou menos usuários e não precisar dos recursos mais avançados do Microsoft Active Directory.

Conector AD

O AD Connector é a melhor opção quando você deseja usar o diretório onpremises existente com os serviços em nuvem da AWS.

Serviço de Gerenciamento de Chaves da AWS

O serviço de gerenciamento de chaves da AWS (KMS) e o gerenciamento de chaves do AWS CloudHSM são o gerenciamento de chaves criptográficas em um sistema de criptografia. Isso inclui lidar com a geração, troca, armazenamento, uso e substituição de chaves.

A AWS oferece dois serviços que oferecem a capacidade de gerenciar suas próprias chaves criptográficas simétricas ou assimétricas:

- AWS KMS: um serviço que permite gerar, armazenar, ativar / desativar e excluir chaves simétricas
- AWS CloudHSM: um serviço que fornece armazenamento de chaves criptográficas seguro, disponibilizando os Módulos de segurança de hardware (HSMs) na nuvem da AWS Serviço de gerenciamento de chaves da AWS (AWS KMS) O AWS KMS é um serviço gerenciado que facilita a criação e o controle do chaves de criptografia usadas para criptografar seus dados.

O AWS KMS permite criar chaves que nunca podem ser exportadas do serviço e que podem ser usadas para criptografar e descriptografar dados com base nas políticas definidas por você.

Ao usar o AWS KMS, você obtém mais controle sobre o acesso aos dados criptografados. Você pode usar os principais recursos de criptografia e gerenciamento diretamente em seus aplicativos ou através dos serviços de nuvem da AWS integrados ao AWS KMS.

Esteja você escrevendo aplicativos para a AWS ou usando os serviços de nuvem da AWS, o AWS KMS permite manter o controle sobre quem pode usar suas chaves e obter acesso aos dados criptografados.

Chaves gerenciadas pelo cliente

O AWS KMS usa um tipo de chave chamada CMK (Customer Master Key) para criptografar e descriptografar dados. CMKs são os recursos fundamentais que o AWS KMS gerencia.

Eles podem ser usados dentro do AWS KMS para criptografar ou descriptografar até 4 KB de dados diretamente. Eles também podem ser usados para criptografar chaves de dados geradas que são usadas para criptografar ou descriptografar grandes quantidades de dados fora do serviço.

As CMKs nunca podem deixar o AWS KMS descriptografado, mas as chaves de dados podem deixar o serviço não criptografado.

Chaves de Dados

Você usa chaves de dados para criptografar objetos de dados grandes em seu próprio aplicativo fora do AWS KMS. Quando você chama GenerateDataKey, o AWS KMS retorna uma versão em texto sem formatação da chave e do texto cifrado que contém a chave criptografada no CMK especificado. O AWS KMS rastreia qual CMK foi usado para criptografar a chave de dados.

Você usa a chave de dados de texto sem formatação no seu aplicativo para criptografar dados e normalmente armazena a chave criptografada ao lado dos dados criptografados.

As práticas recomendadas de segurança sugerem que você remova a chave de texto sem formatação da memória assim que possível após o uso. Para descriptografar dados em seu aplicativo, passe a chave de dados criptografados para a função Descriptografar.

O AWS KMS usa o CMK associado para descriptografar e recuperar sua chave de dados de texto sem formatação. Use a chave de texto sem formatação para descriptografar seus dados e remova a chave da memória.

Criptografia de Envelope

O AWS KMS usa criptografia de envelope para proteger os dados. O AWS KMS cria uma chave de dados, criptografa-a em um CMK e retorna versões em texto sem formatação e criptografadas da chave de dados para você.

Você usa a chave de texto sem formatação para criptografar dados e armazenar os dados criptografados.

chave ao lado dos dados criptografados. A chave deve ser removida da memória assim que possível após o uso. Você pode recuperar uma chave de dados em texto sem formatação apenas se tiver a chave de dados criptografados e tiver permissão para usar a chave mestra correspondente.

Contexto de criptografia

Todas as operações criptográficas do AWS KMS aceitam um mapa de chave / valor opcional de informações contextuais adicionais, denominado contexto de criptografia. O contexto especificado deve ser o mesmo para as operações de criptografia e descriptografia, ou a descriptografia não terá êxito.

O contexto de criptografia é registrado, pode ser usado para auditoria adicional e está disponível como contexto no idioma de política da AWS para autorização refinada baseada em políticas.

AWS CloudHSM

O AWS CloudHSM ajuda a atender às exigências corporativas, contratuais e requisitos de conformidade regulatórios para segurança de dados usando dispositivos HSM dedicados na nuvem da AWS.

Um HSM é um dispositivo de hardware que fornece armazenamento seguro de chaves e operações criptográficas dentro de um módulo de hardware resistente a violações. Os HSMs são projetados para armazenar com segurança o material da

chave criptográfica e usá-lo sem expô-lo fora dos limites criptográficos do dispositivo.

O AWS CloudHSM permite proteger suas chaves de criptografia nos HSMs projetados e validados de acordo com os padrões governamentais para o gerenciamento seguro de chaves.

Você pode gerar, armazenar e gerenciar com segurança as chaves criptográficas usadas para criptografia de dados de uma maneira que garanta que apenas você tenha acesso às chaves. O AWS CloudHSM ajuda a cumprir rigorosamente os principais requisitos de gerenciamento na nuvem da AWS sem sacrificar o desempenho do aplicativo.

Casos de Uso

Os serviços de gerenciamento de chaves da AWS atendem a várias necessidades de segurança que exigiriam grande esforço para implantar e gerenciar de outra forma, incluindo, entre outros:

Distribuição de chave simétrica escalável

Os algoritmos de criptografia simétrica exigem que a mesma chave seja usada para criptografar e descriptografar os dados. Isso é problemático porque a transferência da chave do remetente para o receptor deve ser feita por um canal seguro conhecido ou por algum processo "fora da banda".

Criptografia validada pelo governo

Certos tipos de dados (por exemplo, Indústria de cartões de pagamento - PCI - ou registros de informações de saúde) devem ser protegidos com criptografia validada por uma parte externa em conformidade com o (s) algoritmo (s) declarado (s) pela parte requerente.

AWS CloudTrail

O AWS CloudTrail fornece visibilidade da atividade do usuário registrando as chamadas de API feitas em sua conta. O AWS CloudTrail registra informações importantes sobre cada chamada da API, incluindo o nome da API, a identidade do chamador, a hora da chamada da API, os parâmetros de solicitação, e os elementos de resposta retornados pelo serviço da AWS.

Essas informações ajudam a rastrear as alterações feitas nos recursos da AWS e a solucionar problemas operacionais. O AWS CloudTrail facilita garantir a conformidade com políticas internas e padrões regulatórios.

O AWS CloudTrail captura chamadas da API da AWS e eventos relacionados feitos por ou em nome de uma conta da AWS e entrega arquivos de log para um bucket do Amazon S3 que você especificar.

Opcionalmente, você pode configurar o AWS CloudTrail para entregar eventos para um grupo de logs monitorado pelo Amazon CloudWatch Logs. Você também pode optar por receber o Amazon Simple Notification Service (Amazon SNS) sempre que um arquivo de log é entregue ao seu bucket.

Você pode criar uma trilha com o console do AWS CloudTrail, a AWS Command Line Interface (CLI) ou a API do AWS CloudTrail. Uma trilha é uma configuração que permite o registro da atividade da API da AWS e eventos relacionados em sua conta.

Você pode criar dois tipos de trilhas:

Uma trilha que se aplica a todas as regiões Quando você cria uma trilha que se aplica a todas as regiões da AWS, o AWS CloudTrail cria a mesma trilha em cada região, registra os arquivos de log em cada região e entrega os arquivos de log para o único bucket do Amazon S3 (e opcionalmente ao grupo de logs do Amazon CloudWatch Logs) que você especificar.

Essa é a opção padrão quando você cria uma trilha usando o console do AWS CloudTrail. Se você optar por receber notificações do Amazon SNS para entregas de arquivos de log, um tópico do Amazon SNS será suficiente para todas as regiões.

Se você optar por fazer com que o AWS CloudTrail envie eventos de uma trilha que se aplique a todas as regiões para um grupo de logs do Amazon CloudWatch Logs, os eventos de todas as regiões serão enviados para o único grupo de logs.

Uma trilha que se aplica a uma região Você especifica um bucket que recebe eventos somente dessa região. O bucket pode estar em qualquer região que você especificar. Se você criar trilhas individuais adicionais que se aplicam a regiões específicas, poderá fazer com que essas trilhas entreguem logs de eventos em um único bucket do Amazon S3.

Por padrão, seus arquivos de log são criptografados usando o Amazon S3 SSE. Você pode armazenar seus arquivos de log no seu depósito pelo tempo que quiser, mas também pode definir regras de ciclo de vida do Amazon S3 para arquivar ou excluir arquivos de log automaticamente.

O AWS CloudTrail normalmente fornece arquivos de log em 15 minutos após uma chamada de API. Além disso, o serviço publica novos arquivos de log várias vezes por hora, geralmente a cada cinco minutos.

Esses arquivos de log contêm chamadas de API de todos os serviços da conta que oferecem suporte ao AWS CloudTrail.

Ative o AWS CloudTrail em todas as suas contas da AWS. Em vez de configurar uma trilha para uma região, você deve habilitar trilhas para todas as regiões.

Casos de Uso

O AWS CloudTrail é benéfico para vários casos de uso:

Auditorias externas de conformidade

Sua empresa deve demonstrar conformidade com um conjunto de regulamentos pertinentes a alguns ou todos os dados transmitidos, processados e armazenados

nas suas contas da AWS. Os eventos do AWS CloudTrail podem ser usados para mostrar o grau em que você está em conformidade com os regulamentos.

Acesso não autorizado à sua conta da AWS

O AWS CloudTrail registra todas as tentativas de logon na sua conta da AWS, incluindo tentativas de login do AWS Management Console, chamadas da API do AWS Software Development Kit (SDK) e chamadas da API da AWS CLI. O exame de rotina dos eventos do AWS CloudTrail fornecerá as informações necessárias para determinar se sua conta da AWS está sendo direcionada para acesso não autorizado.

Google Analytics

O Analytics e os big data associados que ele exige, apresentam uma lista exclusiva de desafios para um arquiteto de soluções. O big data deve ser ingerido a uma taxa muito alta, armazenado em volume muito alto e processado com uma quantidade enorme de computação.

Freqüentemente, a necessidade de executar análises no big data é esporádica, com uma grande quantidade de infraestrutura de computação necessária regularmente por períodos muito pequenos.

A nuvem, com seu fácil acesso à computação e capacidade de armazenamento quase ilimitada, é ideal para enfrentar esses desafios de análise. Esta seção abrange vários serviços em nuvem da AWS que ajudarão você a resolver problemas de análise e big data no exame.

Amazon Kinesis

O Amazon Kinesis é uma plataforma para lidar com dados de streaming massivos na AWS, oferecendo serviços avançados para facilitar o carregamento e a análise

de dados de streaming, além de oferecer a capacidade de criar aplicativos de dados de streaming personalizados para necessidades especializadas.

O Amazon Kinesis é uma plataforma de dados de streaming que consiste em três serviços que abordam diferentes desafios de dados de streaming em tempo real:

- Amazon Kinesis Firehose: um serviço que permite carregar grandes volumes de dados de streaming na AWS
- Amazon Kinesis Streams: um serviço que permite criar aplicativos personalizados para análises mais complexas de dados de streaming em tempo real
- Amazon Kinesis Analytics: um serviço que permite analisar facilmente dados de streaming em tempo real com SQL padrão

Cada um desses serviços pode ser dimensionado para lidar com dados praticamente ilimitados.

Amazon Kinesis Firehose

O Amazon Kinesis Firehose recebe dados de fluxo e os armazena no Amazon S3, Amazon Redshift ou Amazon Elasticsearch.

Você não precisa escrever nenhum código, basta criar um fluxo de entrega e configurar o destino para seus dados. Os clientes gravam dados no fluxo usando uma chamada da API da AWS e os dados são automaticamente enviados para o destino apropriado.

Quando configurado para salvar um fluxo no Amazon S3, o Amazon Kinesis Firehose envia os dados diretamente para o Amazon S3. Para um destino Amazon Redshift, os dados são gravados primeiro no Amazon S3 e, em seguida, um comando COPY do Amazon Redshift é executado para carregar os dados no Amazon Redshift.

O Amazon Kinesis Firehose também pode gravar dados no Amazon Elasticsearch, com a opção de fazer backup dos dados simultaneamente no Amazon S3.

Amazon Kinesis Streams

O Amazon Kinesis Streams permite coletar e processar grandes fluxos de registros de dados em tempo real. Usando os AWS SDKs, você pode criar um aplicativo Amazon Kinesis Streams que processa os dados à medida que se movem pelo fluxo. Porque como o tempo de resposta para entrada e processamento de dados é quase em tempo real, o processamento geralmente é leve.

O Amazon Kinesis Streams pode ser dimensionado para oferecer suporte a fluxos de dados quase ilimitados, distribuindo os dados recebidos por vários shards. Se um fragmento ficar muito ocupado, ele poderá ser dividido em mais fragmentos para distribuir ainda mais a carga. O processamento é então executado nos consumidores, que lêem dados dos shards e executam o aplicativo Amazon Kinesis Streams.

Amazon Kinesis Analytics

Os serviços Amazon Kinesis suportam muitas cargas de trabalho estratégicas que, de outra forma, exigiriam grande esforço para implantar e gerenciar, incluindo, entre outros:

Ingestão de dados.

O primeiro desafio com um enorme fluxo de dados é aceitá-los de maneira confiável.

Sejam dados de usuários de sites altamente trafegados, dados de entrada de milhares de dispositivos de monitoramento ou quaisquer outras fontes de fluxos enormes, o Amazon Kinesis Firehose é uma excelente opção para garantir que todos os seus dados sejam armazenados com êxito na infraestrutura da AWS.

Processamento em tempo real de fluxos de dados maciços

As empresas geralmente precisam agir imediatamente com base no conhecimento obtido de um grande fluxo de dados, seja para alimentar um aplicativo de painel,

alterar estratégias de publicidade com base nas tendências de mídia social, alocar ativos com base em situações em tempo real ou uma série de outros cenários. O Amazon Kinesis Streams permite reunir esse conhecimento a partir dos dados em seu fluxo em tempo real.

É bom lembrar que, embora o Amazon Kinesis seja ideal para ingestão e processamento de fluxos de dados, é menos apropriado para tarefas em lote, como processos noturnos de extração, transformação, carga (ETL).

Amazon Elastic MapReduce (Amazon EMR)

O Amazon Elastic MapReduce (Amazon EMR) fornece uma estrutura Hadoop sob demanda totalmente gerenciada. O Amazon EMR reduz a complexidade e os custos iniciais da configuração do Hadoop e, combinado com a escala da AWS, oferece a capacidade de ativar o Hadoop e uma grande quantidade de clusters instantaneamente e inicie o processamento em minutos.

Ao iniciar um cluster do Amazon EMR, você especifica várias opções, sendo a mais importante:

- O tipo de instância dos nós no seu cluster
- O número de nós no seu cluster

A versão do Hadoop que você deseja executar (o Amazon EMR suporta várias versões recentes do Apache Hadoop e também várias versões do MapR Hadoop.)

Existem dois tipos de armazenamento que podem ser usados com o Amazon EMR:

Sistema de arquivos distribuídos do Hadoop (HDFS)

HDFS é o sistema de arquivos padrão que acompanha o Hadoop. Todos os dados são replicados em várias instâncias para garantir durabilidade. Amazonas

O EMR pode usar o armazenamento de instância do Amazon EC2 ou o Amazon EBS para HDFS. Quando um cluster é desligado, o armazenamento da instância é

perdido e os dados não persistem. O HDFS também pode fazer uso do armazenamento do Amazon EBS, negociando a relação custo-benefício do armazenamento de instância pela capacidade de desligar um cluster sem perder dados.

Sistema de arquivos EMR (EMRFS)

O EMRFS é uma implementação do HDFS que permite que os clusters armazenem dados no Amazon S3. O EMRFS permite obter a durabilidade e o baixo custo do Amazon S3, preservando seus dados, mesmo que o cluster esteja desligado.

Um fator chave que impulsiona o tipo de armazenamento que um cluster usa é se o cluster é persistente ou transitório. Um cluster persistente continua em execução 24 × 7 após o lançamento. Clusters persistentes são apropriados quando a análise contínua será executada nos dados. Para clusters persistentes, o HDFS é uma escolha comum. Clusters persistentes aproveitam a baixa latência do HDFS, especialmente no armazenamento de instância, quando operação constante significa que não há perda de dados ao desligar um cluster. Em outras situações, as cargas de trabalho de big data são frequentemente executadas inconsistentemente, e pode ser rentável desativar o cluster quando não estiver em uso.

Os clusters iniciados quando necessário e, em seguida, interrompidos imediatamente quando concluídos são chamados de clusters transitórios. O EMRFS é adequado para clusters transitórios, pois os dados persistem independentemente da vida útil do cluster. Você também pode optar por usar uma combinação de HDFS e EMRFS local para atender às suas necessidades de carga de trabalho.

Como o Amazon EMR é uma instância do Apache Hadoop, você pode usar o extenso ecossistema de ferramentas que funcionam sobre o Hadoop, como Hive, Pig e Spark. Muitas dessas ferramentas são suportadas nativamente e podem ser incluídas automaticamente quando você inicia o cluster, enquanto outros podem ser instalados através de ações de inicialização.

Casos de Uso

O Amazon EMR é adequado para um grande número de casos de uso, incluindo, entre outros:

Processamento de Log

O Amazon EMR pode ser usado para processar logs gerados por aplicativos da web e móveis. O Amazon EMR ajuda os clientes a transformar petabytes de dados não estruturados ou semiestruturados em informações úteis sobre seus aplicativos ou usuários.

Análise de fluxo de cliques

O Amazon EMR pode ser usado para analisar dados de fluxo de cliques para segmentar usuários e entender suas preferências. Os anunciantes também podem analisar fluxos de cliques e logs de impressões de publicidade para exibir anúncios mais eficazes.

Genômica e Ciências da Vida

O Amazon EMR pode ser usado para processar grandes quantidades de dados genômicos e outros grandes conjuntos de dados científicos de maneira rápida e eficiente. Processos que exigem anos de computação podem ser concluídos em um dia quando escalados em grandes clusters.

Pipeline de dados da AWS

O AWS Data Pipeline é um serviço da Web que ajuda a processar e mover dados de maneira confiável entre diferentes serviços de computação e armazenamento da AWS, e também fontes de dados locais, em intervalos especificados.

Com o AWS Data Pipeline, você pode acessar regularmente seus dados onde estão armazenados, transformá-los e processá-los em escala e transferir com eficiência

os resultados para serviços da AWS, como Amazon S3, Amazon Relational Database Service (Amazon RDS), Amazon DynamoDB e Amazon EMR.

Tudo no AWS Data Pipeline começa com o próprio pipeline. Um pipeline agenda e executa tarefas de acordo com a definição do pipeline. O agendamento é flexível e pode ser executado a cada 15 minutos, todos os dias, todas as semanas e assim por diante.

O pipeline interage com os dados armazenados nos nós de dados. Nós de dados são locais em que o pipeline lê dados de entrada ou grava dados de saída, como Amazon S3, um banco de dados MySQL ou um cluster Amazon Redshift. Os nós de dados podem estar na AWS ou nas suas instalações.

O pipeline executará atividades que representam cenários comuns, como mover dados de um local para outro, executar consultas do Hive e assim por diante.

As atividades podem exigir recursos adicionais para execução, como um cluster do Amazon EMR ou uma instância do Amazon EC2. Nessas situações, o AWS Data Pipeline inicia automaticamente os recursos necessários e os destrói quando a atividade é concluída.

Os fluxos de dados distribuídos geralmente têm dependências; só porque uma atividade está programada para ser executada não significa que há dados aguardando para serem processados. Para situações como essa, o AWS Data Pipeline suporta pré-condições, que são declarações condicionais que devem ser verdadeiras antes que uma atividade possa ser executada.

Isso inclui cenários como se uma chave do Amazon S3 está presente, se uma tabela do Amazon DynamoDB contém dados e assim por diante.

Se uma atividade falhar, a nova tentativa é automática. A atividade continuará tentando novamente até o limite que você configurar. Você pode definir ações a serem executadas no evento quando a atividade atingir esse limite sem êxito.

Casos de Uso

O AWS Data Pipeline pode ser usado para praticamente qualquer processo ETL em modo de lote.

Importação / Exportação da AWS

Um dos principais desafios do big data na nuvem da AWS é obter enormes conjuntos de dados na nuvem em primeiro lugar ou recuperá-los de volta ao local quando necessário.

Independentemente da quantidade de largura de banda configurada para fora do seu datacenter, há momentos em que há mais dados a serem transferidos do que podem se mover pela conexão em um período de tempo razoável.

O AWS Import / Export é um serviço que acelera a transferência de grandes quantidades de dados para dentro e fora da AWS usando dispositivos de armazenamento físico, ignorando a Internet. Os dados são copiados para um dispositivo na origem (seu data center ou uma região da AWS), enviados por mecanismos de envio padrão e, em seguida, copiados para o destino (seu data center ou uma região da AWS).

O AWS Import / Export possui dois recursos que oferecem suporte ao envio e recebimento de dados de sua infraestrutura da AWS: AWS Import / Export Snowball (AWS Snowball) e AWS Import / Export Disk.

AWS Snowball

O AWS Snowball usa dispositivos de armazenamento entregáveis fornecidos pela Amazon e enviados pela UPS. Cada AWS Snowball é protegido pelo AWS KMS e

fabricado com resistência física para proteger e proteger seus dados enquanto o dispositivo está em trânsito. No momento da redação deste artigo, o AWS Snowballs possui dois tamanhos: 50 TB e 80 TB, e a disponibilidade de cada um varia de acordo com a região.

O AWS Snowball fornece os seguintes recursos:

- Você pode importar e exportar dados entre os locais de armazenamento de dados local e o Amazon S3.
- A criptografia é imposta, protegendo seus dados em repouso e em trânsito físico.
- Você não precisa comprar ou manter seus próprios dispositivos de hardware.
- Você pode gerenciar suas tarefas através do console da AWS Snowball.

O AWS Snowball é seu próprio contêiner de remessa e a etiqueta de remessa é uma tela E Ink que mostra automaticamente o endereço correto quando o AWS Snowball está pronto para ser enviado. Você pode entregá-lo com a UPS, sem necessidade de caixa.

Com o AWS Snowball, você pode importar ou exportar terabytes ou mesmo petabytes de dados.

Disco de importação / exportação da AWS

O AWS Import / Export Disk suporta transferências de dados diretamente para fora dos dispositivos de armazenamento que você possui usando a rede interna de alta velocidade da Amazon. Coisas importantes a entender sobre o AWS Import / Export Disk incluem:

- Você pode importar seus dados para o Amazon Glacier e o Amazon EBS, além do Amazon S3.
- Você pode exportar dados do Amazon S3.
- A criptografia é opcional e não aplicada.
- Você compra e mantém seus próprios dispositivos de hardware.
- Você não pode gerenciar suas tarefas através do console do AWS Snowball.

Diferentemente do AWS Snowball, o AWS Import / Export Disk tem um limite superior de 16 TB.

Casos de Uso

O AWS Import / Export pode ser usado para praticamente qualquer situação em que você tenha mais dados para mover do que na conexão à Internet em um tempo razoável, incluindo, mas não se limitando a:

Migração de armazenamento

Quando as empresas encerram um data center, geralmente precisam mover grandes quantidades de armazenamento para outro local. O AWS Import / Export é uma tecnologia adequada para esse requisito.

Migrando aplicativos

A migração de um aplicativo para a nuvem geralmente envolve mover grandes quantidades de dados. Isso pode ser acelerado usando o AWS Import / Export.

DevOps

À medida que as organizações criavam aplicativos de software cada vez mais complexos, as equipes de desenvolvimento de TI evoluíram suas práticas de criação de software para obter mais flexibilidade, passando de modelos em cascata para práticas de desenvolvimento ágeis ou enxutas.

Essa mudança também se propagou para as equipes de operações, que embaçaram as linhas entre as equipes tradicionais de desenvolvimento e operações. A AWS fornece um ambiente flexível que facilitou o sucesso de organizações como Netflix, Airbnb, General Electric e muitas outras que adotaram o DevOps.

AWS OpsWorks

O AWS OpsWorks é um serviço de gerenciamento de configuração que ajuda a configurar e operar aplicativos usando o Chef. O AWS OpsWorks funcionará com aplicativos de qualquer nível de complexidade e é independente de qualquer padrão arquitetural específico.

Você pode definir a arquitetura de um aplicativo e a especificação de cada componente, incluindo instalação de pacotes, configuração de software e recursos como armazenamento.

O AWS OpsWorks suporta servidores Linux ou Windows, incluindo instâncias existentes do Amazon EC2 ou servidores em execução em seu próprio data center. Isso permite que as organizações usem um único serviço de gerenciamento de configuração para implantar e operar aplicativos nas arquiteturas híbridas.

Muitas soluções na AWS geralmente envolvem grupos de recursos, como instâncias do Amazon EC2 e Amazon RDS, que devem ser criados e gerenciados coletivamente. Por exemplo, essas arquiteturas geralmente requerem servidores de aplicativos, servidores de banco de dados, balanceadores de carga e assim por diante. Esse grupo de recursos geralmente é chamado de pilha.

Além de criar as instâncias e instalar os pacotes necessários, você normalmente precisa de uma maneira de distribuir aplicativos para os servidores de aplicativos, monitorar o desempenho da pilha, gerenciar segurança e permissões e assim por diante. O AWS OpsWorks fornece uma maneira simples e flexível de criar e gerenciar pilhas e aplicativos.

A pilha é o principal componente do AWS OpsWorks. É basicamente um contêiner para recursos da AWS - instâncias do Amazon EC2, instâncias do banco de dados Amazon RDS e assim por diante - que têm um objetivo comum e fazem sentido serem gerenciados logicamente juntos.

A pilha ajuda a gerenciar esses recursos como um grupo e define algumas configurações padrão, como o sistema operacional das instâncias do Amazon EC2 e a região da AWS. Se você deseja isolar alguns componentes da pilha da interação direta do usuário, é possível executar a pilha em uma nuvem virtual privada da Amazon (Amazon VPC). Cada pilha permite conceder aos usuários permissão para acessar a pilha e especificar quais ações eles podem executar.

Você pode usar o AWS OpsWorks ou o IAM para gerenciar permissões de usuário.

Observe que as duas opções não são mutuamente exclusivas; às vezes é desejável usar os dois.

Você define os elementos de uma pilha adicionando uma ou mais camadas. Uma camada representa um conjunto de recursos que atendem a uma finalidade específica, como balanceamento de carga, aplicativos Web ou hospedagem de um servidor de banco de dados. Você pode personalizar ou estender as camadas modificando as configurações padrão ou adicionando receitas do Chef para executar tarefas como instalar pacotes adicionais.

As camadas oferecem controle completo sobre quais pacotes estão instalados, como estão configurados, como os aplicativos são implantados e muito mais.

As camadas dependem das receitas do Chef para lidar com tarefas como instalar pacotes em instâncias, implantar aplicativos e executar scripts.

Um dos principais recursos do AWS OpsWorks é um conjunto de eventos do ciclo de vida que executam automaticamente um conjunto especificado de receitas no momento apropriado em cada instância.

Uma instância representa um único recurso de computação, como uma instância do Amazon EC2. Ele define a configuração básica do recurso, como sistema operacional e tamanho. Outras definições de configuração, como endereços IP Elastic ou volumes Amazon EBS, são definidas pelas camadas da instância. As receitas da camada concluem a configuração executando tarefas, como instalar e configurar pacotes e implantar aplicativos.

Você armazena aplicativos e arquivos relacionados em um repositório, como um bucket do Amazon S3 ou repositório Git.

Cada aplicativo é representado por um aplicativo, que especifica o tipo de aplicativo e contém as informações necessárias para implantar o aplicativo do repositório em suas instâncias, como a URL e a senha do repositório.

Quando você implanta um aplicativo, o AWS OpsWorks aciona um evento Deploy, que executa as receitas Deploy nas instâncias da pilha. Usando os conceitos de pilhas, camadas e aplicativos, você pode modelar e visualizar seu aplicativo e recursos de forma organizada.

Por fim, o AWS OpsWorks envia todas as suas métricas de recursos para o Amazon CloudWatch, facilitando a visualização de gráficos e a definição de alarmes para ajudá-lo a solucionar problemas e executar ações automatizadas com base no estado de seus recursos.

O AWS OpsWorks fornece muitas métricas personalizadas, como inativo da CPU, total de memória, carga média por um minuto e mais. Cada instância da pilha

possui monitoramento detalhado para fornecer informações sobre sua carga de trabalho.

Casos de Uso

O AWS OpsWorks oferece suporte a muitos esforços de DevOps, incluindo, entre outros:

Hospede aplicativos da web de várias camadas

O AWS OpsWorks permite modelar e visualizar seu aplicativo com camadas que definem como configurar um conjunto de recursos gerenciados em conjunto. Como o AWS OpsWorks usa a estrutura do Chef, você pode criar suas próprias receitas ou aproveitar centenas de configurações criadas pela comunidade.

Suporte à integração contínua

O AWS OpsWorks suporta os princípios do DevOps, como a integração contínua. Tudo no seu ambiente pode ser automatizado.

AWS CloudFormation

O AWS CloudFormation é um serviço que ajuda a modelar e configurar seus recursos da AWS, para que você possa gastar menos tempo gerenciando esses recursos e mais tempo concentrando-se nos aplicativos executados na AWS.

O AWS CloudFormation permite que as organizações implantem, modifiquem, e atualize os recursos de maneira controlada e previsível, aplicando o controle de versão à infraestrutura da AWS da mesma maneira que faria com o software.

O AWS CloudFormation oferece aos desenvolvedores e administradores de sistemas uma maneira fácil de criar e gerenciar uma coleção de recursos

relacionados da AWS, provisionando e atualizando-os de maneira ordenada e previsível.

Ao usar o AWS CloudFormation, você trabalha com modelos e pilhas.

Você cria modelos do AWS CloudFormation para definir seus recursos da AWS e suas propriedades. Um modelo é um arquivo de texto cujo formato está em conformidade com o padrão JSON. O AWS CloudFormation usa esses modelos como modelos para criar seus recursos da AWS.

Ao usar o AWS CloudFormation, você pode reutilizar seu modelo para configurar seus recursos de forma consistente e repetida. Apenas descreva seus recursos uma vez e, em seguida, provisione os mesmos recursos repetidamente em várias regiões.

Ao usar o AWS CloudFormation, você gerencia recursos relacionados como uma única unidade chamada pilha. Você cria, atualiza e exclui uma coleção de recursos criando, atualizando e excluindo pilhas. Todos os recursos em uma pilha são definidos pelo modelo AWS CloudFormation da pilha.

Suponha que você tenha criado um modelo que inclua um grupo de Auto Scaling, um balanceador de carga do Elastic Load Balancing e uma instância de banco de dados do Amazon RDS.

Para criar esses recursos, você cria uma pilha enviando seu modelo que define esses recursos e o AWS CloudFormation lida com todo o provisionamento para você. Após a criação de todos os recursos, o AWS CloudFormation informa que sua pilha foi criada.

Você pode começar a usar os recursos na sua pilha. Se a criação da pilha falhar, o AWS CloudFormation reverterá suas alterações excluindo os recursos que ele criou.

Freqüentemente, você precisará iniciar pilhas do mesmo modelo, mas com pequenas variações, como dentro de um Amazon VPC diferente ou usando AMIs de uma região diferente. Essas variações podem ser tratadas usando parâmetros.

Você pode usar parâmetros para personalizar aspectos do seu modelo em tempo de execução, quando a pilha é construída. Por exemplo, você pode passar o

tamanho do banco de dados Amazon RDS, tipos de instância do Amazon EC2, banco de dados e números de porta do servidor Web para o AWS CloudFormation quando criar uma pilha.

Ao alavancar os parâmetros do modelo, você pode usar um único modelo para muitas implantações de infraestrutura com diferentes valores de configuração.

Por exemplo, os tipos de instância do Amazon EC2, os limites de alarme do Amazon CloudWatch e as configurações de réplica de leitura do Amazon RDS podem diferir entre as regiões da AWS, se você receber mais tráfego de clientes nos Estados Unidos do que na Europa.

Você pode usar os parâmetros do modelo para ajustar as configurações e os limites em cada região separadamente e ainda assim garantir que o aplicativo seja implantado de forma consistente nas regiões.

Como os ambientes são dinâmicos por natureza, você inevitavelmente precisará atualizar os recursos da sua pilha de tempos em tempos. Não há necessidade de criar uma nova pilha e excluir a antiga; você pode simplesmente modificar o modelo da pilha existente.

Para atualizar uma pilha, crie um conjunto de alterações enviando uma versão modificada do modelo de pilha original, diferentes valores de parâmetros de entrada ou ambos. O AWS CloudFormation compara o modelo modificado com o modelo original e gera um conjunto de alterações.

O conjunto de mudanças lista as mudanças propostas. Após revisar as alterações, você pode executar o conjunto de alterações para atualizar sua pilha.

Quando chegar a hora e você precisar excluir uma pilha, o AWS CloudFormation excluirá a pilha e todos os recursos dessa pilha.

Se você deseja excluir uma pilha, mas ainda reter alguns recursos nessa pilha, poderá usar uma política de exclusão para reter esses recursos. Se um recurso não tiver uma política de exclusão, o AWS CloudFormation excluirá o recurso por padrão.

Após a exclusão de todos os recursos, o AWS CloudFormation sinaliza que sua pilha foi excluída com sucesso. Se o AWS CloudFormation não puder excluir um recurso, a pilha não será excluída. Quaisquer recursos que não foram excluídos permanecerão até que você possa excluir com êxito a pilha.

Caso de Uso

Ao permitir que você replique toda a sua pilha de infraestrutura com facilidade e rapidez, o AWS CloudFormation permite uma variedade de casos de uso, incluindo, entre outros:

Inicie rapidamente novos ambientes de teste

O AWS CloudFormation permite que as equipes de teste criem rapidamente um ambiente limpo para executar testes sem perturbar os esforços em andamento em outros ambientes.

Replicar Confiavelmente a Configuração Entre Ambientes

Como o AWS CloudFormation cria scripts para todo o ambiente, o erro humano é eliminado ao criar novas pilhas.

Iniciar aplicativos em novas regiões da AWS

Um único script pode ser usado em várias regiões para lançar pilhas de maneira confiável em diferentes mercados.

AWS Elastic Beanstalk

O AWS Elastic Beanstalk é a maneira mais rápida e simples de colocar um aplicativo em funcionamento na AWS. Os desenvolvedores podem simplesmente fazer upload do código do aplicativo e o serviço lida automaticamente com todos os

detalhes, como provisionamento de recursos, balanceamento de carga, dimensionamento automático e monitoramento.

A AWS compreende dezenas de serviços de blocos de construção, cada um dos quais expõe uma área de funcionalidade. Embora a variedade de serviços ofereça flexibilidade sobre como as organizações desejam gerenciar sua infraestrutura da AWS, pode ser um desafio descobrir quais serviços usar e como provisioná-los.

Com o AWS Elastic Beanstalk, você pode implantar e gerenciar rapidamente aplicativos na nuvem da AWS sem se preocupar com a infraestrutura que executa esses aplicativos.

O AWS Elastic Beanstalk reduz a complexidade do gerenciamento sem restringir a escolha ou o controle.

Existem componentes-chave que compõem o AWS Elastic Beanstalk e trabalham juntos para fornecer os serviços necessários para implantar e gerenciar aplicativos facilmente na nuvem.

Um aplicativo AWS Elastic Beanstalk é a coleção lógica desses componentes do AWS Elastic Beanstalk, que inclui ambientes, versões e configurações de ambiente.

No AWS Elastic Beanstalk, um aplicativo é conceitualmente semelhante a uma pasta. Uma versão do aplicativo refere-se a uma iteração específica e rotulada de código implementável para um aplicativo da web.

Uma versão do aplicativo aponta para um objeto Amazon S3 que contém o código implantável. Os aplicativos podem ter muitas versões e cada versão do aplicativo é única.

Em um ambiente de execução, as organizações podem implantar qualquer versão do aplicativo que já tenha carregado no aplicativo ou podem fazer upload e implantar imediatamente uma nova versão do aplicativo.

As organizações podem fazer upload de várias versões de aplicativos para testar as diferenças entre uma versão do aplicativo da Web e outra.

Um ambiente é uma versão de aplicativo implantada nos recursos da AWS. Cada ambiente executa apenas uma única versão do aplicativo por vez; no entanto, a

mesma versão ou versões diferentes podem ser executadas em tantos ambientes ao mesmo tempo, conforme necessário.

Quando um ambiente é criado, o AWS Elastic Beanstalk fornece os recursos necessários para executar a versão do aplicativo especificada.

Uma configuração de ambiente identifica uma coleção de parâmetros e configurações que definem como um ambiente e seus recursos associados se comportam.

Quando as definições de configuração de um ambiente são atualizadas, o AWS Elastic Beanstalk aplica automaticamente as alterações de recursos existentes ou exclui e implanta novos recursos, dependendo do tipo de alteração.

Quando um ambiente do AWS Elastic Beanstalk é iniciado, a camada, a plataforma e o tipo de ambiente são especificados. A camada de ambiente escolhida determina se o AWS Elastic Beanstalk fornece recursos para dar suporte a um aplicativo Web que lida com solicitações HTTP (S) ou um aplicativo que lida com tarefas de processamento em segundo plano.

Uma camada de ambiente cujo aplicativo da Web processa solicitações da Web é conhecida como camada do servidor da Web. Uma camada de ambiente cujo aplicativo executa tarefas em segundo plano é conhecida como camada de trabalho.

No momento da redação deste artigo, o AWS Elastic Beanstalk fornece suporte de plataforma para as linguagens de programação Java, Node.js, PHP, Python, Ruby e Go, com suporte para os contêineres da Web Tomcat, Passenger, Puma e Docker.

Casos de Uso

Uma empresa fornece um site para potenciais compradores de imóveis residenciais, vendedores e locatários a procurarem listagens de residências e apartamentos em mais de 110 milhões de residências. O site processa mais de três milhões de novas imagens diariamente. Ele recebe mais de 17.000 solicitações de

imagem por segundo em seu site durante o pico de tráfego de clientes para computadores e dispositivos móveis.

A empresa estava procurando maneiras de ser mais ágil com as implantações e capacitou seus desenvolvedores a se concentrarem mais na escrita de código, em vez de gastar tempo gerenciando e configurando servidores, bancos de dados, balanceadores de carga, firewalls e redes. Começou a usar o AWS Elastic Beanstalk como o serviço para implantar e dimensionar aplicativos e serviços da Web.

Os desenvolvedores tiveram o poder de fazer upload de código no AWS Elastic Beanstalk, que então lidava automaticamente com a implantação, desde o provisionamento de capacidade, o balanceamento de carga e o Auto Scaling, até o monitoramento da integridade do aplicativo.

Como a empresa ingere dados de maneira aleatória, executando feeds que despejam uma tonelada de trabalho no sistema de processamento de imagens de uma só vez, precisa aumentar sua frota de conversores de imagens para atender ao pico de demanda.

A empresa determinou que uma frota de trabalhadores do AWS Elastic Beanstalk para executar uma Python Imaging Library com código personalizado era a maneira mais simples de atender aos requisitos. Isso eliminou a necessidade de ter várias instâncias estáticas ou, pior ainda, tentar escrever sua própria configuração de Auto Scaling.

Ao mudar para o AWS Elastic Beanstalk, a empresa conseguiu reduzir os custos operacionais e aumentar a agilidade e a escalabilidade do seu sistema de processamento e entrega de imagens.

Características principais

O AWS Elastic Beanstalk fornece vários recursos de gerenciamento que facilitam a implantação e o gerenciamento de aplicativos na AWS.

As organizações têm acesso a métricas de monitoramento integradas do Amazon CloudWatch, como utilização média da CPU, contagem de solicitações e média

latência. Eles podem receber notificações por email através do Amazon SNS quando alterações de integridade ou servidores de aplicativos são adicionados ou removidos. Os logs do servidor para os servidores de aplicativos podem ser acessados sem a necessidade de efetuar login.

As organizações podem até optar por aplicar as atualizações automaticamente à plataforma subjacente que executa o aplicativo, como AMI, sistema operacional, idioma e estrutura e servidor de aplicativos ou proxy.

Além disso, os desenvolvedores mantêm controle total sobre os recursos da AWS que alimentam seus aplicativos e podem executar uma variedade de funções simplesmente ajustando as definições de configuração. Isso inclui configurações como:

- Selecionando o tipo de instância Amazon EC2 mais apropriado que corresponda aos requisitos de CPU e memória de seus aplicativos
- Escolhendo as opções corretas de banco de dados e armazenamento, como Amazon RDS, Amazon DynamoDB, Microsoft SQL Server e Oracle
- Habilitando o acesso de login às instâncias do Amazon EC2 para solução de problemas imediatos e diretos
- Aprimorando a segurança do aplicativo ativando o protocolo HTTPS no balanceador de carga
- Ajustando as configurações do servidor de aplicativos (por exemplo, configurações da JVM) e transmitindo variáveis de ambiente
- Ajuste as configurações do Auto Scaling para controlar as métricas e os limites usados para determinar quando adicionar ou remover instâncias de um ambiente
- Com o AWS Elastic Beanstalk, as organizações podem implantar um aplicativo rapidamente, mantendo o controle que desejam sobre a infraestrutura subjacente.

AWS Trusted Advisor

O AWS Trusted Advisor utiliza as melhores práticas aprendidas no histórico operacional agregado de atendimento a mais de um milhão de clientes da AWS. O AWS Trusted Advisor inspeciona seu ambiente de AW e faz recomendações quando existem oportunidades para economizar dinheiro, melhorar a disponibilidade e o desempenho do sistema ou ajudar a fechar brechas na segurança.

Você pode visualizar o status geral de seus recursos e estimativas de economia da AWS no painel do AWS Trusted Advisor.

O AWS Trusted Advisor é acessado no AWS Management Console. Além disso, o acesso programático ao AWS Trusted Advisor está disponível com o AWS

API de suporte.

O AWS Trusted Advisor fornece práticas recomendadas em quatro categorias: otimização de custos, segurança, tolerância a falhas e melhoria de desempenho.

O código de cores reflete as seguintes informações:

- Vermelho: ação recomendada
- Amarelo: Investigação recomendada
- Verde: nenhum problema detectado

Para cada verificação, você pode revisar uma descrição detalhada das melhores práticas recomendadas, um conjunto de critérios de alerta, diretrizes para ação e uma lista de recursos úteis sobre o tópico.

Todos os clientes da AWS têm acesso a quatro verificações do AWS Trusted Advisor, sem nenhum custo. As quatro verificações padrão do AWS Trusted Advisor são:

Limites de serviço

Verifica se há mais de 80% do limite de serviço. Esses valores são baseados em uma captura instantânea, portanto, o uso atual pode ser diferente e pode levar até 24 horas para refletir as alterações.

Portas específicas de grupos de segurança irrestritas

Verifica grupos de segurança quanto a regras que permitem acesso irrestrito (0.0.0.0/0) a portas específicas

Uso do IAM

Verifica o uso do AWS IAM MFA na conta raiz Verifica a conta raiz e avisa se o MFA não está ativado Os clientes com um plano de suporte comercial ou corporativo da AWS podem ver todos os AWS Trusted

Verificações do orientador - mais de 50 verificações.

Pode haver ocasiões em que uma verificação específica não seja relevante para alguns recursos em seu ambiente da AWS. Você pode excluir itens de uma verificação e, opcionalmente, restaurá-los mais tarde, a qualquer momento. O AWS Trusted Advisor atua como um especialista em nuvem personalizado e ajuda as organizações a fornecer seus recursos, seguindo as práticas recomendadas, identificando ineficiências, desperdícios, possíveis economias de custos e problemas de segurança.

AWS Config

O AWS Config é um serviço totalmente gerenciado que fornece um inventário de recursos da AWS, histórico de configuração e notificações de alterações na configuração para habilitar a segurança e a governança.

Com o AWS Config, você pode descobrir os recursos existentes e excluídos da AWS, determinar sua conformidade geral com as regras e mergulhar nos detalhes da configuração de um recurso a qualquer momento. Esses recursos permitem auditoria de conformidade, análise de segurança, rastreamento de alterações de recursos e solução de problemas.

O AWS Config fornece uma visão detalhada da configuração dos recursos da AWS em sua conta da AWS. Isso inclui como os recursos estão relacionados e como

foram configurados no passado, para que você possa ver como as configurações e os relacionamentos mudam ao longo do tempo.

O AWS Config define um recurso como uma entidade com a qual você pode trabalhar na AWS, como uma instância do Amazon EC2, um volume do Amazon EBS, um grupo de segurança ou um Amazon VPC.

Quando você ativa o AWS Config, ele primeiro descobre os recursos suportados da AWS existentes na sua conta e gera um item de configuração para cada recurso.

Um item de configuração representa uma exibição pontual dos vários atributos de um recurso suportado da AWS que existe em sua conta.

Os componentes de um item de configuração incluem metadados, atributos, relacionamentos, configuração atual e eventos relacionados.

O AWS Config gera itens de configuração quando a configuração de um recurso é alterada e mantém registros históricos dos itens de configuração dos seus recursos a partir do momento em que você inicia o gravador de configuração.

O gravador de configuração armazena as configurações dos recursos suportados em sua conta como itens de configuração. Por padrão, o AWS Config cria itens de configuração para todos os recursos suportados na região.

Se você não deseja que o AWS Config crie itens de configuração para todos os recursos suportados, você pode especificar os tipos de recursos que deseja acompanhar.

As organizações geralmente precisam avaliar a conformidade geral e o status de risco de uma perspectiva de configuração, visualizar tendências de conformidade ao longo do tempo e identificar quais alterações na configuração fizeram com que um recurso se afastasse da conformidade.

Uma regra de configuração da AWS representa as configurações desejadas para recursos específicos da AWS ou para uma conta inteira da AWS. Enquanto o AWS Config rastreia continuamente suas alterações na configuração de recursos, ele verifica se essas alterações violam alguma das condições de suas regras. Se um

recurso viola uma regra, o AWS Config sinaliza o recurso e a regra como não compatível e o notifica por meio do Amazon SNS.

O AWS Config facilita o rastreamento da configuração de recursos sem a necessidade de investimentos iniciais e evita a complexidade da instalação e atualização de agentes para coleta de dados ou manutenção de grandes bancos de dados.

Depois que o AWS Config é ativado, as organizações podem exibir detalhes atualizados continuamente de todos os atributos de configuração associados aos recursos da AWS.

Casos de Uso

Algumas das tarefas de gerenciamento de infraestrutura que o AWS Config permite incluem:

Descoberta

O AWS Config descobrirá os recursos existentes em sua conta, registrará sua configuração atual e capturará quaisquer alterações nessas configurações. O AWS Config também manterá os detalhes de configuração dos recursos que foram excluídos. Um snapshot abrangente de todos os recursos e seus atributos de configuração fornece um inventário completo dos recursos da sua conta.

Mudança de Configuração

Quando seus recursos são criados, atualizados ou excluídos, o AWS Config transmite essas alterações de configuração ao Amazon SNS para que você seja notificado de todas as alterações de configuração. O AWS Config representa relacionamentos entre recursos, para que você possa avaliar como uma alteração em um recurso pode afetar outros recursos.

Auditoria e conformidade contínuas

O AWS Config e o AWS Config Rules foram projetados para ajudá-lo a avaliar a conformidade com políticas internas e padrões regulatórios, fornecendo visibilidade da configuração de um recurso a qualquer momento e avaliando alterações relevantes na configuração em relação às regras que você pode definir.

Solução de problemas

Usando o AWS Config, você pode solucionar rapidamente problemas operacionais, identificando as alterações recentes na configuração de seus recursos.

Análise de Segurança e Incidentes

Recursos configurados corretamente melhoram sua postura de segurança. Os dados do AWS Config permitem monitorar as configurações de seus recursos continuamente e avaliar essas configurações quanto a possíveis falhas de segurança. Após um possível evento de segurança, o AWS Config permite examinar a configuração de seus recursos em qualquer ponto único do passado.

Características principais

No passado, as organizações precisavam pesquisar APIs de recursos e manter seu próprio banco de dados externo para gerenciamento de mudanças. O AWS Config resolve essa necessidade anterior e registra automaticamente as informações de configuração de recursos e avalia todas as regras acionadas por uma alteração.

A configuração do recurso e sua conformidade geral com as regras são apresentadas em um painel.

O AWS Config se integra ao AWS CloudTrail, um serviço que registra as chamadas da API da AWS para uma conta e entrega arquivos de log de uso da API a um bucket do Amazon S3. Se a alteração na configuração de um recurso foi resultado de uma chamada da API, o AWS Config também registra o ID do evento do AWS CloudTrail que corresponde à chamada da API que alterou a configuração do recurso.

As organizações podem aproveitar os logs do AWS CloudTrail para obter detalhes da chamada da API que foi feita - incluindo quem fez a chamada da API, a que horas e de qual endereço IP - a fim de solucionar problemas.

Quando uma alteração de configuração é feita em um recurso ou quando a conformidade de uma regra de configuração da AWS é alterada, é entregue uma mensagem de notificação que contém a configuração atualizada do recurso ou estado de conformidade da regra e informações importantes, como os valores antigos e novos de cada atributo alterado.

Além disso, o AWS Config envia notificações quando um arquivo de histórico de configuração é entregue ao Amazon S3 e quando o cliente inicia um snapshot de configuração. Todas essas mensagens são transmitidas para um tópico do Amazon SNS que você especificar.

As organizações podem usar o Console de Gerenciamento da AWS, a API ou a AWS CLI para obter detalhes de como era a configuração de um recurso em qualquer ponto do passado.

O AWS Config também entregará automaticamente um arquivo de histórico para o bucket do Amazon S3 especificado a cada seis horas, contendo todas as alterações nas configurações de recursos.

Segurança na AWS

Modelo de Responsabilidade Compartilhada

Antes de entrarmos nos detalhes de como a AWS protege seus recursos, devemos falar sobre como a segurança na nuvem é um pouco diferente da segurança nos seus datacenters locais.

Quando você move dados e sistemas de computador para a nuvem, as responsabilidades de segurança são compartilhadas entre você e seu provedor de serviços em nuvem. Nesse caso, a AWS é responsável por proteger a infraestrutura subjacente que suporta a nuvem, e você é responsável por qualquer coisa que você colocar na nuvem ou conectar-se à nuvem.

Esse modelo de responsabilidade compartilhada pode reduzir sua carga operacional de várias maneiras e, em alguns casos, pode até melhorar sua postura de segurança padrão sem ação adicional de sua parte

Programa de conformidade da AWS

A conformidade da AWS permite que os clientes entendam os controles robustos em vigor na AWS para manter a segurança e a proteção de dados na nuvem.

Ao criar sistemas sobre a infraestrutura em nuvem da AWS, você compartilha responsabilidades de conformidade com a AWS. Ao vincular os recursos de serviço amigáveis e focados em governança com os padrões de conformidade ou auditoria aplicáveis, os facilitadores de conformidade da AWS se baseiam em programas

tradicionais, ajudando você a estabelecer e operar em um ambiente de controle de segurança da AWS.

A infraestrutura de TI fornecida pela AWS é projetada e gerenciada em alinhamento com as melhores práticas de segurança e uma variedade de padrões de segurança de TI, incluindo (no momento da redação deste documento):

- Controle da organização de serviços (SOC) 1 / Declaração sobre normas para compromissos de atestado (SSAE) 16 / Normas internacionais para compromissos de garantia nº 3402 (ISAE) 3402 (anteriormente Declaração sobre normas de auditoria [SAS] 70)
- SOC 2
- SOC 3
- Lei Federal de Gerenciamento de Segurança da Informação (FISMA)
- Departamento de Defesa (DoD)
- Processo de Certificação e Acreditação de Garantia da Informação (DIACAP)
- Programa Federal de Gerenciamento de Riscos e Autorizações (FedRAMP)
- Guia de Requisitos de Segurança (SRG) do DoD Cloud Computing Níveis 2 e 4
- Padrão de segurança de dados da indústria de cartões de pagamento (PCI DSS) Nível 1
- Organização Internacional de Normalização (ISO) 9001 e ISO 27001
- Regulamentos sobre o tráfego internacional de armas (ITAR)
- Padrão Federal de Processamento de Informações (FIPS) 140-2

Além disso, a flexibilidade e o controle fornecidos pela plataforma da AWS permitem que os clientes implantem soluções que atendem a vários padrões específicos do setor, incluindo:

- Serviços de Informação da Justiça Criminal (CJIS)
- Aliança de Segurança na Nuvem (CSA)
- Lei de Privacidade e Direitos Educacionais da Família (FERPA)
- Lei de Portabilidade e Responsabilidade do Seguro de Saúde (HIPAA)
- Associação de Cinema da América (MPAA)

A AWS fornece uma ampla gama de informações sobre seu ambiente de controle de TI para os clientes por meio de whitepapers, relatórios, certificações, acreditações e outros atestados de terceiros.

Segurança de infraestrutura global da AWS

A AWS opera a infraestrutura de nuvem global usada para provisionar uma variedade de recursos básicos de computação, como processamento e armazenamento. A infraestrutura global da AWS inclui as instalações, a rede, o hardware e o software operacional (por exemplo, host operacional, sistema e software de virtualização) que oferecem suporte ao provisionamento e uso desses recursos.

A infraestrutura global da AWS é projetada e gerenciada de acordo com as melhores práticas de segurança, bem como com uma variedade de padrões de conformidade de segurança. Como cliente da AWS, você pode ter certeza de que está construindo arquiteturas da Web sobre algumas das infraestruturas de computação mais seguras do mundo.

Segurança Física e Ambiental

Os data centers da AWS são avançados, usando abordagens inovadoras de arquitetura e engenharia. A Amazon tem muitos anos de experiência no design, construção e operação de data centers em larga escala. Essa experiência foi aplicada à plataforma e infraestrutura da AWS.

Os datacenters da AWS estão alojados em instalações não-descritas. O acesso físico é estritamente controlado tanto no perímetro quanto nos pontos de entrada da construção pela equipe de segurança profissional, usando videovigilância, sistemas de detecção de intrusão e outros meios eletrônicos.

A equipe autorizada deve passar pela autenticação de dois fatores no mínimo duas vezes para acessar os andares do datacenter. Todos os visitantes e contratados são

obrigados a apresentar identificação e são assinados e acompanhados continuamente por pessoal autorizado.

A AWS fornece apenas acesso e informações ao datacenter a funcionários e contratados que possuem uma necessidade legítima dos negócios por esses privilégios. Quando um funcionário não tem mais uma necessidade comercial desses privilégios, seu acesso é imediatamente revogado, mesmo se continuar sendo funcionário da Amazon ou da AWS. Todo o acesso físico aos datacenters pelos funcionários da AWS é registrado e auditado rotineiramente.

Detecção e Supressão de Incêndio

Os datacenters da AWS possuem equipamentos automáticos de detecção e supressão de incêndio para reduzir riscos. O sistema de detecção de incêndio utiliza sensores de detecção de fumaça em todos os ambientes de data center, espaços de infraestrutura mecânica e elétrica, salas de resfriadores e salas de equipamentos de gerador.

Essas áreas são protegidas por sistemas de tubulação úmida, pré-ação com bloqueio duplo ou aspersores gasosos.

Energia

Os sistemas de energia elétrica do datacenter da AWS foram projetados para serem totalmente redundantes e mantidos sem impacto nas operações, 24 horas por dia e 7 dias por semana. As unidades de fonte de alimentação ininterrupta (UPS) fornecem energia de backup no caso de uma falha elétrica para cargas críticas e essenciais nas instalações. Os datacenters da AWS usam geradores para fornecer energia de backup para toda a instalação.

Clima e Temperatura

O controle climático é necessário para manter uma temperatura operacional constante para servidores e outros hardwares, o que evita o superaquecimento e reduz a possibilidade de interrupções no serviço.

Os data centers da AWS são criados para manter as condições atmosféricas em níveis ideais. O pessoal e os sistemas monitoram e controlam a temperatura e a umidade em níveis adequados.

Gerenciamento

A AWS monitora sistemas e equipamentos elétricos, mecânicos e de suporte à vida, para que quaisquer problemas sejam identificados imediatamente. A equipe da AWS realiza manutenção preventiva para manter a operacionalidade contínua dos equipamentos.

Desativação do dispositivo de armazenamento

Quando um dispositivo de armazenamento atinge o fim de sua vida útil, os procedimentos da AWS incluem um processo de descomissionamento desenvolvido para impedir que os dados do cliente sejam expostos a pessoas não autorizadas.

Gestão de Continuidade de Negócios

A infraestrutura da Amazon tem um alto nível de disponibilidade e fornece aos clientes os recursos para implantar uma arquitetura de TI resiliente. A AWS projetou seus sistemas para tolerar falhas de sistema ou hardware com impacto mínimo no cliente.

Continuidade de negócios do data center

O gerenciamento da AWS está sob a direção do Amazon Infrastructure Group.

Disponibilidade

Os datacenters são construídos em clusters em várias regiões globais. Todos os data centers estão online e atendem aos clientes; nenhum data center está "frio".

Em caso de falha, processos automatizados afastam o tráfego de dados da área afetada.

Os aplicativos principais são implantados em uma configuração N + 1, para que, no caso de uma falha no data center, haja capacidade suficiente para permitir que o tráfego seja balanceado por carga nos sites restantes.

A AWS oferece a seus clientes a flexibilidade de colocar instâncias e armazenar dados em várias regiões geográficas e também em várias zonas de disponibilidade em cada região.

Cada zona de disponibilidade é projetada como uma zona de falha independente.

Isso significa que as zonas de disponibilidade são fisicamente separadas dentro de uma região metropolitana típica e localizadas em planícies de inundação de menor risco (a categorização específica das zonas de inundação varia de acordo com a região). Além de possuir instalações discretas de geração de no-break e backup no local, elas são alimentadas por diferentes grades de utilidades independentes para reduzir ainda mais os pontos únicos de falha. As zonas de disponibilidade são todas redundantemente conectadas a vários provedores de transporte de nível 1

Você deve arquitetar o uso da AWS para tirar proveito de várias regiões e zonas de disponibilidade. A distribuição de aplicativos em várias zonas de disponibilidade oferece a capacidade de permanecer resiliente diante da maioria dos modos de falha, incluindo desastres naturais ou falhas no sistema.

Resposta a Incidentes

A equipe de gerenciamento de incidentes da Amazon emprega procedimentos de diagnóstico padrão do setor para direcionar a resolução durante eventos com impacto nos negócios.

Os operadores da equipe oferecem cobertura 24 × 7 × 365 para detectar incidentes e gerenciar o impacto e a resolução.

Comunicação

A AWS implementou vários métodos de comunicação interna em nível global para ajudar os funcionários a entender suas funções e responsabilidades individuais e comunicar eventos significativos em tempo hábil.

Esses métodos incluem programas de orientação e treinamento para funcionários recém-contratados, reuniões regulares de gerenciamento para atualizações sobre o desempenho dos negócios e outros assuntos e meios eletrônicos, como videoconferência, mensagens de correio eletrônico e postagem de informações pela intranet da Amazon.

A AWS também implementou vários métodos de comunicação externa para apoiar sua base de clientes e a comunidade.

Existem mecanismos para permitir que a equipe de suporte ao cliente seja notificada sobre problemas operacionais que afetam a experiência do cliente. Um painel de integridade do serviço está disponível e é mantido pela equipe de suporte ao cliente para alertar os clientes sobre quaisquer problemas que possam ter amplo impacto.

O Centro de segurança da AWS está disponível para fornecer detalhes de segurança e conformidade sobre a AWS. Os clientes também podem se inscrever nas ofertas de suporte da AWS, que incluem comunicação direta com a equipe de suporte ao cliente e alertas proativos para quaisquer problemas que impactem o cliente.

Segurança de rede

A rede da AWS foi arquitetada para permitir que você selecione o nível de segurança e resiliência apropriado para sua carga de trabalho. Para permitir a criação de arquiteturas da Web geograficamente dispersas e tolerantes a falhas com recursos de nuvem, a AWS implementou uma infraestrutura de rede de classe mundial que é cuidadosamente monitorada e gerenciada.

Arquitetura de rede segura

Dispositivos de rede, incluindo firewall e outros dispositivos de limite, existem para monitorar e controlar as comunicações nos limites externos da rede e nos principais limites internos da rede. Esses dispositivos de limite empregam conjuntos de regras, listas de controle de acesso (ACLs) e configurações para impor o fluxo de informações a serviços específicos do sistema de informações.

As ACLs ou políticas de fluxo de tráfego são estabelecidas em cada interface gerenciada, que gerencia e aplica o fluxo de tráfego. As políticas da ACL são aprovadas pelo Amazon Information Security. Essas políticas são enviadas automaticamente para garantir que essas interfaces gerenciadas apliquem as ACLs mais atualizadas.

Pontos de acesso seguros

A AWS colocou estrategicamente um número limitado de pontos de acesso na nuvem para permitir um monitoramento mais abrangente das comunicações de entrada e saída e do tráfego da rede.

Esses pontos de acesso do cliente são chamados de endpoint API (Application Programming Interface) e permitem acesso HTTP seguro (HTTPS), o que permite estabelecer uma sessão de comunicação segura com suas instâncias de armazenamento ou computação na AWS.

Para oferecer suporte a clientes com requisitos criptográficos do Federal Information Processing Standard (FIPS), os balanceadores de carga que terminam com Secure Sockets Layer (SSL) no AWS GovCloud (US) são compatíveis com FIPS 140-2.

Além disso, a AWS implementou dispositivos de rede dedicados ao gerenciamento de comunicações de interface com os ISPs (Internet Service Providers). A AWS emprega uma conexão redundante com mais de um serviço de comunicação em cada extremidade da rede da AWS voltada para a Internet. Essas conexões possuem dispositivos de rede dedicados.

Proteção de transmissão

Você pode conectar-se a um ponto de acesso da AWS via HTTP ou HTTPS usando SSL, um protocolo criptográfico desenvolvido para proteger contra interceptação, adulteração e falsificação de mensagens.

Monitoramento e proteção de rede

A rede da AWS fornece proteção significativa contra problemas de segurança de rede tradicionais e você pode implementar mais proteção. A seguir estão alguns exemplos:

Ataques de negação de serviço distribuída (DDoS)

Os endpoint da API da AWS são hospedados em uma grande infraestrutura de classe mundial em escala da Internet, que se beneficia da mesma experiência em engenharia que transformou a Amazon no maior varejista on-line do mundo.

Técnicas proprietárias de mitigação de DDoS são usadas. Além disso, as redes da AWS têm hospedagem múltipla em vários fornecedores para alcançar a diversidade de acesso à Internet.

Ataques intermediários (MITM)

Todas as APIs da AWS estão disponíveis por endpoint protegidos por SSL que fornecem autenticação de servidor.

As AMIs do Amazon Elastic Compute Cloud (Amazon EC2) geram automaticamente novos certificados de host Secure Shell (SSH) na primeira inicialização e os registram no console da instância. Você pode usar as APIs seguras para chamar o console e acessar os certificados de host antes de efetuar login na instância pela primeira vez. A AWS incentiva você a usar SSL para todas as suas interações.

Falsificação de IP

As instâncias do Amazon EC2 não podem enviar tráfego de rede falsificado. A infraestrutura de firewall controlada por host e controlada pela AWS não permitirá que uma instância envie tráfego com um endereço IP ou MAC (Machine Access Control) de origem diferente do seu.

Verificação de porta

Verificações de porta não autorizadas por clientes do Amazon EC2 são uma violação da Política de uso aceitável da AWS. As violações da Política de uso aceitável da AWS são levadas a sério e todas as violações relatadas são investigadas.

Os clientes podem denunciar suspeitas de abuso por meio dos contatos disponíveis no site da AWS. Quando a varredura de porta não autorizada é detectada pela AWS, ela é parada e bloqueada. As verificações de porta das instâncias do Amazon EC2 geralmente são ineficazes porque, por padrão, todas as portas de entrada nas instâncias do Amazon EC2 são fechadas e são abertas apenas pelo cliente.

O gerenciamento rigoroso de grupos de segurança pode reduzir ainda mais a ameaça de varreduras portuárias. Se você configurar o grupo de segurança para permitir o tráfego de qualquer origem para uma porta específica, essa porta específica ficará vulnerável a uma verificação de porta. Nesses casos, você deve usar as medidas de segurança apropriadas para proteger os serviços de escuta que possam ser essenciais para seus serviços de aplicativos sejam descobertos por uma verificação de porta não autorizada.

Por exemplo, um servidor Web deve claramente ter a porta 80 (HTTP) aberta ao mundo, e o administrador deste servidor é responsável pela segurança do software do servidor HTTP, como o Apache. Você pode solicitar permissão para realizar verificações de vulnerabilidade, conforme necessário, para atender aos seus requisitos de conformidade específicos.

Essas verificações devem ser limitadas às suas próprias instâncias e não devem violar a Política de uso aceitável da AWS. A aprovação avançada para esses tipos de verificações pode ser iniciada enviando uma solicitação pelo site da AWS.

Detecção de pacotes

Embora você possa colocar suas interfaces no modo promíscuo, o hipervisor não fornecerá nenhum tráfego para eles que não lhes seja endereçado. Mesmo duas instâncias virtuais pertencentes ao mesmo cliente localizadas no mesmo host físico não podem escutar o tráfego um do outro. Embora o Amazon EC2 forneça ampla proteção contra um cliente que, inadvertidamente ou maliciosamente, tente visualizar os dados de outro cliente, como prática padrão, você deve criptografar o tráfego confidencial.

Recursos de segurança da conta da AWS

A AWS fornece uma variedade de ferramentas e recursos que você pode usar para manter sua conta e recursos da AWS protegidos contra uso não autorizado. Isso inclui credenciais para controle de acesso, endpoint HTTPS para transmissão de dados criptografados, a criação de contas de usuário separadas do AWS Identity and Access Management (IAM) e o registro de atividades do usuário para monitoramento de segurança. Você pode tirar proveito de todas essas ferramentas de segurança, independentemente dos serviços da AWS selecionados.

Credenciais da AWS

Para ajudar a garantir que apenas usuários e processos autorizados acessem sua conta e recursos da AWS, a AWS usa vários tipos de credenciais para autenticação. Isso inclui senhas, chaves criptográficas, assinaturas digitais e certificados. A AWS também oferece a opção de exigir a autenticação multifator (MFA) para efetuar login na sua conta da AWS ou contas de usuário do IAM.

Por motivos de segurança, se suas credenciais foram perdidas ou esquecidas, você não pode recuperá-las ou fazer o download novamente. No entanto, você pode criar novas credenciais e, em seguida, desabilitar ou excluir o conjunto antigo de credenciais. De fato, a AWS recomenda que você altere (gire) suas chaves de acesso e certificados regularmente.

Para ajudá-lo a fazer isso sem afetar potencialmente a disponibilidade do seu aplicativo, a AWS oferece suporte a várias chaves e certificados de acesso simultâneo.

Com esse recurso, você pode alternar chaves e certificados para dentro e fora de operação regularmente, sem tempo de inatividade para o seu aplicativo. Isso pode ajudar a reduzir o risco de chaves ou certificados de acesso perdidos ou comprometidos.

A API do AWS IAM permite que você gire as chaves de acesso da sua conta da AWS e também das contas de usuário do IAM.

Senhas

As senhas são necessárias para acessar sua conta da AWS, contas de usuário individuais do IAM, fóruns de discussão da AWS e o Centro de suporte da AWS.

Você especifica a senha quando cria a conta e pode alterá-la a qualquer momento, acessando a página Credenciais de segurança.

As senhas da AWS podem ter até 128 caracteres e conter caracteres especiais, permitindo a criação de senhas muito fortes.

Você pode definir uma política de senha para suas contas de usuário do IAM para garantir que senhas fortes sejam usadas e que sejam alteradas com frequência. Uma política de senha é um conjunto de regras que definem o tipo de senha que um usuário do IAM pode definir.

Autenticação multifator da AWS (AWS MFA)

O AWS MFA é uma camada adicional de segurança para acessar os serviços da AWS Cloud. Ao ativar esse recurso opcional, você precisará fornecer um código de uso único de seis dígitos, além das credenciais padrão de nome de usuário e senha, antes que o acesso seja concedido às configurações da sua conta da AWS ou aos serviços e recursos da AWS Cloud.

Você obtém esse código de uso único de um dispositivo de autenticação que você mantém em sua posse física. Esse é o MFA porque mais de um fator de autenticação é verificado antes do acesso ser concedido: uma senha (algo que você sabe) e o código preciso do seu dispositivo de autenticação (algo que você possui).

Você pode ativar os dispositivos MFA para sua conta da AWS e para os usuários que você criou na sua conta da AWS com o AWS IAM. Além disso, você pode adicionar proteção MFA para acesso nas contas da AWS, pois quando você deseja permitir que um usuário criado em uma conta da AWS use uma função do IAM para acessar recursos em outra conta da AWS.

Você pode exigir que o usuário use o MFA antes de assumir a função como uma camada adicional de segurança.

O AWS MFA suporta o uso de tokens de hardware e dispositivos MFA virtuais. Os dispositivos virtuais MFA usam os mesmos protocolos que os dispositivos físicos MFA, mas podem ser executados em qualquer dispositivo de hardware móvel, incluindo um telefone inteligente.

Um dispositivo MFA virtual usa um aplicativo de software que gera códigos de autenticação de seis dígitos que são compatíveis com o padrão TOTP (Time-Based One-Time Password), conforme descrito na RFC 6238.

A maioria dos aplicativos MFA virtuais permite hospedar mais de um dispositivo MFA virtual, o que os torna mais convenientes que os dispositivos MFA de hardware. No entanto, você deve estar ciente de que, como um MFA virtual pode ser executado em um dispositivo menos seguro, como um telefone inteligente, um MFA virtual pode não fornecer o mesmo nível de segurança que um dispositivo MFA de hardware.

Você também pode aplicar a autenticação MFA para as APIs de serviço da AWS Cloud para fornecer uma camada extra de proteção contra ações poderosas ou privilegiadas, como encerrar instâncias do Amazon EC2 ou ler dados confidenciais armazenados no Amazon S3. Você faz isso adicionando um requisito de MFA a uma política de acesso do IAM.

Você pode anexar essas políticas de acesso a usuários, grupos do IAM ou recursos do IAM que suportam ACLs como buckets do Amazon S3, filas do Amazon Simple Queue Service (Amazon SQS) e tópicos do Amazon Simple Notification Service (Amazon SNS).

Chaves de Acesso

As chaves de acesso são criadas pelo AWS IAM e entregues como um par: o ID da chave de acesso (AKI) e a chave de acesso secreta (SAK). A AWS exige que todas as solicitações de API sejam assinadas pela SAK; ou seja, eles devem incluir uma assinatura digital que a AWS possa usar para verificar a identidade do solicitante.

Você calcula a assinatura digital usando uma função de hash criptográfico. Se você usar qualquer um dos SDKs da AWS para gerar solicitações, o cálculo da assinatura digital será feito para você.

O processo de assinatura não apenas ajuda a proteger a integridade da mensagem, impedindo a violação da solicitação enquanto ela está em trânsito, mas também ajuda a proteger contra possíveis ataques de reprodução.

Uma solicitação deve chegar à AWS dentro de 15 minutos do carimbo de data / hora na solicitação. Caso contrário, a AWS nega a solicitação.

A versão mais recente do processo de cálculo de assinatura digital no momento da redação deste documento é a versão 4 da assinatura, que calcula a assinatura usando o protocolo HMAC - Hashed Message Authentication Mode (HMAC) - Algoritmo de hash seguro (SHA) -256. A versão 4 fornece uma medida adicional de proteção em relação às versões anteriores, exigindo que você assine a mensagem usando uma chave derivada da sua SAK em vez de usar a própria SAK. Além disso, você obtém a chave de assinatura com base no escopo da credencial, o que facilita o isolamento criptográfico da chave de assinatura.

Como as chaves de acesso podem ser mal utilizadas se caírem em mãos erradas, a AWS recomenda que você as salve em um local seguro e não as incorpore ao seu código. Para clientes com grandes frotas de instâncias do Amazon EC2 com escala

elástica, o uso de funções do IAM pode ser uma maneira mais segura e conveniente de gerenciar a distribuição de chaves de acesso.

As funções do IAM fornecem credenciais temporárias, que não apenas são carregadas automaticamente na instância de destino, mas também são rotacionadas automaticamente várias vezes ao dia.

O Amazon EC2 usa um perfil de instância como um contêiner para uma função do IAM.

Quando você cria uma função do IAM usando o AWS Management Console, o console cria um perfil de instância automaticamente e atribui o mesmo nome à função à qual corresponde. Se você usar a AWS CLI, API ou um AWS SDK para criar uma função, crie o perfil da função e da instância como ações separadas e poderá atribuir nomes diferentes a eles.

Para iniciar uma instância com uma função do IAM, especifique o nome do seu perfil de instância. Ao iniciar uma instância usando o console do Amazon EC2, você pode selecionar uma função a ser associada à instância; no entanto, a lista exibida é na verdade uma lista de nomes de perfis de instância.

Pares de chaves

O Amazon EC2 suporta chaves SSA RSA 2048 para obter o primeiro acesso a uma instância do Amazon EC2. Em uma instância do Linux, o acesso é concedido mostrando a posse da chave privada SSH. Em uma instância do Windows, o acesso é concedido mostrando a posse da chave privada SSH para descriptografar a senha do administrador.

A chave pública está incorporada à sua instância e você usa a chave privada para entrar com segurança sem uma senha. Depois de criar suas próprias AMIs, você pode escolher outros mecanismos para efetuar logon em suas novas instâncias com segurança.

Você pode ter um par de chaves gerado automaticamente para você ao iniciar a instância ou fazer upload de seu próprio. Salve a chave privada em um local seguro no seu sistema e registre o local onde a salvou.

No Amazon CloudFront, você usa pares de chaves para criar URLs assinados para conteúdo privado, como quando você deseja distribuir conteúdo restrito pelo qual alguém pagou. Você cria pares de chaves do Amazon CloudFront usando a página Credenciais de segurança. Os pares de chaves do Amazon CloudFront podem ser criados apenas pela conta raiz e não podem ser criados pelos usuários do IAM.

Certificados X.509

Os certificados X.509 são usados para assinar solicitações baseadas em SOAP. Os certificados X.509 contêm uma chave pública associada a uma chave privada. Ao criar uma solicitação, você cria uma assinatura digital com sua chave privada e a inclui na solicitação, juntamente com seu certificado.

A AWS verifica se você é o remetente descriptografando a assinatura com a chave pública que está no seu certificado. A AWS também verifica se o certificado que você enviou corresponde ao certificado que você enviou para a AWS.

Para sua conta da AWS, você pode fazer com que a AWS crie um certificado X.509 e uma chave privada que você possa baixar ou fazer upload de seu próprio certificado usando a página Credenciais de segurança.

Para usuários do IAM, você deve criar o certificado X.509 (certificado de assinatura) usando software de terceiros. Ao contrário das credenciais da conta raiz, a AWS não pode criar um certificado X.509 para usuários do IAM.

Depois de criar o certificado, você o anexa a um usuário do IAM usando o IAM. Além das solicitações SOAP, os certificados X.509 são usados como certificados de servidor SSL / Transport Layer Security (TLS) para clientes que desejam usar HTTPS para criptografar suas transmissões.

Para usá-los para HTTPS, você pode usar uma ferramenta de código aberto como o OpenSSL para criar uma chave privada exclusiva. Você precisará da chave privada para criar a Solicitação de Assinatura de Certificado (CSR) enviada a uma Autoridade de Certificação (CA) para obter o certificado do servidor.

Você usará a CLI da AWS para fazer upload do certificado, chave privada e cadeia de certificados no IAM.

Você também precisará de um certificado X.509 para criar uma AMI do Linux personalizada para instâncias do Amazon EC2. O certificado é necessário apenas para criar uma AMI suportada por instância (em oposição a uma AMI suportada pelo Amazon Elastic Block Store [Amazon EBS]). Você pode fazer com que a AWS crie um certificado X.509 e uma chave privada que você possa baixar ou fazer upload de seu próprio certificado usando a página Credenciais de segurança.

AWS CloudTrail

O AWS CloudTrail é um serviço da web que registra as chamadas de API feitas em sua conta e entrega arquivos de log ao seu bucket do Amazon S3. O benefício do AWS CloudTrail é a visibilidade da atividade da conta, registrando as chamadas de API feitas em sua conta. O AWS CloudTrail registra as seguintes informações sobre cada chamada de API:

- O nome da API
- A identidade do chamador
- A hora da chamada da API
- Os parâmetros de solicitação
- Os elementos de resposta retornados pelo serviço de nuvem da AWS

Essas informações ajudam a rastrear as alterações feitas nos recursos da AWS e a solucionar problemas operacionais. O AWS CloudTrail facilita garantir a conformidade com políticas internas e padrões regulatórios.

O AWS CloudTrail oferece suporte à integridade do arquivo de log, o que significa que você pode provar a terceiros (por exemplo, auditores) que o arquivo de log enviado pelo AWS CloudTrail não foi alterado.

Arquivos de log validados são inestimáveis em investigações forenses e de segurança. Esse recurso foi criado usando algoritmos padrão do setor: SHA-256 para hash e SHA-256 com RSA para assinatura digital.

Isso inviabiliza computacionalmente modificar, excluir ou forjar arquivos de log do AWS CloudTrail sem detecção.

Segurança específica do serviço de nuvem da AWS

Não apenas a segurança é incorporada a todas as camadas da infraestrutura da AWS, mas também a cada um dos serviços disponíveis nessa infraestrutura. Os serviços em nuvem da AWS são projetados para trabalhar de maneira eficiente e segura com todas as redes e plataformas da AWS.

Cada serviço fornece recursos de segurança adicionais para proteger dados e aplicativos confidenciais.

Serviços de computação

A AWS fornece uma variedade de serviços de computação baseados em nuvem que incluem uma ampla seleção de instâncias de computação que podem ser aumentadas e diminuídas automaticamente para atender às necessidades de seu aplicativo ou empresa.

Segurança do Amazon Elastic Compute Cloud (Amazon EC2)

O Amazon EC2 é um componente essencial na infraestrutura como serviço (IaaS) da Amazon, fornecendo capacidade de computação redimensionável usando instâncias de servidor nos datacenters da AWS. O Amazon EC2 foi desenvolvido para facilitar a computação em escala da Web, permitindo obter e configurar a capacidade com atrito mínimo. Você cria e inicia instâncias, que são coleções de hardware e software da plataforma.

Vários níveis de segurança

A segurança no Amazon EC2 é fornecida em vários níveis: o sistema operacional (SO) da plataforma host, o SO de instância virtual ou SO convidado, um firewall e chamadas de API assinadas. Cada um desses itens se baseia nos recursos dos outros. O objetivo é impedir que os dados contidos no Amazon EC2 sejam interceptados por usuários ou sistemas não autorizados e tornar as instâncias do Amazon EC2 tão seguras quanto possível, sem sacrificar a flexibilidade na configuração exigida pelos clientes.

Hypervisor

Atualmente, o Amazon EC2 usa uma versão altamente personalizada do Xen hypervisor, aproveitando a paravirtualização (no caso de convidados Linux). Como os convidados paravirtualizados contam com o hipervisor para fornecer suporte para operações que normalmente exigem acesso privilegiado, o SO convidado não tem acesso elevado à CPU.

A CPU fornece quatro modos de privilégio separados: 0–3, chamados anéis. O anel 0 é o mais privilegiado e 3 o menos. O sistema operacional host executa no Ring 0. No entanto, em vez de executar no Ring 0 como a maioria dos sistemas operacionais, o SO convidado é executado no Ring 1 com menos privilégios e os aplicativos com menos privilégios no Ring 3.

Essa virtualização explícita dos recursos físicos leva a uma separação clara entre convidado e hipervisor, resultando em separação de segurança adicional entre os dois.

Isolamento de Instância

Diferentes instâncias em execução na mesma máquina física são isoladas uma da outra por meio do hipervisor Xen. A Amazon está ativa na comunidade Xen, que fornece à AWS o conhecimento dos últimos desenvolvimentos. Além disso, o

firewall da AWS reside na camada do hipervisor, entre a interface de rede física e a interface virtual da instância.

Todos os pacotes devem passar por essa camada; portanto, os vizinhos de uma instância não têm mais acesso a ela do que qualquer outro host na Internet e podem ser tratados como se estivessem em hosts físicos separados.

A RAM física é separada usando mecanismos semelhantes. As instâncias do cliente não têm acesso aos dispositivos de disco bruto, mas são apresentadas com discos virtualizados. A camada de virtualização de disco proprietária da AWS redefine automaticamente todos os blocos de armazenamento usados pelo cliente, para que os dados de um cliente nunca sejam expostos involuntariamente a outro cliente. Além disso, a memória alocada para os convidados é limpa (definida como zero) pelo hipervisor quando não é alocada para um convidado.

A memória não é retornada ao conjunto de memória livre disponível para novas alocações até que a limpeza da memória seja concluída.

Sistema Operacional Host

Os administradores com uma empresa que precisam acessar o plano de gerenciamento precisam usar o MFA para obter acesso aos hosts de administração criados com finalidade específica. Esses hosts administrativos são sistemas projetados, construídos, configurados e protegidos especificamente para proteger o plano de gerenciamento da nuvem.

Todo esse acesso é registrado e auditado. Quando um funcionário não tem mais uma necessidade comercial de acessar o plano de gerenciamento, os privilégios e o acesso a esses hosts e sistemas relevantes podem ser revogados.

Operador convidado

As instâncias virtuais do sistema são completamente controladas por você, o cliente. Você tem acesso root completo ou controle administrativo sobre contas, serviços e aplicativos.

A AWS não possui direitos de acesso às suas instâncias ou ao SO convidado. A AWS recomenda um conjunto básico de práticas recomendadas de segurança para incluir a desativação do acesso somente por senha aos seus convidados e o uso de alguma forma de MFA para obter acesso às suas instâncias (ou no mínimo, no acesso SSH Versão 2 com base em certificado).

Além disso, você deve empregar um mecanismo de escalonamento de privilégios com o registro por usuário. Por exemplo, se o sistema operacional convidado for Linux, após a proteção, sua instância deverá usar o SSHv2 baseado em certificado para acessar a instância virtual, desabilitar o logon raiz remoto, usar o log da linha de comando e usar o sudo para escalar privilégios.

Você deve gerar seus próprios pares de chaves para garantir que eles sejam exclusivos e não sejam compartilhados com outros clientes ou com a AWS. A AWS também oferece suporte ao uso do protocolo de rede SSH para permitir o login seguro nas instâncias do Amazon EC2 do UNIX / Linux.

A autenticação do SSH usado com a AWS é feita através de um par de chaves pública / privada para reduzir o risco de acesso não autorizado à sua instância. Você também pode conectar-se remotamente às instâncias do Windows usando o RDP (Remote Desktop Protocol) usando um certificado RDP gerado para sua instância. Você também controla a atualização e o patch do sistema operacional convidado, incluindo atualizações de segurança.

As AMIs baseadas em Windows e Linux fornecidas pela Amazon são atualizadas regularmente com os patches mais recentes. Portanto, se você não precisar preservar dados ou personalizações nas instâncias em execução da Amazon AMI, basta reiniciar novas instâncias com a AMI atualizada mais recente. Além disso, são fornecidas atualizações para o Amazon Linux AMI por meio dos repositórios yum do Amazon Linux.

Firewall

O Amazon EC2 fornece um firewall de entrada obrigatório configurado no modo de negar tudo padrão; Os clientes do Amazon EC2 devem abrir explicitamente as

portas necessárias para permitir o tráfego de entrada. O tráfego pode ser restrito pelo protocolo, pela porta de serviço e pelo endereço IP de origem (IP individual ou bloco CIDR).

O firewall pode ser configurado em grupos, permitindo que diferentes classes de instâncias tenham regras diferentes. Considere, por exemplo, o caso de um aplicativo da web tradicional de três camadas.

O grupo para os servidores da Web teria a porta 80 (HTTP) e / ou a porta 443 (HTTPS) aberta para a Internet. O grupo para os servidores de aplicativos teria a porta 8000 (específica do aplicativo) acessível apenas ao grupo de servidores da web.

O grupo para os servidores de banco de dados teria a porta 3306 (MySQL) aberta apenas para o grupo de servidores de aplicativos. Todos os três grupos permitiriam acesso administrativo na porta 22 (SSH), mas apenas da rede corporativa do cliente.

O nível de segurança oferecido pelo firewall é uma função de quais portas você abre e por qual duração e finalidade. Gerenciamento de tráfego bem informado e design de segurança ainda são necessários por instância. A AWS ainda recomenda que você aplique filtros de instância adicionais a firewalls baseados em host, como tabelas de IP ou Firewall do Windows e VPNs.

O estado padrão é negar todo o tráfego recebido, e você deve planejar cuidadosamente o que abrirá ao criar e proteger seus aplicativos.

Acesso à API

As chamadas de API para iniciar e encerrar instâncias, alterar parâmetros de firewall e executar outras funções são todas assinadas pela sua Amazon Secret Access Key, que pode ser a Chave de acesso secreto da conta da AWS ou a Chave de acesso secreto de um usuário criado com o AWS IAM.

Sem acesso à sua chave de acesso secreto, as chamadas da API do Amazon EC2 não podem ser feitas em seu nome. As chamadas de API também podem ser

criptografadas com SSL para manter a confidencialidade. A AWS recomenda sempre o uso de terminais de API protegidos por SSL.

Amazon Elastic Block Storage (Amazon EBS)

O Amazon EBS permite criar volumes de armazenamento de 1 GB a 16 TB que podem ser montados como dispositivos pelas instâncias do Amazon EC2. Os volumes de armazenamento se comportam como dispositivos de bloco não formatados brutos, com nomes de dispositivos fornecidos pelo usuário e uma interface de dispositivo de bloco. Você pode criar um sistema de arquivos sobre os volumes do Amazon EBS ou usá-los de qualquer outra maneira que usaria um dispositivo de bloco (como um disco rígido).

O acesso ao volume Amazon EBS é restrito à conta da AWS que criou o volume e aos usuários da conta da AWS criada com o AWS IAM (se o usuário tiver acesso concedido às operações do EBS). Todas as outras contas e usuários da AWS têm permissão negada para exibir ou acessar o volume.

Os dados armazenados nos volumes do Amazon EBS são redundantemente armazenados em vários locais físicos, como parte da operação normal desses serviços e sem custo adicional. No entanto, a replicação do Amazon EBS é armazenada na mesma zona de disponibilidade, não em várias zonas; portanto, é altamente recomendável que você realize snapshots regulares no Amazon S3 para durabilidade dos dados a longo prazo.

Para clientes que arquitetaram bancos de dados transacionais complexos usando o Amazon EBS, recomenda-se que os backups do Amazon S3 sejam executados por meio do sistema de gerenciamento de banco de dados, para que as transações e logs distribuídos possam ser verificados.

A AWS não executa automaticamente backups de dados mantidos em discos virtuais conectados a instâncias em execução no Amazon EC2.

Você pode disponibilizar publicamente os snapshots de volume do Amazon EBS para outras contas da AWS para usar como base para a criação de volumes duplicados.

O compartilhamento de snapshots de volume do Amazon EBS não fornece a outras contas da AWS a permissão para alterar ou excluir o snapshot original, pois esse direito é explicitamente reservado para a conta da AWS que criou o volume. Um snapshot do Amazon EBS é uma visualização em nível de bloco de um volume inteiro do Amazon EBS.

Observe que dados que não são visíveis no sistema de arquivos no volume, como arquivos que foram excluídos, podem estar presentes no snapshot do Amazon EBS. Se você deseja criar snapshots compartilhados, faça com tanto cuidado. Se um volume contiver dados confidenciais ou tiver arquivos excluídos, você deverá criar um novo volume do Amazon EBS para compartilhar.

Os dados a serem contidos na captura instantânea compartilhada devem ser copiados para o novo volume e a captura instantânea criada a partir do novo volume.

Os volumes do Amazon EBS são apresentados a você como dispositivos de bloco não formatados brutos que foram limpos antes de serem disponibilizados para uso. A limpeza ocorre imediatamente antes da reutilização, para que você possa ter certeza de que o processo de limpeza foi concluído. Se você tiver procedimentos que exijam a limpeza de todos os dados por meio de um método específico, poderá fazê-lo no Amazon EBS.

Você deve realizar um procedimento de limpeza especializado antes de excluir o volume para conformidade com os requisitos estabelecidos.

A criptografia de dados confidenciais geralmente é uma boa prática de segurança, e a AWS fornece a capacidade de criptografar volumes do Amazon EBS e seus snapshots com o Advanced Encryption Standard (AES) -256. A criptografia ocorre nos servidores que hospedam as instâncias do Amazon EC2, fornecendo criptografia de dados à medida que eles se movem entre as instâncias do Amazon EC2 e o armazenamento do Amazon EBS. Para poder fazer isso de forma eficiente

e com baixa latência, o recurso de criptografia do Amazon EBS está disponível apenas nos tipos de instância mais poderosos do Amazon EC2.

Segurança em Rede

A AWS fornece uma gama de serviços de rede que permitem criar uma rede logicamente isolada que você define, estabelecer uma conexão de rede privada com a AWS Cloud, usar um serviço DNS (Sistema de Nomes de Domínio) altamente disponível e escalável e fornecer conteúdo para seus usuários finais com baixa latência e altas velocidades de transferência de dados com um serviço da Web de entrega de conteúdo.

Segurança de balanceamento de carga elástico

O Elastic Load Balancing é usado para gerenciar o tráfego em uma frota de instâncias do Amazon EC2, distribuindo o tráfego para instâncias em todas as zonas de disponibilidade em uma região.

O Elastic Load Balancing possui todas as vantagens de um balanceador de carga local, além de vários benefícios de segurança:

Assume o trabalho de criptografia e descriptografia das instâncias do Amazon EC2 e o gerencia centralmente no balanceador de carga.

Oferece aos clientes um único ponto de contato e também pode servir como a primeira linha de defesa contra ataques à sua rede.

Quando usado em um Amazon VPC, oferece suporte à criação e gerenciamento de grupos de segurança associados ao seu Elastic Load Balancing para fornecer opções adicionais de rede e segurança.

Suporta criptografia de tráfego de ponta a ponta usando TLS (anteriormente SSL) nas redes que usam conexões HTTP seguras (HTTPS). Quando o TLS é usado, o

certificado do servidor TLS usado para finalizar as conexões do cliente pode ser gerenciado centralmente no balanceador de carga, em vez de em cada instância individual.

HTTPS / TLS usa uma chave secreta de longo prazo para gerar uma chave de sessão de curto prazo a ser usada entre o servidor e o navegador para criar a mensagem criptografada. O Elastic Load Balancing configura seu balanceador de carga com um conjunto de códigos predefinido usado para negociação TLS quando uma conexão é estabelecida entre um cliente e seu balanceador de carga.

O conjunto de cifras predefinido fornece compatibilidade com uma ampla variedade de clientes e usa algoritmos criptográficos robustos. No entanto, alguns clientes podem ter requisitos para permitir apenas cifras e protocolos específicos (por exemplo, PCI DSS], Sarbanes-Oxley Act [SOX]) dos clientes para garantir que os padrões sejam atendidos. Nesses casos, o Elastic Load Balancing fornece opções para selecionar configurações diferentes para protocolos e cifras TLS. Você pode optar por ativar ou desativar as cifras, dependendo de seus requisitos específicos.

Segurança da nuvem virtual privada da Amazon (Amazon VPC)

Normalmente, cada instância do Amazon EC2 iniciada recebe aleatoriamente um endereço IP público no espaço de endereço do Amazon EC2.

O Amazon VPC permite criar uma parte isolada da nuvem da AWS e iniciar instâncias do Amazon EC2 com endereços privados (RFC 1918) no intervalo de sua escolha (por exemplo, 10.0.0.0/16).

Você pode definir sub-redes no Amazon VPC, agrupando tipos semelhantes de instâncias com base no intervalo de endereços IP e, em seguida, configurar o roteamento e segurança para controlar o fluxo de tráfego dentro e fora das instâncias e sub-redes.

Os recursos de segurança no Amazon VPC incluem grupos de segurança, ACLs de rede, tabelas de roteamento e gateways externos. Cada um desses itens é complementar ao fornecimento de uma rede segura e isolada que pode ser

estendida por meio da habilitação seletiva de acesso direto à Internet ou conectividade privada a outra rede.

As instâncias do Amazon EC2 em execução no Amazon VPC herdam todos os benefícios descritos abaixo relacionados ao SO convidado e à proteção contra detecção de pacotes. Observe, no entanto, que você deve criar grupos de segurança especificamente para o Amazon VPC; quaisquer grupos de segurança do Amazon EC2 que você criou não funcionarão dentro do seu Amazon VPC.

Além disso, os grupos de segurança do Amazon VPC possuem recursos adicionais que os grupos de segurança do Amazon EC2 não possuem, como alterar o grupo de segurança após o lançamento da instância e especificar qualquer protocolo com um número de protocolo padrão (em vez de apenas TCP, User Datagram Protocol [UDP] ou Internet Control Message Protocol [ICMP]).

Cada Amazon VPC é uma rede distinta e isolada dentro da nuvem; o tráfego de rede em cada Amazon VPC é isolado de todos os outros Amazon VPCs. No momento da criação, você seleciona um intervalo de endereços IP para cada Amazon VPC. Você pode criar e conectar um gateway da Internet, gateway virtual privado ou ambos para estabelecer conectividade externa, sujeito aos seguintes controles.

Chamadas de acesso à API para criar e excluir Amazon VPCs; alterar os parâmetros de roteamento, grupo de segurança e ACL da rede; e executar outras funções são todas assinadas pela chave de acesso secreto da Amazon, que pode ser a chave de acesso secreto da conta da AWS ou a chave de acesso secreto de um usuário criado com o AWS IAM. Sem acesso à sua chave de acesso secreto, as chamadas à API do Amazon VPC não podem ser feitas em seu nome. Além disso, as chamadas de API podem ser criptografadas com SSL para manter a confidencialidade.

A AWS recomenda sempre o uso de terminais de API protegidos por SSL. O AWS IAM também permite que um cliente controle ainda mais quais APIs um usuário recém-criado tem permissão para chamar.

Sub-redes e tabelas de rotas Você cria uma ou mais sub-redes dentro de cada Amazon VPC; cada instância iniciada no Amazon VPC é conectada a uma sub-rede. Os ataques de segurança tradicionais da camada 2, incluindo falsificação de MAC e

falsificação de ARP, são bloqueados. Cada sub-rede em um Amazon VPC é associada a uma tabela de roteamento e todo o tráfego de rede que sai da sub-rede é processado pela tabela de roteamento para determinar o destino.

Firewall (grupos de segurança)

Como o Amazon EC2, o Amazon VPC suporta um firewall de solução completa, permitindo a filtragem no tráfego de entrada e saída de uma instância. O grupo padrão permite a comunicação de entrada de outros membros do mesmo grupo e a comunicação de saída para qualquer destino.

O tráfego pode ser restringido por qualquer protocolo IP, porta de serviço e endereço IP de origem / destino (bloco individual de IP ou CIDR). O firewall não é controlado pelo sistema operacional convidado; em vez disso, só pode ser modificado através da invocação de APIs do Amazon VPC.

A AWS suporta a capacidade de conceder acesso granular a diferentes funções administrativas nas instâncias e no firewall, permitindo, assim, implementar segurança adicional por meio da separação de tarefas.

O nível de segurança oferecido pelo firewall é uma função de quais portas você abre e por qual duração e finalidade. O gerenciamento de tráfego e o design de segurança bem informados ainda são necessários por instância.

A AWS ainda recomenda que você aplique filtros adicionais por instância com firewalls baseados em host, como IPtables ou Windows Firewall.

ACLs de rede

Para adicionar uma camada adicional de segurança ao Amazon VPC, você pode configurar as ACLs de rede. Esses são filtros de tráfego sem estado que se aplicam a todo o tráfego de entrada ou saída de uma sub-rede no Amazon VPC.

Essas ACLs podem conter regras ordenadas para permitir ou negar tráfego com base no protocolo IP, por porta de serviço e endereço IP de origem / destino.

Como grupos de segurança, as ACLs de rede são gerenciadas por meio de APIs do Amazon VPC, adicionando uma camada adicional de proteção e permitindo segurança adicional por meio da separação de tarefas.

Gateway Privado Virtual

Um gateway privado virtual permite conectividade privada entre o Amazon VPC e outra rede. O tráfego de rede em cada gateway privado virtual é isolado do tráfego de rede em todos os outros gateways privados virtuais.

Você pode estabelecer conexões VPN para o gateway privado virtual a partir de dispositivos de gateway em suas instalações. Cada conexão é protegida por uma chave pré-compartilhada em conjunto com o endereço IP do cliente de dispositivo de gateway.

Gateway de Internet

Um gateway da Internet pode ser anexado a um Amazon VPC para permitir conectividade direta ao Amazon S3, outros serviços da AWS e a Internet. Cada instância que deseja esse acesso deve ter um IP Elastic associado a ele ou rotear o tráfego através de uma instância de Network Address Translation (NAT)

Instâncias dedicadas

Em um Amazon VPC, você pode iniciar instâncias do Amazon EC2 que são fisicamente isolados no nível do hardware do host (ou seja, eles serão executados no hardware de um único inquilino).

Um Amazon VPC pode ser criado com locação "dedicada", para que todas as instâncias iniciadas no Amazon VPC usem esse recurso.

Como alternativa, um Amazon VPC pode ser criado com locação "padrão", mas você pode especificar a locação dedicada para instâncias específicas lançado nele.

Amazon CloudFront Security

O Amazon CloudFront oferece aos clientes uma maneira fácil de distribuir conteúdo para usuários finais com baixa latência e altas velocidades de transferência de dados. Ele fornece conteúdo dinâmico, estático e de streaming usando uma rede global de locais de borda.

As solicitações de objetos dos clientes são roteadas automaticamente para o local da borda mais próximo, para que o conteúdo seja entregue com o melhor desempenho possível.

O Amazon CloudFront é otimizado para trabalhar com outros serviços da AWS, como Amazon S3, Amazon EC2, Elastic Load Balancing e Amazon Route 53. Ele também funciona perfeitamente com qualquer servidor de origem que não seja da AWS que armazene as versões definitivas originais de seus arquivos.

O Amazon CloudFront exige que todas as solicitações feitas à sua API de controle sejam autenticadas, para que apenas usuários autorizados possam criar, modificar ou excluir suas próprias distribuições do Amazon CloudFront. As solicitações são assinadas com uma assinatura HMAC-SHA-1 calculada a partir da solicitação e da chave privada do usuário.

Além disso, a API de controle Amazon CloudFront é acessível apenas por endpoint habilitados para SSL.

Não há garantia de durabilidade dos dados mantidos nos locais de borda do Amazon CloudFront. Às vezes, o serviço pode remover objetos de locais de borda se esses objetos não forem solicitados com frequência. A durabilidade é fornecida pelo Amazon S3, que funciona como servidor de origem do Amazon CloudFront, mantendo as cópias definitivas originais dos objetos entregues pelo Amazon CloudFront.

Se você deseja controlar quem pode baixar o conteúdo do Amazon CloudFront, pode ativar o recurso de conteúdo privado do serviço. Esse recurso tem dois componentes.

A primeira controla como o conteúdo é entregue a partir da localização de borda do Amazon CloudFront para os espectadores na Internet. O segundo controla como os locais de borda do Amazon CloudFront acessam objetos no Amazon S3. O Amazon CloudFront também oferece suporte à restrição geográfica, que restringe o acesso ao seu conteúdo com base na localização geográfica dos seus visualizadores.

Para controlar o acesso às cópias originais de seus objetos no Amazon S3, o Amazon CloudFront permite criar uma ou mais identidades de acesso à origem e associá-las às suas distribuições. Quando uma identidade de acesso de origem é associada a uma distribuição do Amazon CloudFront, a distribuição usa essa identidade para recuperar objetos do Amazon S3.

Você pode usar o recurso ACL do Amazon S3, que limita o acesso a essa identidade de acesso de origem para que a cópia original do objeto não seja legível publicamente.

Para controlar quem pode baixar objetos de locais de borda do Amazon CloudFront, o serviço usa um sistema de verificação de URL assinado. Para usar esse sistema, primeiro crie um par de chaves público-privado e faça o upload da chave pública em sua conta pelo AWS Management Console.

Em seguida, você configura sua distribuição do Amazon CloudFront para indicar quais contas você autorizaria a assinar solicitações - você pode indicar até cinco contas da AWS nas quais confia para assinar solicitações. Ao receber solicitações, você criará documentos de política indicando as condições sob as quais deseja que o Amazon CloudFront sirva seu conteúdo.

Esses documentos de política podem especificar o nome do objeto solicitado, a data e a hora da solicitação e o IP de origem (ou intervalo CIDR) do cliente que está fazendo a solicitação. Você então calcula o hash SHA-1 do seu documento de política e assina isso usando sua chave privada.

Por fim, você inclui o documento de política codificado e a assinatura como parâmetros da sequência de consulta ao fazer referência a seus objetos. Quando o Amazon CloudFront recebe uma solicitação, ele decodifica a assinatura usando sua

chave pública. O Amazon CloudFront atenderá apenas solicitações que tenham um documento de política válido e assinatura correspondente.

Observe que o conteúdo privado é um recurso opcional que deve ser ativado quando você configura sua distribuição do Amazon CloudFront. O conteúdo entregue sem esse recurso ativado será publicamente legível.

O Amazon CloudFront oferece a opção de transferir conteúdo por uma conexão criptografada (HTTPS). Por padrão, o Amazon CloudFront aceita solicitações nos protocolos HTTP e HTTPS. No entanto, você também pode configurar o Amazon CloudFront para exigir HTTPS para todas as solicitações ou fazer com que o Amazon CloudFront redirecione solicitações HTTP para HTTPS.

Você pode até configurar distribuições do Amazon CloudFront para permitir HTTP para alguns objetos, mas exigir HTTPS para outros objetos.

Armazenamento

A AWS fornece armazenamento de dados de baixo custo com alta durabilidade e disponibilidade. A AWS oferece opções de armazenamento para backup, arquivamento e recuperação de desastres e também para armazenamento de blocos e objetos.

Segurança do Amazon Simple Storage Service (Amazon S3)

O Amazon S3 permite fazer upload e recuperar dados a qualquer momento, de qualquer lugar da Web. O Amazon S3 armazena dados como objetos dentro de buckets. Um objeto pode ser qualquer tipo de arquivo: um arquivo de texto, uma foto, um vídeo e muito mais.

Ao adicionar um arquivo ao Amazon S3, você tem a opção de incluir metadados com o arquivo e definir permissões para controlar o acesso ao arquivo. Para cada depósito, você pode controlar o acesso ao depósito (quem pode criar, excluir e listar objetos no depósito), exibir logs de acesso ao depósito e seus objetos e

escolher a região geográfica onde o Amazon S3 armazenará o depósito e seus itens. conteúdo.

Acesso de dados

O acesso aos dados armazenados no Amazon S3 é restrito por padrão; somente proprietários de bucket e objeto têm acesso aos recursos do Amazon S3 que eles criam. (Observe que o proprietário do bloco / objeto é o proprietário da conta da AWS, não o usuário que criou o bloco / objeto.) Existem várias maneiras de controlar o acesso a buckets e objetos:

Políticas do IAM

O AWS IAM permite que organizações com muitos funcionários criem e gerenciem vários usuários em uma única conta da AWS. As políticas do IAM são anexadas aos usuários, permitindo o controle centralizado das permissões dos usuários na sua conta da AWS para acessar buckets ou objetos. Com as políticas do IAM, você só pode conceder usuários em sua própria conta da AWS com permissão para acessar seus recursos do Amazon S3.

ACLs

No Amazon S3, você pode usar ACLs para fornecer acesso de leitura ou gravação em buckets ou objetos a grupos de usuários. Com as ACLs, você só pode conceder a outras contas da AWS (usuários não específicos) acesso aos seus recursos do Amazon S3.

Políticas de bucket

As políticas de bucket no Amazon S3 podem ser usadas para adicionar ou negar permissões em alguns ou todos os objetos em um único bucket. As políticas podem ser anexadas a usuários, grupos ou buckets do Amazon S3, permitindo o gerenciamento centralizado de permissões. Com as políticas de bucket, você pode

conceder aos usuários da sua conta da AWS ou de outras contas da AWS acesso aos recursos do Amazon S3.

Autenticação de string de consulta

Você pode usar uma string de consulta para expressar uma solicitação inteiramente em uma URL. Nesse caso, você usa parâmetros de consulta para fornecer informações de solicitação, incluindo as informações de autenticação. Como a assinatura da solicitação faz parte da URL, esse tipo de URL geralmente é chamado de URL pré-assinado. Você pode usar URLs pré-assinados para incorporar links clicáveis, que podem ser válidos por até sete dias, em HTML.

Você pode restringir ainda mais o acesso a recursos específicos com base em determinadas condições. Por exemplo, você pode restringir o acesso com base no horário da solicitação (Data Condição), se a solicitação foi enviada usando SSL (Condições Booleanas), no endereço IP de um solicitante (Condição de Endereço IP) ou no aplicativo cliente do solicitante (Condições da String). Para identificar essas condições, você usa chaves de política.

O Amazon S3 também oferece aos desenvolvedores a opção de usar a autenticação de string de consulta, o que permite compartilhar objetos do Amazon S3 por meio de URLs válidas por um período predefinido.

A autenticação de cadeia de consulta é útil para fornecer ao HTTP acesso do navegador a recursos que normalmente exigiriam autenticação. A assinatura na cadeia de consulta protege a solicitação.

Transferência de dados

Para segurança máxima, você pode fazer upload / download de dados com segurança para o Amazon S3 através dos endpoint criptografados por SSL. Os endpoint criptografados são acessíveis na Internet e no Amazon EC2, para que os dados sejam transferidos com segurança na AWS e para e de fontes fora da AWS.

Armazenamento de dados

O Amazon S3 fornece várias opções para proteger os dados em repouso. Para os clientes que preferem gerenciar sua própria criptografia, eles podem usar uma biblioteca de criptografia de cliente como o Amazon S3 Encryption Client para criptografar dados antes de fazer o upload para o Amazon S3.

Como alternativa, você pode usar o SSE (Amazon S3 Server Side Encryption) se preferir que o Amazon S3 gerencie o processo de criptografia para você. Os dados são criptografados com uma chave gerada pela AWS ou com uma chave fornecida, dependendo dos seus requisitos.

Com o Amazon S3 SSE, você pode criptografar dados no upload simplesmente adicionando um cabeçalho de solicitação adicional ao gravar o objeto.

A descriptografia acontece automaticamente quando os dados são recuperados. Observe que os metadados, que você pode incluir no seu objeto, não são criptografados.

O Amazon S3 SSE usa uma das cifras de bloco mais fortes disponíveis: AES-256. Com o Amazon S3 SSE, todos os objetos protegidos são criptografados com uma chave de criptografia exclusiva.

Essa chave de objeto é criptografada com uma chave mestre rotacionada regularmente. O Amazon S3 SSE fornece segurança adicional armazenando os dados criptografados e as chaves de criptografia em diferentes hosts.

O Amazon S3 SSE também possibilita a imposição de requisitos de criptografia.

Por exemplo, você pode criar e aplicar políticas de buckets que exigem que apenas dados criptografados possam ser carregados em seus buckets.

Quando um objeto é excluído do Amazon S3, a remoção do mapeamento do nome público para o objeto inicia imediatamente e geralmente é processada no sistema distribuído em alguns segundos.

Após a remoção do mapeamento, não há acesso remoto ao objeto excluído. A área de armazenamento subjacente é então recuperada para uso pelo sistema.

O Amazon S3 Standard foi projetado para fornecer 99,9999999999% de durabilidade dos objetos em um determinado ano. Esse nível de durabilidade corresponde a uma perda média anual esperada de 0,000000001 por cento dos objetos. Por exemplo, se você armazenar 10.000 objetos no Amazon S3, poderá esperar, em média, a perda de um único objeto a cada 10.000.000 anos.

Além disso, o Amazon S3 foi projetado para sustentar a perda simultânea de dados em duas instalações.

Logs de acesso

Um bucket do Amazon S3 pode ser configurado para registrar o acesso ao bucket e aos objetos dentro dele. O log de acesso contém detalhes sobre cada solicitação de acesso, incluindo o tipo de solicitação, o recurso solicitado, o IP do solicitante e a hora e data da solicitação.

Quando o log é ativado para um bucket, os registros são periodicamente agregados aos arquivos de log e entregues no bucket especificado do Amazon S3.

Compartilhamento de recursos entre origens (CORS)

Os clientes da AWS que usam o Amazon S3 para hospedar páginas da Web estáticas ou armazenar objetos usados por outras páginas da Web podem carregar conteúdo com segurança, configurando um bucket do Amazon S3 para ativar explicitamente solicitações de origem cruzada.

Navegadores modernos usam a política Same Origin para bloquear JavaScript ou HTML5 permite que solicitações carreguem conteúdo de outro site ou domínio como uma maneira de garantir que o conteúdo malicioso não seja carregado de uma fonte menos respeitável (como durante ataques de script entre sites).

Com a política de compartilhamento de recursos de origem cruzada (CORS) ativada, ativos como fontes da web e imagens armazenadas em um bucket do Amazon S3 podem ser referenciados com segurança por páginas da web externas, folhas de estilo e aplicativos HTML5.

Amazon Glacier Security

Como o Amazon S3, o serviço Amazon Glacier fornece armazenamento de baixo custo, seguro e durável.

Onde o Amazon S3 foi projetado para recuperação rápida, no entanto, o Amazon Glacier deve ser usado como um serviço de arquivamento de dados que não são acessados com frequência e para os quais os tempos de recuperação de várias horas são adequados.

O Amazon Glacier armazena arquivos como arquivos dentro de cofres. Os arquivos podem ser quaisquer dados, como uma foto, vídeo ou documento, e podem conter um ou vários arquivos.

Você pode armazenar um número ilimitado de arquivos em um único cofre e criar até 1.000 cofres por região. Cada arquivo pode conter até 40 TB de dados.

Transferência de dados

Para segurança máxima, você pode fazer upload / download de dados com segurança no Amazon Glacier por meio dos endpoint criptografados SSL. Os endpoint criptografados são acessíveis na Internet e no Amazon EC2, para que os dados sejam transferidos com segurança na AWS e para e de fontes externas à AWS.

Recuperação de dados

A recuperação de arquivos do Amazon Glacier requer o início de um trabalho de recuperação, que geralmente é concluído em três a cinco horas. Você pode acessar os dados por meio de solicitações HTTP GET. Os dados permanecerão disponíveis para você por 24 horas.

Você pode recuperar um arquivo inteiro ou vários arquivos de um arquivo. Se você deseja recuperar apenas um subconjunto de um archive, pode usar uma solicitação de recuperação para especificar o intervalo do archive que contém os arquivos nos

quais está interessado ou pode iniciar várias solicitações de recuperação, cada uma com um intervalo para um ou mais arquivos.

Você também pode limitar o número de itens de inventário do Vault recuperados filtrando um período de criação de arquivo morto ou definindo um limite máximo de itens.

Seja qual for o método escolhido, ao recuperar partes do arquivo morto, você pode usar a soma de verificação fornecida para ajudar a garantir a integridade dos arquivos, desde que o intervalo recuperado esteja alinhado com o hash da árvore do arquivo morto geral.

Armazenamento de dados

O Amazon Glacier criptografa automaticamente os dados usando o AES-256 e os armazena de forma durável de forma imutável. O Amazon Glacier foi projetado para fornecer durabilidade média anual de 99,9999999999% para um arquivo morto.

Ele armazena cada arquivo em várias instalações e vários dispositivos. Diferentemente dos sistemas tradicionais, que podem exigir verificação de dados trabalhosa e reparo manual, o Amazon Glacier realiza verificações sistemáticas e regulares da integridade dos dados e foi desenvolvido para ser auto-reparável.

Acesso de dados

Somente sua conta pode acessar seus dados no Amazon Glacier. Para controlar o acesso aos seus dados no Amazon Glacier, você pode usar o AWS IAM para especificar quais usuários da sua conta têm direitos para operações em um determinado cofre.

Segurança do AWS Storage Gateway

O serviço AWS Storage Gateway conecta seu dispositivo de software local ao armazenamento baseado em nuvem para fornecer integração perfeita e segura entre seu ambiente de TI e a infraestrutura de armazenamento da AWS.

O serviço permite fazer upload de dados com segurança para o AWS escalável, serviço de armazenamento confiável e seguro do Amazon S3 para backup econômico e recuperação rápida de desastres.

Transferência de dados

Os dados são transferidos de forma assíncrona do seu hardware de armazenamento local para a AWS sobre SSL.

Armazenamento de dados

Os dados são armazenados criptografados no Amazon S3 usando o AES 256, um padrão de criptografia de chave simétrica usando chaves de criptografia de 256 bits. O AWS Storage Gateway apenas carrega dados que foram alterados, minimizando a quantidade de dados enviados pela Internet.

Base de dados

A AWS fornece várias soluções de banco de dados para desenvolvedores e empresas, desde serviços gerenciados de banco de dados relacional e NoSQL a cache de memória como serviço e serviço de armazém de dados em escala de petabytes.

Segurança do Amazon DynamoDB

O Amazon DynamoDB é um serviço de banco de dados NoSQL gerenciado que fornece desempenho rápido e previsível com escalabilidade perfeita. O Amazon DynamoDB permite descarregar os encargos administrativos de operação e dimensionamento de bancos de dados distribuídos para a AWS, para que você não precisa se preocupar com provisionamento de hardware, instalação e configuração, replicação, aplicação de patches de software ou dimensionamento de cluster.

Você pode criar uma tabela de banco de dados que possa armazenar e recuperar qualquer quantidade de dados e atender a qualquer nível de tráfego de solicitação. O Amazon DynamoDB espalha automaticamente os dados e o tráfego para a tabela sobre um número suficiente de servidores para lidar com a capacidade de solicitação especificada e a quantidade de dados armazenados, mantendo um desempenho rápido e consistente.

Todos os itens de dados são armazenados em unidades de estado sólido (SSDs) e são replicados automaticamente em várias zonas de disponibilidade em uma região para fornecer alta disponibilidade e durabilidade de dados.

Você pode configurar backups automáticos usando um modelo especial no AWS Data Pipeline que foi criado apenas para copiar tabelas do Amazon DynamoDB. Você pode escolher backups completos ou incrementais para uma tabela na mesma região ou em uma região diferente.

Você pode usar a cópia para recuperação de desastre no caso de um erro no seu código danificar a tabela original ou federar dados do Amazon DynamoDB entre regiões para oferecer suporte a um aplicativo com várias regiões.

Para controlar quem pode usar os recursos e a API do Amazon DynamoDB, configure as permissões no AWS IAM. Além de controlar o acesso no nível do recurso com o IAM, você também pode controlar o acesso no nível do banco de dados - você pode criar permissões no nível do banco de dados que permitem ou

negam o acesso a itens (linhas) e atributos (colunas) com base nas necessidades de sua aplicação.

Essas permissões no nível do banco de dados são chamadas de controles de acesso refinados, e você as cria usando uma política do IAM que especifica sob quais circunstâncias um usuário ou aplicativo pode acessar uma tabela do Amazon DynamoDB. A política do IAM pode restringir o acesso a itens individuais em uma tabela, o acesso aos atributos nesses itens ou a ambos ao mesmo tempo.

Além de exigir permissões de banco de dados e de usuário, cada solicitação ao serviço Amazon DynamoDB deve conter uma assinatura HMAC-SHA-256 válida ou a solicitação é rejeitada.

Os AWS SDKs assinam automaticamente suas solicitações; no entanto, se você quiser escrever suas próprias solicitações HTTP POST, deverá fornecer a assinatura no cabeçalho da sua solicitação ao Amazon DynamoDB. Para calcular a assinatura, você deve solicitar credenciais de segurança temporárias do AWS Security Token Service.

Use as credenciais de segurança temporárias para assinar suas solicitações no Amazon DynamoDB. O Amazon DynamoDB pode ser acessado por terminais criptografados em SSL, e os terminais criptografados podem ser acessados na Internet e no Amazon EC2.

Segurança do Amazon RDS

O Amazon Relational Database Service (Amazon RDS) permite criar rapidamente uma Instância de banco de dados relacional (Instância de banco de dados) e escalar com flexibilidade os recursos de computação associados e a capacidade de armazenamento para atender à demanda de aplicativos.

O Amazon RDS gerencia a instância do banco de dados em seu nome, executando backups, manipulando o failover e mantendo o software do banco de dados.

Até o momento em que este artigo foi escrito, o Amazon RDS estava disponível para os mecanismos de banco de dados MySQL, Oracle, Microsoft SQL Server, MariaDB, Amazon Aurora e PostgreSQL.

O Amazon RDS possui vários recursos que aprimoram a confiabilidade de bancos de dados críticos de produção, incluindo grupos de segurança de banco de dados, permissões, conexões SSL, backups automatizados, snapshots de banco de dados e várias implantações da Zona de Disponibilidade (Multi-AZ).

As instâncias de banco de dados também podem ser implantadas em um Amazon VPC para isolamento adicional da rede.

Controle de acesso

Quando você cria uma Instância de banco de dados pela primeira vez no Amazon RDS, cria uma conta de usuário principal, que é usada apenas no contexto do Amazon RDS para controlar o acesso às suas Instâncias de banco de dados.

A conta de usuário principal é uma conta de usuário nativa do banco de dados que permite fazer logon na sua Instância de Banco de Dados com todos os privilégios do banco de dados.

Você pode especificar o nome de usuário mestre e a senha que deseja associar a cada Instância de banco de dados ao criar a Instância de banco de dados. Depois de criar sua Instância de banco de dados, você pode se conectar ao banco de dados usando as credenciais de usuário principal. Posteriormente, você pode criar contas de usuário adicionais para restringir quem pode acessar sua instância de banco de dados.

Você pode controlar o acesso à instância do Amazon RDS DB via grupos de segurança do DB, que são semelhantes aos grupos de segurança do Amazon EC2, mas não são intercambiáveis. Os grupos de segurança do banco de dados agem como um firewall que controla o acesso da rede à sua instância de banco de dados. Os grupos de segurança do banco de dados são padrão para negar todo o modo de acesso, e os clientes devem autorizar especificamente a entrada na rede.

Há duas maneiras de fazer isso:

- Autorizando um intervalo de IP de rede
- Autorizando um grupo de segurança existente do Amazon EC2

Os grupos de segurança do banco de dados permitem apenas o acesso à porta do servidor de banco de dados (todos os outros estão bloqueados) e podem ser atualizados sem reiniciar a Instância de banco de dados do Amazon RDS, o que fornece controle contínuo do acesso ao banco de dados.

Usando o AWS IAM, você pode controlar ainda mais o acesso às suas instâncias do Amazon RDS DB. O AWS IAM permite controlar as operações do Amazon RDS que cada usuário do AWS IAM tem permissão para chamar.

Isolamento de rede

Para controle de acesso à rede adicional, você pode executar suas instâncias de banco de dados em um Amazon VPC. O Amazon VPC permite isolar suas instâncias de banco de dados especificando o intervalo de IPs que você deseja usar e se conectar à sua infraestrutura de TI existente por meio da VPN IPsec criptografada padrão do setor.

A execução do Amazon RDS em uma VPC permite que você tenha uma instância de banco de dados em uma sub-rede privada. Você também pode configurar um gateway privado virtual que estenda sua rede corporativa à sua VPC e permita acesso à instância do RDS DB nessa VPC.

Para implantações Multi-AZ, a definição de uma sub-rede para todas as zonas de disponibilidade em uma região permitirá que o Amazon RDS crie um novo modo de espera em outra zona de disponibilidade, se necessário. Você pode criar grupos de sub-rede de banco de dados, que são coleções de sub-redes que você pode designar para suas instâncias de banco de dados do Amazon RDS em um Amazon VPC.

Cada grupo de sub-rede de banco de dados deve ter pelo menos uma sub-rede para cada zona de disponibilidade em uma determinada região. Nesse caso, quando

você cria uma instância de banco de dados em um Amazon VPC, você seleciona um grupo de sub-rede de banco de dados; O Amazon RDS usa esse grupo de sub-redes do banco de dados e sua Zona de disponibilidade preferida para selecionar uma sub-rede e um endereço IP dentro dessa sub-rede.

O Amazon RDS cria e associa uma interface de rede elástica à sua instância de banco de dados com esse endereço IP.

As instâncias de banco de dados implantadas em um Amazon VPC podem ser acessadas da Internet ou de instâncias do Amazon EC2 fora do Amazon VPC por meio de hosts VPN ou bastiões que você pode iniciar em sua sub-rede pública.

Para usar um host bastião, você precisará configurar uma sub-rede pública com uma instância do Amazon EC2 que atue como bastião SSH. Essa sub-rede pública deve ter um gateway da Internet e regras de roteamento que permitam direcionar o tráfego através do host SSH, que deve encaminhar solicitações para o endereço IP privado da sua instância do Amazon RDS DB.

Grupos de segurança de banco de dados podem ser usados para ajudar a proteger instâncias de banco de dados dentro de um Amazon VPC. Além disso, o tráfego de rede que entra e sai de cada sub-rede pode ser permitido ou negado por meio de ACLs da rede. Todo o tráfego de rede que entra ou sai do Amazon VPC por meio da conexão VPN IPsec pode ser inspecionado pela infraestrutura de segurança local, incluindo firewalls de rede e sistemas de detecção de intrusão.

Criptografia

Você pode criptografar conexões entre seu aplicativo e sua Instância de banco de dados usando SSL. Para MySQL e SQL Server, o Amazon RDS cria um certificado SSL e instala o certificado na instância do banco de dados quando a instância é provisionada.

Para o MySQL, você inicia o cliente MySQL usando o parâmetro --ssl_ca para referenciar a chave pública para criptografar as conexões. Para o SQL Server, baixe a chave pública e importe o certificado para o sistema operacional Windows. O

Oracle RDS usa criptografia de rede nativa Oracle com uma instância de banco de dados.

Você simplesmente adiciona a opção de criptografia de rede nativa a um grupo de opções e associa esse grupo de opções à instância do banco de dados. Depois que uma conexão criptografada é estabelecida, os dados transferidos entre a Instância do banco de dados e seu aplicativo serão criptografados durante a transferência.

Você também pode exigir que sua Instância de banco de dados aceite apenas conexões criptografadas.

O Amazon RDS suporta criptografia de dados transparente (TDE) para SQL Server (SQL Server Enterprise Edition) e Oracle (parte da opção Oracle Advanced Security disponível no Oracle Enterprise Edition).

O recurso TDE criptografa automaticamente os dados antes de serem gravados para armazenamento e descriptografa automaticamente os dados quando são lidos do armazenamento. Se você precisar que seus dados MySQL sejam criptografados enquanto estiver descansando no banco de dados, seu aplicativo deverá gerenciar a criptografia e descriptografia de dados.

Observe que o suporte a SSL no Amazon RDS é para criptografar a conexão entre seu aplicativo e sua instância de banco de dados; não deve ser invocado para autenticar a própria instância do banco de dados. Embora o SSL ofereça benefícios de segurança, lembre-se de que a criptografia SSL é uma operação intensiva em computação e aumentará a latência da sua conexão com o banco de dados.

Backups automatizados e snapshots de banco de dados O Amazon RDS fornece dois métodos diferentes para fazer backup e restaurar suas instâncias de banco de dados: backups automatizados e snapshots de banco de dados (snapshots de banco de dados). Ativado por padrão, o recurso de backup automatizado do Amazon RDS permite a recuperação point-in-time para sua Instância de banco de dados.

O Amazon RDS fará backup do banco de dados e dos logs de transações e armazenará ambos por um período de retenção especificado pelo usuário. Isso permite restaurar a instância do banco de dados a qualquer segundo durante o período de retenção, até os últimos cinco minutos.

Seu período de retenção de backup automático pode ser configurado para até 35 dias. Snapshots de banco de dados são backups iniciados pelo usuário da sua instância de banco de dados.

Esses backups completos do banco de dados são armazenados pelo Amazon RDS até que você os exclua explicitamente. Você pode copiar snapshots de banco de dados de qualquer tamanho e movê-los entre qualquer uma das regiões públicas da AWS ou copiar o mesmo snapshot para várias regiões simultaneamente. Você pode criar uma nova instância de banco de dados a partir de um snapshot de banco de dados sempre que desejar.

Durante a janela de backup, a E / S de armazenamento pode ser suspensa enquanto o backup dos dados está sendo feito. Essa suspensão de E / S normalmente dura alguns minutos. Essa suspensão de E / S é evitada nas implantações do Multi-AZ DB, porque o backup é retirado do modo de espera.

Replicação de Instância de Banco de Dados

Os recursos de computação em nuvem da AWS estão alojados em instalações de data center altamente disponíveis em diferentes regiões do mundo, e cada região contém vários locais distintos chamados Zonas de Disponibilidade.

Cada zona de disponibilidade é projetada para se isolar de falhas em outras zonas de disponibilidade e fornecer uma rede de baixo custo e baixa latência de conectividade com outras zonas de disponibilidade na mesma região.

Para projetar a alta disponibilidade de seus bancos de dados Oracle, PostgreSQL ou MySQL, você pode executar a instância do Amazon RDS DB em várias zonas de disponibilidade, uma opção chamada implantação Multi-AZ.

Quando você seleciona essa opção, a AWS provisiona e mantém automaticamente uma réplica síncrona em espera da sua Instância de banco de dados em uma zona de disponibilidade diferente.

A instância de banco de dados principal é replicada de forma síncrona nas zonas de disponibilidade para a réplica em espera. No caso de falha da instância do banco

de dados ou da zona de disponibilidade, o Amazon RDS fará failover automaticamente no modo de espera, para que as operações do banco de dados possam ser retomadas rapidamente sem intervenção administrativa.

Para clientes que usam o MySQL e precisam escalar além das restrições de capacidade de uma única instância de banco de dados para cargas de trabalho de banco de dados com muita leitura, o Amazon RDS fornece uma opção de réplica de leitura.

Depois de criar uma réplica de leitura, as atualizações do banco de dados na Instância de banco de dados de origem são replicadas para a réplica de leitura usando a replicação assíncrona nativa do MySQL.

Você pode criar várias réplicas de leitura para uma determinada instância de banco de dados de origem e distribuir o tráfego de leitura do seu aplicativo entre elas. As réplicas de leitura podem ser criadas com implantações Multi-AZ para obter benefícios de escala de leitura, além da disponibilidade aprimorada de gravação no banco de dados e durabilidade dos dados fornecidos pelas implantações Multi-AZ.

Correção automática de software

O Amazon RDS garantirá que o software de banco de dados relacional que alimenta sua implantação permaneça atualizado com os patches mais recentes.

Quando necessário, os patches são aplicados durante uma janela de manutenção que você pode controlar. Você pode pensar na janela de manutenção do Amazon RDS como uma oportunidade de controlar quando ocorrem modificações na Instância do banco de dados (como a classe de instância do banco de dados de escala) e correções de software, em que o evento é solicitado ou necessário.

Se um evento de manutenção for agendado para uma determinada semana, ele será iniciado e concluído em algum momento durante a janela de manutenção de 30 minutos que você identificar.

Os únicos eventos de manutenção que exigem que o Amazon RDS coloque sua Instância de banco de dados offline são operações de computação em escala (que

geralmente levam apenas alguns minutos do início ao fim) ou aplicação de patches de software.

O patch necessário é agendado automaticamente apenas para patches relacionados à segurança e durabilidade. Essas correções ocorrem com pouca frequência (geralmente uma vez a cada poucos meses) e raramente exigem mais do que uma fração da sua janela de manutenção.

Se você não especificar uma janela de manutenção semanal preferida ao criar sua Instância de banco de dados, um valor padrão de 30 minutos será atribuído. Se você deseja modificar quando a manutenção é executada em seu nome, você pode fazê-lo modificando sua Instância de banco de dados no AWS Management Console ou usando a API ModifyDBInstance.

Cada uma das suas instâncias de banco de dados pode ter diferentes janelas de manutenção preferenciais, se você escolher.

A execução da sua Instância de banco de dados em uma implantação Multi-AZ pode reduzir ainda mais o impacto de um evento de manutenção, pois o Amazon RDS realizará a manutenção através das seguintes etapas:

1. Execute a manutenção no modo de espera.
2. Promova o modo de espera para o primário.
3. Execute a manutenção no primário antigo, que se torna o novo modo de espera.

Quando uma API de exclusão da instância do Amazon RDS DB (DeleteDBInstance) é executada, a instância do banco de dados é marcada para exclusão. Depois que a instância não indica mais o status de exclusão, ela foi removida. Nesse momento, a instância não está mais acessível e, a menos que uma cópia final do snapshot tenha sido solicitada, ela não poderá ser restaurada e não será listada por nenhuma das ferramentas ou APIs.

Amazon Redshift Security

O Amazon Redshift é um serviço de data warehouse SQL em escala de petabytes que é executado em recursos de computação e armazenamento altamente otimizados e gerenciados da AWS.

O serviço foi arquitetado não apenas para aumentar ou diminuir rapidamente, mas também para melhorar significativamente as velocidades de consulta, mesmo em conjuntos de dados extremamente grandes.

Para aumentar o desempenho, o Amazon Redshift usa técnicas como armazenamento colunar, compactação de dados e mapas de zona para reduzir a quantidade de E / S necessária para executar consultas. Ele também possui uma arquitetura MPP (Massively Parallel Processing), paralelizando e distribuindo operações SQL para aproveitar todos os recursos disponíveis.

Acesso ao Cluster

Por padrão, os clusters que você cria são fechados para todos. O Amazon Redshift permite configurar regras de firewall (grupos de segurança) para controlar o acesso de rede ao cluster de data warehouse. Você também pode executar o Amazon Redshift dentro de um Amazon VPC para isolar o cluster de data warehouse em sua própria rede virtual e conectá-lo à sua infraestrutura de TI existente usando a VPN IPsec criptografada padrão do setor.

A conta da AWS que cria o cluster tem acesso total ao cluster. Na sua conta da AWS, você pode usar o AWS IAM para criar contas de usuário e gerenciar permissões para essas contas. Ao usar o IAM, você pode conceder permissão a diferentes usuários para executar apenas as operações de cluster necessárias para o trabalho delas.

Como todos os bancos de dados, você deve conceder permissão no Amazon Redshift no nível do banco de dados, além de conceder acesso no nível do recurso.

Os usuários do banco de dados são denominados contas de usuário que podem se conectar a um banco de dados e são autenticadas quando efetuam login no

Amazon Redshift. No Amazon Redshift, você concede permissões de usuário de banco de dados por cluster, em vez de por tabela.

No entanto, os usuários podem ver dados apenas nas linhas da tabela que foram geradas por suas próprias atividades; linhas geradas por outros

os usuários não são visíveis para eles.

O usuário que cria um objeto de banco de dados é seu proprietário. Por padrão, apenas um superusuário ou o proprietário de um objeto pode consultar, modificar ou conceder permissões ao objeto. Para que os usuários usem um objeto, você deve conceder as permissões necessárias ao usuário ou ao grupo que contém o

do utilizador. Além disso, apenas o proprietário de um objeto pode modificá-lo ou excluí-lo.

Backups de dados

O Amazon Redshift distribui seus dados por todos os nós de computação em um cluster. Quando você executa um cluster com pelo menos dois nós de computação, os dados em cada nó sempre serão espelhados em discos em outro nó, reduzindo o risco de perda de dados.

Além disso, é feito backup contínuo de todos os dados gravados em um nó do cluster no Amazon S3 usando snapshots. O Amazon Redshift armazena seus snapshots por um período definido pelo usuário, que pode ser de 1 a 35 dias.

Você também pode tirar suas próprias capturas instantâneas a qualquer momento; esses snapshots aproveitam todos os snapshots do sistema existentes e são mantidos até que você os exclua explicitamente.

O Amazon Redshift monitora continuamente a integridade do cluster e replica automaticamente os dados de unidades com falha e substitui os nós conforme necessário. Tudo isso acontece sem nenhum esforço de sua parte, embora você

possa observar uma ligeira degradação do desempenho durante o processo de replicação.

Você pode usar qualquer snapshot do sistema ou do usuário para restaurar seu cluster usando o AWS Management Console ou as APIs do Amazon Redshift.

Seu cluster estará disponível assim que os metadados do sistema forem restaurados e você poderá iniciar a execução de consultas enquanto os dados do usuário estão em spool.

Criptografia de Dados

Ao criar um cluster, você pode optar por criptografá-lo para fornecer proteção adicional aos seus dados em repouso. Quando você ativa a criptografia em seu cluster, o Amazon Redshift armazena todos os dados em tabelas criadas pelo usuário em um formato criptografado usando chaves de criptografia de bloco AES-256 aceleradas por hardware. Isso inclui todos os dados gravados no disco e todos os backups.

O Amazon Redshift usa uma arquitetura baseada em chave de quatro camadas para criptografia. Essas chaves consistem em chaves de criptografia de dados, uma chave de banco de dados, uma chave de cluster e uma chave mestra.

As chaves de criptografia de dados criptografam os blocos de dados no cluster. Cada bloco de dados recebe uma chave AES256 gerada aleatoriamente. Essas chaves são criptografadas usando a chave do banco de dados do cluster.

A chave do banco de dados criptografa as chaves de criptografia de dados no cluster. A chave do banco de dados é uma chave AES-256 gerada aleatoriamente. Ele é armazenado em disco em uma rede separada do cluster Amazon Redshift e criptografado por uma chave mestra. O Amazon Redshift passa a chave do banco de dados por um canal seguro e a mantém na memória no cluster.

A chave do cluster criptografa a chave do banco de dados do cluster Amazon Redshift. Você pode usar a AWS ou um HSM (Hardware Security Module) para armazenar a chave do cluster.

Os HSMs fornecem controle direto da geração e gerenciamento de chaves e tornam o gerenciamento de chaves separado e distinto do aplicativo e do banco de dados.

A chave mestra criptografa a chave do cluster se estiver armazenada na AWS. A chave mestra criptografa a chave de banco de dados criptografada pela chave do cluster se a chave do cluster estiver armazenada em um HSM.

Você pode fazer com que o Amazon Redshift gire as chaves de criptografia dos seus clusters criptografados a qualquer momento. Como parte do processo de rotação, as chaves também são atualizadas para todos os snapshots automáticos e manuais do cluster.

Observe que a ativação da criptografia em seu cluster afetará o desempenho, mesmo que seja acelerado por hardware.

A criptografia também se aplica aos backups. Quando você estiver restaurando a partir de um snapshot criptografado, o novo cluster também será criptografado.

Para criptografar sua tabela, carregue os arquivos de dados ao carregá-los no Amazon S3, você pode usar a criptografia no servidor do Amazon S3. Quando você carrega os dados do Amazon S3, o comando COPY descriptografa os dados à medida que carrega a tabela.

Log de auditoria de banco de dados

O Amazon Redshift registra todas as operações SQL, incluindo tentativas de conexão, consultas e alterações no seu banco de dados. Você pode acessar esses logs usando consultas SQL em tabelas do sistema ou optar por fazer o download para um bucket seguro do Amazon S3. Em seguida, você pode usar esses logs de auditoria para monitorar seu cluster para fins de segurança e solução de problemas.

Correção automática de software

O Amazon Redshift gerencia todo o trabalho de configurar, operar e dimensionar seu data warehouse, incluindo capacidade de provisionamento, monitoramento do

cluster e aplicação de patches e atualizações no mecanismo Amazon Redshift. Os patches são aplicados apenas durante as janelas de manutenção especificadas.

Conexões SSL

Para proteger seus dados em trânsito na nuvem da AWS, Amazon Redshift usa SSL acelerado por hardware para se comunicar com o Amazon S3 ou o Amazon DynamoDB para operações de COPY, UNLOAD, backup e restauração. Você pode criptografar a conexão entre seu cliente e o cluster especificando SSL no grupo de parâmetros associado ao cluster.

Para que seus clientes também autentiquem o servidor Amazon Redshift, você pode instalar a chave pública (arquivo .pem) do certificado SSL em seu cliente e usar a chave para conectar-se aos seus clusters.

O Amazon Redshift oferece os conjuntos de cifras mais novos e mais fortes que usam o protocolo Elliptic Curve Diffie-Hellman Ephemeral (ECDHE). O ECDHE permite que os clientes SSL forneçam o Perfect Forward Secrecy entre o cliente e o cluster do Amazon Redshift.

O Perfect Forward Secrecy usa chaves de sessão efêmeras e não armazenadas em qualquer lugar, o que impede a decodificação de dados capturados por terceiros não autorizados, mesmo se a própria chave secreta de longo prazo estiver comprometida.

Você não precisa configurar nada no Amazon Redshift para ativar o ECDHE. Se você se conectar a partir de uma ferramenta de cliente SQL que usa ECDHE para criptografar a comunicação entre o cliente e o servidor, o Amazon Redshift usará a lista de criptografia fornecida para estabelecer a conexão apropriada.

Amazon ElastiCache Security

O Amazon ElastiCache é um serviço da Web que facilita a configuração, o gerenciamento e a escalabilidade de ambientes de cache na memória distribuídos na nuvem.

O serviço melhora o desempenho dos aplicativos da Web, permitindo recuperar informações de um sistema de cache em memória gerenciado e rápido, em vez de depender inteiramente de um disco mais lento, baseado em disco de bancos de dados.

Ele pode ser usado para melhorar significativamente a latência e a taxa de transferência para muitas cargas de trabalho de aplicativos pesados (como redes sociais, jogos, compartilhamento de mídia e portais de perguntas e respostas) ou cargas de trabalho intensivas em computação (como um mecanismo de recomendação).

O armazenamento em cache melhora o desempenho do aplicativo, armazenando dados críticos na memória para acesso de baixa latência. As informações em cache podem incluir os resultados de consultas de banco de dados intensivas em E / S ou os resultados de cálculos computacionalmente intensivos.

O serviço Amazon ElastiCache automatiza tarefas de gerenciamento demoradas para ambientes de cache de memória, como gerenciamento de patches, detecção de falhas e recuperação. Funciona em conjunto com outros serviços da AWS Cloud (como Amazon EC2, Amazon CloudWatch e Amazon SNS) para fornecer um cache de memória seguro, de alto desempenho e gerenciado.

Por exemplo, um aplicativo em execução no Amazon EC2 pode acessar com segurança um cluster do Amazon ElastiCache na mesma região com latência muito baixa.

Usando o serviço Amazon ElastiCache, você cria um cluster de cache, que é uma coleção de um ou mais nós de cache, cada um executando uma instância do serviço Memcached. Um nó de cache é um pedaço de tamanho fixo de RAM segura e conectada à rede. Cada nó de cache executa uma instância do serviço Memcached e possui seu próprio nome DNS e porta. Vários tipos de nós de cache são suportados, cada um com quantidades variáveis de memória associada.

Um cluster de cache pode ser configurado com um número específico de nós de cache e um grupo de parâmetros de cache que controla as propriedades de cada nó de cache. Todos os nós de cache em um cluster de cache foram projetados para serem do mesmo tipo de nó e terem as mesmas configurações de parâmetro e grupo de segurança.

Acesso de dados

O Amazon ElastiCache permite controlar o acesso aos seus Clusters de cache usando grupos de segurança de cache. Um grupo de segurança de cache age como um firewall, controlando o acesso da rede ao seu cluster de cache. Por padrão, o acesso à rede está desativado nos seus Clusters de cache.

Se você deseja que seus aplicativos acessem seu cluster de cache, ative explicitamente o acesso de hosts em grupos de segurança específicos do Amazon EC2. Após a configuração das regras de entrada, as mesmas regras se aplicam a todos os Clusters de cache associados a esse grupo de segurança de cache.

Para permitir o acesso da rede ao seu cluster de cache, crie um grupo de segurança de cache e use a API ou o comando CLI de entrada de grupo de segurança de cache de autorização para autorizar o grupo de segurança desejado do Amazon EC2 (que por sua vez especifica as instâncias do Amazon EC2 permitidas). O controle de acesso baseado em IPrange atualmente não está ativado para Clusters de cache.

Todos os clientes de um cluster de cache devem estar na rede Amazon EC2 e autorizados por meio de grupos de segurança de cache.

O Amazon ElastiCache for Redis fornece funcionalidade de backup e restauração, onde é possível criar uma captura instantânea de todo o cluster Redis, como existe

em um momento específico. Você pode agendar snapshots diários automáticos e recorrentes ou criar um snapshot manual a qualquer momento.

Para capturas instantâneas automáticas, você especifica um período de retenção; os snapshots manuais são mantidos até que você os exclua. Os snapshots são armazenados no Amazon S3 com alta durabilidade e podem ser usados para inicialização a quente, backups e arquivamento.

Serviços de Aplicação

A AWS oferece uma variedade de serviços gerenciados para uso com seus aplicativos, incluindo serviços que fornecem fluxo de aplicativos, enfileiramento, notificação por push, entrega de email, pesquisa e transcodificação.

Segurança do Amazon Simple Queue Service (Amazon SQS)

O Amazon SQS é um serviço escalável e altamente confiável de enfileiramento de mensagens que permite a comunicação assíncrona baseada em mensagens entre os componentes distribuídos de um aplicativo. Os componentes podem ser computadores ou instâncias do Amazon EC2 ou uma combinação de ambos.

Com o Amazon SQS, você pode enviar qualquer número de mensagens para uma fila do Amazon SQS a qualquer momento a partir de qualquer componente. As mensagens podem ser recuperadas do mesmo componente ou de outro, imediatamente ou posteriormente (em 14 dias).

As mensagens são altamente duráveis; cada mensagem é armazenada persistentemente em filas altamente disponíveis e altamente confiáveis. Vários processos podem ler / gravar de / para uma fila do Amazon SQS ao mesmo tempo sem interferir um no outro.

Acesso de dados

O acesso ao Amazon SQS é concedido com base em uma conta da AWS ou em um usuário criado com o AWS IAM. Depois de autenticada, a conta da AWS tem acesso total a todas as operações do usuário. Um usuário do IAM, no entanto, só tem acesso às operações e filas para as quais foram acessos concedidos via política.

Por padrão, o acesso a cada fila individual é restrito à conta da AWS que o criou. No entanto, você pode permitir outro acesso a uma fila, usando uma política gerada pelo Amazon SQS ou uma política que você escreve.

Criptografia

O Amazon SQS pode ser acessado por terminais criptografados em SSL. Os terminais criptografados são acessíveis pela Internet e pelo Amazon EC2. Os dados armazenados no Amazon SQS não são criptografados pela AWS; no entanto, o usuário pode criptografar os dados antes de fazer o upload para o Amazon SQS, desde que o aplicativo que usa a fila tenha um meio de descriptografar a mensagem quando ela for recuperada.

A criptografia de mensagens antes de enviá-las ao Amazon SQS ajuda a proteger contra o acesso a dados confidenciais de clientes por pessoas não autorizadas, incluindo a AWS.

Segurança do Amazon Simple Notification Service (Amazon SNS)

O Amazon SNS é um serviço da Web que facilita a configuração, a operação e o envio de notificações da nuvem. Ele fornece aos desenvolvedores um recurso altamente escalável, flexível e econômico para publicar mensagens de um aplicativo e entregá-las imediatamente para assinantes ou outros aplicativos.

O Amazon SNS fornece uma interface simples de serviços da web que pode ser usado para criar tópicos sobre os quais os clientes desejam notificar aplicativos (ou

pessoas), inscrever clientes nesses tópicos, publicar mensagens e enviar essas mensagens por protocolo de escolha dos clientes (por exemplo, HTTP / HTTPS, email).

O Amazon SNS entrega notificações aos clientes usando um mecanismo de envio que elimina a necessidade de verificar ou buscar novas informações e atualizações periodicamente.

O Amazon SNS pode ser aproveitado para criar fluxos de trabalho e aplicativos de mensagens altamente confiáveis e controlados por eventos, sem a necessidade de gerenciamento complexo de aplicativos e middleware.

Os usos potenciais para o Amazon SNS incluem aplicativos de monitoramento, sistemas de fluxo de trabalho, atualizações de informações sensíveis ao tempo, aplicativos móveis e muitos outros.

Acesso de dados

O Amazon SNS fornece mecanismos de controle de acesso para que tópicos e mensagens sejam protegidos contra acesso não autorizado. Os proprietários do tópico podem definir políticas para um tópico que restrinja quem pode publicar ou assinar um tópico.

Além disso, os proprietários do tópico podem criptografar a transmissão especificando que o mecanismo de entrega deve ser HTTPS.

O acesso ao Amazon SNS é concedido com base em uma conta da AWS ou em um usuário criado com o AWS IAM. Depois de autenticada, a conta da AWS tem acesso total a todas as operações do usuário.

Um usuário do IAM, no entanto, só tem acesso às operações e tópicos aos quais eles receberam acesso via política.

Por padrão, o acesso a cada tópico individual é restrito à conta da AWS que o criou. No entanto, você pode permitir outro acesso ao Amazon SNS, usando uma política gerada pelo Amazon SNS ou uma política que você escreve.

Serviços de análise

A AWS fornece serviços de análise baseados na nuvem para ajudá-lo a processar e analisar qualquer volume de dados, independentemente de sua necessidade de clusters gerenciados do Hadoop, dados de streaming em tempo real, data warehousing em escala de petabytes ou orquestração.

Segurança do Amazon Elastic MapReduce (Amazon EMR)

O Amazon Elastic MapReduce (Amazon EMR) é um serviço da web gerenciado que você pode usar para executar clusters do Hadoop que processam grandes quantidades de dados, distribuindo o trabalho e os dados entre vários servidores. Ele usa uma versão aprimorada da estrutura do Apache Hadoop em execução na infraestrutura de escala da Web do Amazon EC2 e Amazon S3.

Você simplesmente carrega seus dados de entrada e um aplicativo de processamento de dados no Amazon S3. O Amazon EMR inicia o número de instâncias do Amazon EC2 que você especificar. O serviço inicia a execução do fluxo de trabalho enquanto puxa os dados de entrada do Amazon S3 para as instâncias iniciadas do Amazon EC2.

Após a conclusão do fluxo de trabalho, o Amazon EMR transfere os dados de saída para o Amazon S3, onde você pode recuperá-los ou usá-los como entrada em outro fluxo de trabalho.

Ao iniciar fluxos de trabalho em seu nome, o Amazon EMR configura dois grupos de segurança do Amazon EC2: um para os nós principais e outro para os escravos. O grupo de segurança principal possui uma porta aberta para comunicação com o serviço. Ele também possui a porta SSH aberta para permitir o SSH nas instâncias usando a chave especificada na inicialização.

Os escravos iniciam em um grupo de segurança separado, que permite apenas a interação com a instância principal. Por padrão, os dois grupos de segurança são configurados para não permitir o acesso de fontes externas, incluindo instâncias do Amazon EC2 pertencentes a outros clientes. Como esses grupos são de segurança na sua conta, você pode reconfigurá-los usando as ferramentas ou o painel padrão

do EC2. Para proteger os conjuntos de dados de entrada e saída do cliente, o Amazon EMR transfere dados de e para o Amazon S3 usando SSL.

O Amazon EMR fornece várias maneiras de controlar o acesso aos recursos do seu cluster. Você pode usar o AWS IAM para criar contas e funções de usuário e configurar permissões que controlam quais recursos da AWS esses usuários e funções podem acessar. Ao iniciar um cluster, você pode associar um par de chaves do Amazon EC2 ao cluster, que pode ser usado quando você se conecta ao cluster usando SSH.

Você também pode definir permissões que permitam usuários diferentes do padrão de usuário do Hadoop para enviar tarefas ao seu cluster. Por padrão, se um usuário do IAM iniciar um cluster, esse cluster estará oculto de outros usuários do IAM a conta da AWS.

Essa filtragem ocorre em todas as interfaces do Amazon EMR (console de gerenciamento da AWS, CLI, API e SDKs) e ajuda a impedir que os usuários do IAM acessem e alterem inadvertidamente os clusters criados por outros usuários do IAM.

Para uma camada adicional de proteção, você pode iniciar as instâncias do Amazon EC2 do cluster do Amazon EMR em um Amazon VPC, o que é como iniciá-lo em uma sub-rede privada.

Isso permite que você controle o acesso a toda a sub-rede. Você também pode iniciar o cluster em um Amazon VPC e permitir que o cluster acesse recursos em sua rede interna usando uma conexão VPN.

Você pode criptografar os dados de entrada antes de carregá-los no Amazon S3 usando qualquer ferramenta comum de criptografia de dados. Se você criptografar os dados antes do upload, precisará adicionar uma etapa de descriptografia ao início do seu fluxo de trabalho quando o Amazon EMR buscar os dados do Amazon S3.

Segurança do Amazon Kinesis

O Amazon Kinesis é um serviço gerenciado projetado para lidar com o streaming em tempo real de big data. Ele pode aceitar qualquer quantidade de dados, de qualquer número de fontes, aumentando e diminuindo conforme necessário.

Você pode usar o Amazon Kinesis em situações que exigem ingestão e processamento em larga escala de dados em tempo real, como logs de servidor, mídias sociais ou feeds de dados de mercado e dados de fluxo de cliques na Web.

Os aplicativos leem e gravam registros de dados no Amazon Kinesis em fluxos. Você pode criar qualquer número de fluxos do Amazon Kinesis para capturar, armazenar e transportar dados.

Você pode controlar o acesso lógico aos recursos e funções de gerenciamento do Amazon Kinesis criando usuários na sua conta da AWS usando o AWS IAM e controlando quais operações do Amazon Kinesis esses usuários têm permissão para executar.

Para facilitar a execução de aplicativos produtores ou consumidores em uma instância do Amazon EC2, você pode configurar essa instância com uma função do IAM. Dessa forma, as credenciais da AWS que refletem as permissões associadas à função do IAM são disponibilizadas para aplicativos na instância, o que significa que você não precisa usar suas credenciais de segurança da AWS a longo prazo.

As funções têm o benefício adicional de fornecer credenciais temporárias que expiram dentro de um curto período de tempo, o que adiciona uma medida adicional de proteção.

A API Amazon Kinesis só pode ser acessada por meio de um ponto de extremidade criptografado por SSL (kinesis.us-east-1.amazonaws.com) para ajudar a garantir a transmissão segura de seus dados para a AWS. Você deve se conectar a esse terminal para acessar o Amazon Kinesis, mas poderá usar a API para direcionar o Amazon Kinesis para criar um fluxo em qualquer região da AWS.

Serviços de implantação e gerenciamento

A AWS fornece uma variedade de ferramentas para ajudar na implantação e gerenciamento de seus aplicativos. Isso inclui serviços que permitem criar contas de usuário individuais com credenciais para acesso aos serviços da AWS.

Ele também inclui serviços para criar e atualizar pilhas de recursos da AWS, implantar aplicativos nesses recursos e monitorar a integridade desses recursos da AWS. Outras ferramentas ajudam a gerenciar chaves criptográficas usando HSMs e registrar a atividade da API da AWS para fins de segurança e conformidade.

Segurança do AWS Identity and Access Management (IAM)

O AWS IAM permite criar vários usuários e gerenciar as permissões para cada um desses usuários na sua conta da AWS. Um usuário é uma identidade (dentro de uma conta da AWS) com credenciais de segurança exclusivas que podem ser usadas para acessar os serviços em nuvem da AWS.

O IAM elimina a necessidade de compartilhar senhas ou chaves e facilita a habilitação ou desabilitação do acesso de um usuário, conforme apropriado.

O AWS IAM permite implementar práticas recomendadas de segurança, como privilégios mínimos, concedendo credenciais exclusivas a todos os usuários da sua conta da AWS e concedendo apenas permissão para acessar os serviços e recursos da AWS Cloud necessários para que os usuários realizem suas tarefas.

O IAM é seguro por padrão; novos usuários não têm acesso à AWS até que as permissões sejam concedidas explicitamente.

O AWS IAM também está integrado ao AWS Marketplace, para que você possa controlar quem em sua organização pode se inscrever no software e serviços oferecidos no AWS Marketplace.

Como a assinatura de um determinado software no AWS Marketplace inicia uma instância do Amazon EC2 para executar o software, esse é um recurso importante de controle de acesso.

O uso do IAM para controlar o acesso ao AWS Marketplace também permite que os proprietários da conta da AWS tenham controle refinado sobre o uso e custos de software. O AWS IAM permite minimizar o uso das credenciais da sua conta da AWS.

Depois de criar contas de usuário do IAM, todas as interações com os serviços e recursos da AWS Cloud devem ocorrer com as credenciais de segurança do usuário do IAM.

Funções

Uma função do IAM usa credenciais de segurança temporárias para permitir que você delegue o acesso a usuários ou serviços que normalmente não têm acesso aos seus recursos da AWS.

Uma função é um conjunto de permissões para acessar recursos específicos da AWS, mas essas permissões não estão vinculadas a um usuário ou grupo específico do IAM. Uma entidade autorizada (por exemplo, usuário móvel ou instância do Amazon EC2) assume uma função e recebe credenciais de segurança temporárias para autenticação nos recursos definidos na função.

As credenciais de segurança temporárias fornecem segurança aprimorada devido à sua curta vida útil (a expiração padrão é de 12 horas) e ao fato de que elas não podem ser reutilizadas depois que expiram. Isso pode ser particularmente útil ao fornecer acesso limitado e controlado em determinadas situações:

Acesso de usuário federado (não pertencente à AWS)

Usuários federados são usuários (ou aplicativos) que não possuem contas da AWS. Com as funções, você pode conceder acesso a seus recursos da AWS por um período limitado.

Isso é útil se você tiver usuários que não são da AWS e que podem se autenticar com um serviço externo, como Microsoft Active Directory, LDAP (Lightweight Directory Access Protocol) ou Kerberos.

As credenciais temporárias da AWS usadas com as funções fornecem federação de identidade entre a AWS e seus usuários não pertencentes à AWS em seu sistema de identidade e autorização corporativa.

Linguagem de Marcação de Asserção de Segurança (SAML) 2.0

Se sua organização oferecer suporte ao SAML 2.0, você poderá criar confiança entre sua organização como um provedor de identidade (IdP) e outras organizações como provedores de serviços.

Na AWS, você pode configurar a AWS como o provedor de serviços e usar o SAML para fornecer aos usuários SSO federado (SSO) no AWS Management Console ou obter acesso federado para chamar as APIs da AWS.

As funções também são úteis se você criar um aplicativo móvel ou baseado na Web que acesse os recursos da AWS. Os recursos da AWS exigem credenciais de segurança para solicitações programáticas; no entanto, você não deve incorporar credenciais de segurança de longo prazo em seu aplicativo, pois elas são acessíveis aos usuários do aplicativo e podem ser difíceis de alternar.

Em vez disso, você pode permitir que os usuários façam login no seu aplicativo usando o Login com Amazon, Facebook ou Google e, em seguida, use as informações de autenticação para assumir uma função e obter credenciais de segurança temporárias.

Acesso entre contas

Para organizações que usam várias contas da AWS para gerenciar seus recursos, você pode configurar funções para fornecer aos usuários que têm permissões em uma conta para acessar recursos em outra conta.

Para organizações que possuem funcionários que raramente precisam acessar recursos em outra conta, o uso de funções ajuda a garantir que as credenciais sejam fornecidas temporariamente e somente quando necessário.

Aplicativos em execução em instâncias EC2 que precisam acessar os recursos da AWS

Se um aplicativo é executado em uma instância do Amazon EC2 e precisa fazer solicitações de recursos da AWS, como buckets do Amazon S3 ou uma tabela do DynamoDB, ele deve ter credenciais de segurança. Usando funções em vez de criar contas individuais do IAM para cada aplicativo em cada instância, pode economizar tempo significativo para os clientes que gerenciam um grande número de instâncias ou uma frota de escala elástica usando o AWS Auto Scaling.

As credenciais temporárias incluem um token de segurança, um ID da chave de acesso e uma chave de acesso secreta. Para conceder ao usuário acesso a determinados recursos, você distribui as credenciais de segurança temporárias ao usuário a quem você está concedendo acesso temporário. Quando o usuário faz chamadas para seus recursos, ele passa o token e o ID da chave de acesso e assina a solicitação com a Chave de acesso secreta. O token não funcionará com chaves de acesso diferentes.

O uso de credenciais temporárias fornece proteção adicional para você, porque você não precisa gerenciar ou distribuir credenciais de longo prazo para usuários temporários. Além disso, as credenciais temporárias são carregadas automaticamente na instância de destino, para que você não precise incorporá-las em algum lugar inseguro como o seu código.

As credenciais temporárias são giradas ou alteradas automaticamente várias vezes ao dia, sem nenhuma ação da sua parte, e são armazenadas com segurança por padrão.

Serviços Móveis

Os serviços móveis da AWS facilitam a criação, o envio, a execução, o monitoramento, a otimização e o dimensionamento de aplicativos para dispositivos móveis baseados na nuvem.

Esses serviços também ajudam a autenticar usuários no seu aplicativo móvel, sincronizar dados e coletar e analisar o uso do aplicativo.

Amazon Cognito Security

O Amazon Cognito fornece serviços de identidade e sincronização para aplicativos móveis e baseados na Web. Ele simplifica a tarefa de autenticar usuários e armazenar, gerenciar e sincronizar seus dados em vários dispositivos, plataformas e aplicativos.

Ele fornece credenciais temporárias com privilégios limitados para usuários autenticados e não autenticados sem precisar gerenciar nenhuma infraestrutura de back-end.

O Amazon Cognito trabalha com provedores de identidade conhecidos como Google, Facebook e Amazon para autenticar usuários finais de seus aplicativos móveis e da Web. Você pode tirar proveito dos recursos de identificação e autorização fornecidos por esses serviços, em vez de precisar criar e manter seus próprios.

Seu aplicativo se autentica com um desses provedores de identidade usando o SDK do provedor. Depois que o usuário final é autenticado com o fornecedor, um token OAuth ou OpenID Connect retornado do provedor é passado pelo seu aplicativo para o Amazon Cognito, que retorna um novo ID do Amazon Cognito para o usuário e um conjunto de credenciais temporárias da AWS com privilégios limitados.

Para começar a usar o Amazon Cognito, você cria um pool de identidades por meio do console do Amazon Cognito. O pool de identidades é um armazenamento de informações de identidade do usuário específicas da sua conta da AWS.

Durante a criação do pool de identidades, você será solicitado a criar uma nova função do IAM ou escolher uma existente para seus usuários finais. Uma função do IAM é um conjunto de permissões para acessar recursos específicos da AWS, mas essas permissões não estão vinculadas a um usuário ou grupo específico do IAM.

Uma entidade autorizada (por exemplo, usuário móvel, instância do Amazon EC2) assume uma função e recebe credenciais de segurança temporárias para autenticação nos recursos da AWS definidos na função.

As credenciais de segurança temporárias fornecem segurança aprimorada devido à sua curta vida útil (a expiração padrão é de 12 horas) e ao fato de que elas não podem ser reutilizadas depois que expiram.

A função que você seleciona afeta os serviços da AWS Cloud que seus usuários finais poderão acessar com credenciais temporárias. Por padrão, o Amazon Cognito cria uma nova função com permissões limitadas; os usuários finais têm acesso apenas ao serviço Amazon Cognito Sync e Amazon Mobile Analytics. Se seu aplicativo precisar acessar outros recursos da AWS, como Amazon S3 ou Amazon DynamoDB, você poderá modificar suas funções diretamente no console do IAM.

Com o Amazon Cognito, não há necessidade de criar contas individuais da AWS ou mesmo contas do IAM para todos os usuários finais de aplicativos da Web / dispositivos móveis que precisarão acessar seus recursos da AWS.

Em conjunto com as funções do IAM, os usuários móveis podem acessar com segurança a recursos da AWS e recursos de aplicativos e até mesmo salvar dados na nuvem da AWS sem precisar criar uma conta ou fazer login.

Se optarem por criar uma conta ou fazer login posteriormente, o Amazon Cognito mesclará informações de dados e identificação.

Como o Amazon Cognito armazena dados localmente e também no serviço, seus usuários finais podem continuar interagindo com os dados, mesmo quando estão offline. Seus dados offline podem estar obsoletos, mas eles podem recuperar imediatamente qualquer coisa que colocarem no conjunto de dados, estejam eles online ou não.

O SDK do cliente gerencia um armazenamento SQLite local para que o aplicativo possa funcionar mesmo quando não estiver conectado.

O armazenamento SQLite funciona como um cache e é o alvo de todas as operações de leitura e gravação. O recurso de sincronização do Amazon Cognito compara a versão local dos dados à versão em nuvem e aumenta ou diminui os deltas, conforme necessário. Observe que, para sincronizar dados entre dispositivos, seu pool de identidades deve suportar identidades autenticadas.

Identidades não autenticadas estão vinculadas ao dispositivo, portanto, a menos que um usuário final se autentique, nenhum dado poderá ser sincronizado em vários dispositivos.

Com o Amazon Cognito, seu aplicativo se comunica diretamente com um provedor de identidade pública suportado (Amazon, Facebook ou Google) para autenticar usuários. O Amazon Cognito não recebe ou armazena credenciais de usuário, apenas o token OAuth ou OpenID Connect recebido do provedor de identidade.

Depois que o Amazon Cognito recebe o token, ele retorna um novo ID do Amazon Cognito para o usuário e um conjunto de credenciais temporárias da AWS com privilégios limitados. Cada identidade do Amazon Cognito tem acesso apenas a seus próprios dados no armazenamento de sincronização, e esses dados são criptografados quando armazenados. Além disso, todos os dados de identidade são transmitidos por HTTPS.

O identificador exclusivo do Amazon Cognito no dispositivo é armazenado no local seguro apropriado. Por exemplo, no iOS, o identificador do Amazon Cognito é armazenado no chaveiro do iOS. Os dados do usuário são armazenados em cache em um banco de dados SQLite local na caixa de proteção do aplicativo; se você precisar de segurança adicional, poderá criptografar esses dados de identidade no cache local implementando a criptografia no seu aplicativo.

Aplicativos

Os aplicativos da AWS são serviços gerenciados que permitem fornecer a seus usuários áreas de trabalho e armazenamento centralizadas e seguras na nuvem.

Segurança do Amazon WorkSpaces

O Amazon WorkSpaces é um serviço de desktop gerenciado que permite o provisionamento rápido de desktops baseados em nuvem para seus usuários. Basta escolher um pacote do Windows 7 que melhor atenda às necessidades de seus usuários e ao número de WorkSpaces que você deseja iniciar.

Depois que os WorkSpaces estiverem prontos, os usuários receberão um email informando onde podem fazer o download do cliente relevante e efetuar login no WorkSpace.

Eles podem acessar seus desktops baseados na nuvem a partir de uma variedade de dispositivos de terminal, incluindo PCs, laptops e dispositivos móveis.

No entanto, os dados da sua organização nunca são enviados ou armazenados no dispositivo do usuário final porque o Amazon WorkSpaces usa PC-over-IP (PCoIP), que fornece um fluxo de vídeo interativo sem transmitir dados reais.

O protocolo PCoIP compacta, criptografa e codifica a experiência de computação em desktop dos usuários e transmite como pixels apenas em qualquer rede IP padrão para dispositivos do usuário final.

Para acessar o seu WorkSpace, os usuários devem entrar usando um conjunto de credenciais exclusivas ou suas credenciais regulares do Active Directory. Quando você integra o Amazon WorkSpaces ao Active Directory corporativo, cada WorkSpace ingressa no domínio do Active Directory e pode ser gerenciado como qualquer outra área de trabalho da sua organização.

Isso significa que você pode usar diretivas de grupo do Active Directory para gerenciar os espaços de trabalho dos usuários e especificar opções de configuração que controlam a área de trabalho. Se você optar por não usar o Active Directory ou outro tipo de diretório onpremises para gerenciar seu WorkSpaces do usuário, poderá criar um diretório de nuvem privada no Amazon WorkSpaces que possa ser usado para administração.

Para fornecer uma camada adicional de segurança, você também pode exigir o uso do MFA ao entrar na forma de um token de hardware ou software.

O Amazon WorkSpaces oferece suporte ao MFA usando um servidor RADIUS (Remote Authentication Dial In User Service) ou qualquer provedor de segurança que suporta autenticação RADIUS. Atualmente, ele suporta os protocolos PAP, CHAP, MSCHAP1 e MS-CHAP2, juntamente com proxies RADIUS.

Cada WorkSpace reside em sua própria instância do Amazon EC2 em um Amazon VPC. Você pode criar WorkSpaces em um Amazon VPC que você já possui ou fazer com que o serviço Amazon WorkSpaces crie um para você automaticamente, usando a opção Início rápido do Amazon WorkSpaces.

Quando você usa a opção Início rápido, o Amazon WorkSpaces não apenas cria o Amazon VPC, mas também executa várias outras tarefas de provisionamento e configuração para você, como a criação de um gateway da Internet para o Amazon VPC, a configuração de um diretório no Amazon VPC que é usado para armazenar informações de usuário e espaço de trabalho, criando uma conta de administrador de diretório, criando as contas de usuário especificadas e adicionando-as ao diretório e criando as instâncias do Amazon WorkSpaces.

Ou o Amazon VPC pode ser conectado a uma rede local usando uma conexão VPN segura para permitir o acesso a um Active Directory local existente e outros recursos da intranet. Você pode adicionar um grupo de segurança criado no Amazon VPC a todos os espaços de trabalho que pertencem ao seu Active Directory.

Isso permite controlar o acesso à rede do Amazon WorkSpaces no Amazon VPC para outros recursos no Amazon VPC e na rede local.

O armazenamento persistente do Amazon WorkSpaces é fornecido pelo Amazon EBS e é feito backup automaticamente duas vezes por dia no Amazon S3. Se o Amazon WorkSpaces Sync estiver ativado em um WorkSpace, a pasta que um usuário escolher para sincronizar será continuamente copiada e armazenada no Amazon S3.

Você também pode usar o Amazon WorkSpaces Sync em um Mac ou PC para sincronizar documentos de ou para o seu WorkSpace, para que você possa sempre ter acesso aos seus dados, independentemente do computador que estiver usando.

Por ser um serviço gerenciado, a AWS cuida de várias tarefas de segurança e manutenção, como backups diários e aplicação de patches. As atualizações são entregues automaticamente nos seus WorkSpaces durante uma janela de manutenção semanal.

Você pode controlar como a aplicação de patches é configurada para o espaço de trabalho de um usuário. Por padrão, o Windows Update está ativado, mas você pode personalizar essas configurações ou use uma abordagem alternativa de gerenciamento de patches, se desejar. Para o sistema operacional subjacente, o Windows Update é ativado por padrão no Amazon WorkSpaces e configurado para instalar atualizações semanalmente.

Você pode usar uma abordagem alternativa de patch ou configurar o Windows Update para executar atualizações no momento que você escolher. Você pode usar o IAM para controlar quem em sua equipe pode executar funções administrativas, como criar ou excluir WorkSpaces ou configurar diretórios de usuários.

Você também pode configurar um WorkSpace para administração de diretórios, instalar suas ferramentas de administração favoritas do Active Directory e criar unidades organizacionais e políticas de grupo para aplicar as alterações do Active Directory com mais facilidade a todos os usuários do Amazon WorkSpaces.

Risco e Compliance na AWS

A AWS e seus clientes compartilham o controle sobre o ambiente de TI, para que ambas as partes sejam responsáveis pelo gerenciamento desse ambiente. A parte da AWS nessa responsabilidade compartilhada inclui o fornecimento de seus serviços em uma plataforma altamente segura e controlada e o fornecimento de uma ampla variedade de recursos de segurança que os clientes podem usar.

O cliente é responsável por configurar seu ambiente de TI de maneira segura e controlada para seus propósitos. Embora os clientes não comuniquem seu uso e configurações à AWS, a AWS se comunica com os clientes sobre sua segurança e controle do ambiente, conforme relevante. A AWS divulga essas informações usando três mecanismos principais.

Primeiro, a AWS trabalha diligentemente para obter certificações do setor e atestados independentes de terceiros. Segundo, a AWS publica abertamente informações sobre suas práticas de segurança e controle em whitepapers e conteúdo de sites.

Por fim, a AWS fornece certificados, relatórios e outras documentações diretamente a seus clientes sob os contratos de confidencialidade (NDAs), conforme necessário.

Quando os clientes transferem suas cargas de trabalho de produção para a nuvem da AWS, ambas as partes se tornam responsáveis pelo gerenciamento do ambiente de TI.

Os clientes são responsáveis por configurar seu ambiente de maneira segura e controlada. Os clientes também precisam manter a governança adequada sobre todo o ambiente de controle de TI. Esta seção descreve o modelo de responsabilidade compartilhada da AWS e fornece conselhos sobre como estabelecer uma conformidade forte.

Modelo de Responsabilidade Compartilhada

Esse modelo de responsabilidade compartilhada pode ajudar a diminuir a carga operacional de TI de um cliente, pois é responsabilidade da AWS gerenciar os componentes do sistema operacional host e da virtualização até a segurança física dos datacenters em que esses serviços operam.

O cliente é responsável pelos componentes do sistema operacional convidado para cima (incluindo atualizações, patches de segurança e software antivírus). O cliente também é responsável por qualquer outro software aplicativo, bem como pela configuração de grupos de segurança, Nuvens Privadas Virtuais (VPCs) e assim por diante.

Enquanto a AWS gerencia a segurança da nuvem, a segurança na nuvem é de responsabilidade do cliente. Os clientes mantêm o controle de qual segurança eles escolhem implementar para proteger seu próprio conteúdo, plataforma, aplicativos, sistemas e redes, da mesma forma que seria para aplicativos em um data center no local.

Os clientes precisam estar cientes de quaisquer leis e regulamentos aplicáveis com os quais devem cumprir e, em seguida, devem considerar se os serviços que eles consomem na AWS estão em conformidade com essas leis. Em alguns casos, pode ser necessário melhorar na AWS com medidas de segurança adicionais (como implantar um firewall de aplicativo da web, Sistema de detecção de intrusões [IDS] ou Sistema de prevenção de intrusões [IPS] ou usar alguma forma de criptografia para os dados em repouso)

Esse modelo de responsabilidade compartilhada do cliente / AWS não se limita apenas a considerações de segurança, mas também se estende aos controles de TI. Por exemplo, o gerenciamento, a operação e a verificação dos controles de TI são compartilhados entre a AWS e o cliente.

Antes de migrar para a Nuvem AWS, os clientes eram responsáveis por gerenciar todos os controles de TI em seus ambientes. A AWS gerencia os controles da infra-estrutura física, levando assim a carga pesada indiferenciada dos clientes, permitindo que eles se concentrem no gerenciamento dos controles de TI relevantes. Como cada cliente é implantado de maneira diferente na AWS, os

clientes podem mudar o gerenciamento de determinados controles de TI para a AWS.

Essa mudança no gerenciamento dos controles de TI resulta em um novo ambiente de controle distribuído. Os clientes podem usar a documentação de controle e conformidade da AWS disponível para executar seus procedimentos de avaliação e verificação de controle, conforme necessário.

Forte governança de conformidade

Ainda é responsabilidade dos clientes manter uma governança adequada em todo o ambiente de controle de TI, independentemente de como a TI é implantada (seja no local, na nuvem ou parte de um ambiente híbrido). Ao implantar na nuvem da AWS, os clientes têm opções para aplicar diferentes tipos de controles e vários métodos de verificação.

Para alcançar uma forte conformidade e governança, os clientes podem querer seguir esta metodologia básica:

1. Adote uma abordagem holística. Revise as informações disponíveis na AWS, juntamente com todas as outras informações, para entender o máximo possível do ambiente de TI. Após a conclusão, documente todos os requisitos de conformidade.

2. Projete e implemente objetivos de controle para atender aos requisitos de conformidade da organização.

3. Identifique e documente os controles de propriedade de todos os terceiros.

4. Verifique se todos os objetivos de controle foram alcançados e se todos os controles principais foram projetados e operando com eficiência.

Ao usar essa metodologia básica, os clientes podem entender melhor seu ambiente de controle. Por fim, isso simplificará o processo e ajudará a separar as atividades de verificação que precisam ser executadas.

Avaliando e integrando controles da AWS

A AWS fornece aos clientes uma ampla gama de informações sobre seu ambiente de controle de TI por meio de documentos técnicos, relatórios, certificações e outros atestados de terceiros.

Esta documentação ajuda os clientes a entender os controles em vigor relevantes para os serviços em nuvem da AWS que eles usam e como esses controles foram validados. Essas informações também auxiliam os clientes em seus esforços para contabilizar e validar se os controles em seu ambiente de TI estendido estão operando efetivamente.

Tradicionalmente, o design e a eficácia operacional dos controles e objetivos de controle são validados pelos auditores internos e / ou externos por meio de orientações passo a passo do processo e avaliação de evidências.

Observação direta e verificação, pelo cliente ou auditor externo do cliente, geralmente são realizadas para validar os controles. No caso de fornecedores de serviços como a AWS, empresas solicitam e avaliam atestados e certificações de terceiros a fim de obter garantia razoável do design e da eficácia operacional dos controles e objetivos do controle.

Como resultado, embora os principais controles de um cliente possam ser gerenciados pela AWS, o ambiente de controle ainda pode ser uma estrutura unificada na qual todos os controles são contabilizados e verificados como operando efetivamente. Os atestados e certificações de terceiros da AWS não apenas fornecem um nível mais alto de validação do ambiente de controle, mas também pode aliviar os clientes do requisito de realizar determinados trabalhos de validação.

Informações de controle de TI da AWS

A AWS fornece informações de controle de TI aos clientes das duas maneiras a seguir.

Definição de controle específico

Os clientes da AWS podem identificar os principais controles gerenciados pela AWS. Os controles-chave são críticos para o ambiente de controle do cliente e exigem um atestado externo da eficácia operacional desses controles-chave para atender aos requisitos de conformidade (por exemplo, uma auditoria financeira anual). Para esse fim, a AWS publica uma ampla gama de controles de TI específicos em seu relatório Service Organization Controls 1 (SOC 1) Tipo II.

O relatório SOC 1 Tipo II, anteriormente a Declaração sobre Normas de Auditoria (SAS) Nº 70, é um padrão de auditoria amplamente reconhecido desenvolvido pelo Instituto Americano de Contadores Públicos Certificados

(AICPA). A auditoria SOC 1 é uma auditoria aprofundada da eficácia do projeto e da operação dos objetivos de controle definidos e das atividades de controle da AWS (que incluem objetivos de controle e atividades de controle sobre a parte da infraestrutura que a AWS gerencia). "Tipo II" refere-se ao fato de que cada um dos controles descritos no relatório não é apenas avaliados quanto à adequação do projeto, mas também são testados quanto à eficácia operacional pelo auditor externo. Devido à independência e competência do auditor externo da AWS, os controles identificados no relatório devem fornecer aos clientes um alto nível de confiança no ambiente de controle da AWS.

Os controles da AWS podem ser considerados efetivamente projetados e operacionais para vários fins de conformidade, incluindo auditorias às demonstrações financeiras da Seção 404 da Sarbanes-Oxley (SOX). A utilização de relatórios SOC 1 Tipo II também é geralmente permitida por outros organismos de certificação externos. Por exemplo, os auditores da Organização Internacional de Normalização (ISO) 27001 podem solicitar um relatório SOC 1 Tipo II para concluir suas avaliações para clientes

Conformidade com o padrão de controle geral

Se um cliente da AWS exigir que um amplo conjunto de objetivos de controle seja alcançado, a avaliação das certificações do setor da AWS poderá ser realizada. Com a certificação ISO 27001, a AWS está em conformidade com um padrão de

segurança amplo e abrangente e segue as práticas recomendadas para manter um ambiente seguro. Com a certificação DSS (Data Security Standard) do setor de cartões de pagamento (PCI), a AWS cumpre um conjunto de controles importantes para as empresas que lidam com informações de cartão de crédito.

A conformidade da AWS com os padrões da Federal Information Security Management Act (FISMA) significa que a AWS cumpre uma ampla gama de controles específicos exigidos pelas agências governamentais dos EUA. A conformidade da AWS com esses padrões gerais fornece aos clientes informações detalhadas sobre a natureza abrangente dos controles e processos de segurança em vigor na nuvem da AWS.

Regiões globais da AWS

A infraestrutura da nuvem da AWS é construída em torno de regiões e zonas de disponibilidade. Uma região é um local físico no mundo em que temos várias zonas de disponibilidade. As zonas de disponibilidade consistem em um ou mais data centers distintos, cada um com energia, rede e rede redundantes em conectividade, instalado em instalações separadas. Essas zonas de disponibilidade oferecem aos clientes a capacidade de operar aplicativos e bancos de dados de produção mais altamente disponíveis, tolerantes a falhas e escaláveis do que seria possível usando um único data center.

Até o momento em que este artigo foi escrito, a AWS Cloud opera 33 zonas de disponibilidade em 12 regiões geográficas do mundo. As 12 regiões são Leste dos EUA (Virgínia do Norte), Oeste dos EUA (Oregon), Oeste dos EUA (norte da Califórnia), AWS GovCloud (EUA) (Oregon), UE (Frankfurt), UE (Irlanda), Ásia-Pacífico (Cingapura), Ásia Pacífico (Tóquio), Ásia-Pacífico (Sydney), Ásia-Pacífico (Seul), China (Pequim) e América do Sul (São Paulo).

Programa de conformidade e risco da AWS

O AWS Risk and Compliance foi desenvolvido para aproveitar os programas tradicionais e ajudar os clientes a estabelecer e operar em um ambiente de controle de segurança da AWS.

A AWS fornece informações detalhadas sobre seu programa de risco e conformidade para permitir que os clientes incorporem os controles da AWS em suas estruturas de governança. Essas informações podem ajudar os clientes a documentar estruturas completas de controle e governança nas quais a AWS está incluída como uma parte importante.

As três áreas principais do programa de riscos e conformidade - gerenciamento de riscos, ambiente de controle e segurança da informação - são descritas a seguir.

Gerenciamento de riscos

A AWS desenvolveu um plano estratégico de negócios que inclui a identificação de riscos e a implementação de controles para mitigar ou gerenciar riscos. Uma equipe de gerenciamento da AWS reavalia o plano de risco comercial pelo menos duas vezes por ano.

Como parte desse processo, os membros da equipe de gerenciamento são obrigados a identificar riscos em suas áreas específicas de responsabilidade e implementar controles projetados para tratar e talvez até eliminar esses riscos.

O ambiente de controle da AWS está sujeito a avaliações de risco internas e externas adicionais. As equipes de conformidade e segurança da AWS estabeleceram uma estrutura e políticas de segurança da informação com base na estrutura COBIT (Objetivos de Controle para Tecnologia da Informação e Tecnologia Relacionada) e integraram efetivamente a estrutura certificável ISO 27001 com base nos controles ISO 27002, nos Princípios dos Serviços de Confiança da AICPA, O PCI DSS v3.1 e as Publicações 800–53 do Instituto Nacional de Padrões e Tecnologia (NIST), Revisão 3, Controles de segurança recomendados para sistemas de informações federais.

A AWS mantém a política de segurança e fornece treinamento de segurança aos seus funcionários. Além disso, a AWS realiza revisões regulares de segurança de aplicativos para avaliar a confidencialidade, integridade e disponibilidade de dados e conformidade com a política de segurança da informação.

A equipe de segurança da AWS verifica regularmente todos os endereços IP de endpoints voltados para o público em busca de vulnerabilidades. É importante entender que essas verificações não incluem instâncias do cliente. A segurança da AWS notifica as partes apropriadas para corrigir quaisquer vulnerabilidades identificadas. Além disso, empresas de segurança independentes realizam regularmente avaliações externas de ameaças de vulnerabilidade.

As descobertas e recomendações resultantes dessas avaliações são categorizadas e entregues à liderança da AWS. Essas verificações são feitas de maneira a garantir a integridade e a viabilidade da infraestrutura subjacente da AWS e não têm como objetivo substituir as verificações de vulnerabilidade do cliente necessárias para atender aos requisitos de conformidade específicos.

Conforme mencionado, os clientes podem solicitar permissão para realizar suas próprias verificações de vulnerabilidade em seus próprios ambientes. Essas verificações de vulnerabilidades não devem violar a política de uso aceitável da AWS e devem ser solicitadas antes da verificação.

Ambiente de controle

A AWS gerencia um ambiente de controle abrangente que consiste em políticas, processos e atividades de controle. Esse ambiente de controle existe para a entrega segura de ofertas de serviços da AWS.

O ambiente de controle coletivo inclui pessoas, processos e tecnologia necessários para estabelecer e manter um ambiente que ofereça suporte à eficácia operacional da estrutura de controle da AWS. A AWS integrou controles específicos aplicáveis à nuvem identificados pelos principais organismos do setor de computação em nuvem na estrutura de controle da AWS.

A AWS continua a monitorar esses grupos do setor em busca de idéias sobre quais práticas líderes podem ser implementadas para ajudar melhor os clientes no gerenciamento de seus ambientes de controle.

O ambiente de controle da AWS começa no nível mais alto da empresa. A liderança executiva e sênior desempenha papéis importantes no estabelecimento do tom e dos valores essenciais da empresa.

Todos os funcionários recebem o código de conduta e ética comercial da empresa e concluem o treinamento periódico. As auditorias de conformidade são realizadas para que os funcionários entendam e sigam as políticas estabelecidas.

A estrutura organizacional da AWS fornece uma estrutura para planejar, executar e controlar operações de negócios. A estrutura organizacional atribui funções e responsabilidades para fornecer pessoal adequado, eficiência das operações e segregação de funções.

A gerência também estabeleceu autoridade e linhas de relatório apropriadas para o pessoal-chave. Como parte dos processos de verificação de contratação da empresa, estão incluídos educação, emprego anterior e, em alguns casos, verificações de antecedentes permitidas por lei para funcionários compatíveis com a posição e o nível de acesso do funcionário às instalações da AWS.

A empresa segue um processo de integração estruturado para familiarizar os novos funcionários com as ferramentas, processos, sistemas, políticas e procedimentos da Amazon.

Segurança da Informação

A AWS usa um programa formal de segurança da informação desenvolvido para proteger a confidencialidade, a integridade e a disponibilidade dos sistemas e dados dos clientes. A AWS publica vários documentos técnicos de segurança disponíveis no site principal da AWS. Estes documentos técnicos são de leitura recomendada.

Conclusão

Normalmente, os sistemas de produção vêm com requisitos definidos ou implícitos em termos de tempo de atividade.

Um sistema está altamente disponível quando pode suportar a falha de um componente individual ou múltiplo.

Se você projetar arquiteturas com base na suposição de que algum componente acabará por falhar, os sistemas não falharão quando um componente individual falhar.

A infraestrutura tradicional geralmente requer a previsão da quantidade de recursos de computação que seu aplicativo usará por um período de vários anos.

Se você subestimar, seus aplicativos não terão a potência necessária para lidar com tráfego inesperado, resultando potencialmente em insatisfação do cliente. Se você superestimar, está desperdiçando dinheiro com recursos supérfluos.

A natureza sob demanda e elástica da nuvem permite que a infraestrutura esteja alinhada com a demanda real, aumentando assim a utilização geral e reduzindo os custos.

Enquanto a computação em nuvem fornece capacidade sob demanda praticamente ilimitada, os sistemas e as arquiteturas precisam poder tirar proveito desses recursos sem problemas.

Geralmente, existem duas maneiras de dimensionar uma arquitetura de TI: vertical e horizontalmente.

A nuvem da AWS fornece recursos de governança que permitem o monitoramento contínuo de alterações na configuração de seus recursos de TI.

Como os ativos da AWS são recursos programáveis, sua política de segurança pode ser formalizada e incorporada ao design da sua infraestrutura.

Com a capacidade de ativar ambientes temporários, os testes de segurança agora podem se tornar parte de seu pipeline de entrega contínua.

Os arquitetos de soluções podem aproveitar uma infinidade de recursos nativos de segurança e criptografia da AWS que podem ajudar a alcançar níveis mais altos de proteção e conformidade de dados em todas as camadas de arquiteturas em nuvem.

Como a AWS facilita a paralelização, os arquitetos de soluções precisam internalizar o conceito de paralelização ao projetar arquiteturas na nuvem.

É aconselhável não apenas implementar a paralelização sempre que possível, mas também automatizá-la, pois a nuvem permite criar um processo repetível com muita facilidade.

À medida que a complexidade do aplicativo aumenta, uma característica desejável de um sistema de TI é que ele pode ser dividido em componentes menores e pouco acoplados.

Os arquitetos de soluções devem projetar os sistemas de maneira a reduzir as interdependências, para que uma alteração ou falha em um componente não seja transmitida em cascata a outros componentes.

Quando as organizações tentam mapear suas especificações de sistema existentes para as disponíveis na nuvem, elas percebem que a nuvem pode não ter a especificação exata do recurso que possui no local.

As organizações não devem ter medo e se sentir constrangidas ao usar recursos na nuvem. Mesmo que você não consiga obter uma réplica exata do seu hardware no ambiente em nuvem, você poderá obter mais desses recursos na nuvem para compensar.

Concentrando-se em conceitos e práticas recomendadas - como projetar para falhas, dissociar os componentes do aplicativo, entender e implementar a elasticidade, combiná-lo com paralelização e integrar a segurança em todos os aspectos da arquitetura do aplicativo.

Arquitetos de soluções podem entender as considerações de design necessárias para criar aplicativos em nuvem altamente escaláveis.

Como cada caso de uso é único, os Arquitetos de Soluções precisam permanecer diligentes na avaliação de como as melhores práticas e padrões podem ser aplicados a cada implementação.

O tópico das arquiteturas de computação em nuvem é amplo e está em constante evolução.