**Computação em Nuvem da AWS**

A AWS possui uma infraestrutura de nuvem global abrangente, confiável e segura, com mais de 175 serviços para uma ampla variedade de casos de uso. Esses serviços são criados em uma infraestrutura robusta em que o cliente e a AWS compartilham as responsabilidades de programação e segurança.

A AWS oferece um conjunto abrangente que inclui mais de 175 serviços globais baseados na nuvem. Esses produtos incluem computação, armazenamento, banco de dados, análises, redes, ferramentas do desenvolvedor, aplicações empresariais, ferramentas de gerenciamento, machine learning, internet das coisas e segurança.

A AWS oferece benefícios inovadores aos clientes. Como o acesso sob demanda da AWS, um usuário pode provisionar os serviços necessários, quando precisar deles, em apenas alguns minutos. Com o modelo de preço da AWS de pagamento conforme o uso, o usuário paga por um serviço somente pelo tempo que o utilizar. Novos serviços podem ser provisionados rapidamente, sem despesa de capital inicial nem compromisso. Empresas, startups, clientes do setor público e pequenas e médias empresas podem acessar os componentes dos quais elas precisam.

Dessa forma, elas podem responder rapidamente a mudanças nas necessidades de negócios. Seja qual for o porte da empresa, ela pode ter vários experimentos sem precisar conviver com o efeito colateral decorrente de resultados mal sucedidos.

Com a AWS, o usuário tem acesso a um conjunto de ferramentas de serviços que de outra forma, não conseguirá usar ou comprar. Esses serviços incluem machine learning, análise, IOT e serviços avançados de segurança.

Comportamento dos clientes têm mudado de acordo com o tempo, recentemente, os clientes têm escolhido cada vez mais a busca por uma solução em nuvem para hospedar a infraestrutura deles. Essa migração tem trazido a vantagem de melhorar a performance, segurança e confiabilidade e escalonamento de qualquer serviço que eles possuem.

A AWS vem mantendo consistentemente uma posição de liderança em infraestrutura em nuvem. De 2010 a 2020, a AWS foi avaliada como líder no Quadrante Mágico do Gartner em serviços de infraestrutura e plataforma de nuvem.

Entre os setes principais fornecedores de nuvem, a AWS ficou em primeiro lugar em ambos os eixos de medição: Capacidade de Execução e Integridade de Visão.

**A Infraestrutura Global da AWS**

Os principais componentes da infraestrutura global da AWS são Regiões, zonas de disponibilidade e locais de borda.

* Região - A AWS tem o conceito de Região, que é um local físico ao redor do mundo onde agrupam-se datacenters. Cada grupo de datacenters lógicos é chamado de zona de disponibilidade. Cada região da AWS consiste em várias zonas de disponibilidade isoladas e separadas fisicamente em uma área geográfica. Cada região da AWS consiste em várias zonas de disponibilidade isoladas e separadas fisicamente em uma área geográfica.
* Zona de disponibilidade - É uma área circunscrita dentro de uma região que pode abrigar um ou mais data centers (geralmente três). As zonas de disponibilidade abrigam todos os dispositivos de hardware que a AWS oferece. Com sua própria estrutura de energia, as zonas de disponibilidade são fisicamente separadas por uma distância física significativa (100 KM ou 60 milhas) de todas as outras zonas de disponibilidade na região. As zonas de disponibilidade são interconectadas com redes de alta largura de banda e baixa latência para fornecer uma baixa latência entre as zonas que seja suficiente para realizar a replicação síncrona.
* Locais de Borda - As edge location ou locais de borda são conectados às Regiões AWS por meio de uma rede da AWS ao redor do mundo. Eles se conectam a dezenas de milhares de redes para obter download da origem aprimorada e aceleração dinâmica de conteúdo. Os edge location armazenam cópias em cache de seu conteúdo para entrega mais rápida aos usuários em qualquer local. Eles oferecem suporte a produtos da AWS, como o Amazon Route 53 e o Amazon CloudFront.

**Planejamento para falhas**

Ao construir uma arquitetura na nuvem, é importante preparar-se para falhas. Ao construí-la, é necessário ter um plano para resolver quaisquer falhas que possam ocorrer.

Armazenamento - Quando um arquivo é armazenado no Amazon S3, ele é copiado de forma redundante em todas as zonas de disponibilidade dessa Região. Se uma zona de disponibilidade ficar indisponível, ainda assim o usuário terá duas cópias desse arquivo disponíveis para uso.

Computação - uma das práticas recomendadas é distribuir recursos de computação em várias zonas de disponibilidade para garantir alta disponibilidade. Portanto, mesmo que uma zona de disponibilidade fique inoperante, sua arquitetura vai continuar funcionando.

Banco de dados - o usuário pode configurar o banco de dados para uma implantação multi-AZ. Se houver falha na zona de disponibilidade com o banco de dados primário, um dos bancos de dados standby em uma zona de disponibilidade íntegra se tornará automaticamente o novo banco de dados primário. Portanto, a arquitetura vai manter-se funcionando.

Os planos de falhas podem abranger a infraestrutura global da AWS. Por exemplo, se uma arquitetura em produção estiver localizada na Região Norte da Virgínia. O profissional de TI pode ter um backup que espelhe a arquitetura de produção em menor escala na região do Oregon ou em qualquer região do mundo. O profissional de TI só precisa expandir verticalmente seus recursos na região de backup.

**Vantagens da infraestrutura global da AWS**

Performance - a infraestrutura global da AWS é uma infraestrutura de nuvem de alta performance e baixa latência com capacidade praticamente ilimitada, o que proporciona alta disponibilidade.

Disponibilidade - As zonas de disponibilidade são projetadas para redundância física e para fornecer resiliência. Elas fornecem performance ininterrupta, mesmo em caso de falta de energia, tempo de inatividade da internet, inundações e outros desastres naturais.

Segurança - A infraestrutura é monitorada 2 horas por dia, 7 dias por semana para garantir a confidencialidade, integridade e disponibilidade dos dados dos clientes da AWS. Os clientes podem contar com a infraestrutura global mais segura e saber que eles sempre terão seus dados. Eles podem criptografar os dados, movê-los e gerenciar a retenção.

Confiabilidade - a infraestrutura global da AWS foi projetada e criada para redundância e confiabilidade, desde regiões até links de rede para balanceadores de carga, roteadores e firmware.

Escalabilidade - Com a infraestrutura global da AWS, as empresas podem ser flexíveis e aproveitar a escalabilidade conceitualmente infinita da nuvem. As empresas podem rapidamente disponibilizar recursos conforme a necessidade, implantando centenas ou milhares de servidores em questão de minutos.

A infraestrutura global da AWS fornece o espaço de data center mais extenso do setor. Por isso, mais clientes podem se beneficiar das economias da nuvem e reduzir o custo total de propriedade (TCO) de sua infraestrutura geral de TI.

**Responsabilidade Compartilhada**

O modelo de responsabilidade compartilhada da AWS é um modelo de segurança usado para proteger ambientes de nuvem que usam produtos da AWS. A AWS é responsável pela segurança da nuvem enquanto o cliente é responsável pela segurança na nuvem.

* Responsabilidade do cliente - os clientes são responsáveis por tudo que eles criam e colocam na Nuvem AWS. Ao usar os produtos da AWS, o cliente mantém o controle total sobre o seu conteúdo, o usuário é responsável por gerenciar os requisitos de segurança de seu conteúdo. Esses requisitos incluem o conteúdo que o usuário escolhe armazenar na AWS, quais produtos da AWS o usuário utiliza e quem tem acesso a esse conteúdo. O usuário também controla como os direitos de acesso são concedidos, gerenciados e revogados.
* Responsabilidade da AWS - a AWS gerencia a segurança da nuvem, especificamente a infraestrutura física que hospeda seus recurso, que incluem:
  + Segurança física.
  + Infraestrutura de hardware e software.
  + Infraestrutura de rede.
  + Infraestrutura de virtualização.

Por exemplo, suponha que um usuário crie um bucket usando o Amazon S3 para armazenar arquivos nele. Nesse caso, o usuário é responsável pela segurança na nuvem. O usuários será responsável por conceder permissões de acesso aos usuários e aplicar qualquer requisito de criptografia aos buckets e arquivos.

Dados do cliente:

* plataforma, aplicações, gerenciamento de identidade e acesso.
* configuração de sistema operacional, redes e firewall.
* criptografia de dados no lado do cliente.

A AWS é responsável pela segurança da nuvem. Essa responsabilidade inclui a manutenção do hardware de armazenamento no qual o arquivo está hospedado e a segurança física dos datacenters onde o hardware está localizado.

Serviços básicos da AWS:

* Computação
* Armazenamento
* Banco de dados
* Redes

Infraestrutura global:

* Zonas de disponibilidade.
* Regiões.
* Locais de borda.

**AWS Well-Architected Framework**

A AWS desenvolveu este framework depois de analisar milhares de arquiteturas de clientes na AWS. Ela foi projetada para ajudar ao usuário a criar a infraestrutura que ofereça a mais alta segurança, performance, resiliência, economia possível na nuvem AWS. O framework oferece uma abordagem consistente para avaliar as melhores práticas da AWS. Ela se baseia em cinco pilares, com princípios de design que possibilitam um dimensionamento de acordo com suas necessidades ao longo do tempo.

Ao criar uma arquitetura de TI é muito semelhante a construir um edifício. Se a base não for sólida, isso poderá provocar problemas estruturais que minam a integridade e a função do edifício. Ao projetar soluções tecnológicas, o usuário não pode ignorar os cinco pilares: excelência operacional, segurança, confiabilidade, eficiência de performance e otimização de custos.

**Excelência operacional** - é a capacidade de executar e monitorar sistemas para fornecer valor comercial e melhorar continuamente os processos e procedimentos de apoio. Os princípios de design para a excelência operacional na nuvem incluem: executar operações como código, anotar documentação, antecipar falhas e fazer alterações pequenas e reversíveis com frequência.

**Segurança -**  inclui a capacidade de proteger informações, sistemas e ativos e, ao mesmo tempo, fornecer valor comercial por meio de avaliações de risco e estratégias de atenuação. Ao considerar a segurança de sua arquitetura, implemente estas melhores práticas:

* Automatizar as melhores práticas de segurança sempre que possível.
* Implementar a segurança em todas as camadas.
* Proteger dados em trânsito e em repouso.

**Confiabilidade** - capacidade de um sistema para:

* Recuperar-se de falhas de infraestrutura ou serviços.
* Adquirir dinamicamente recursos de computação para atender à demanda.
* Atenuar interrupções, como configurações inválidas e problemas de rede temporários.

A confiabilidade também inclui:

* Testar procedimentos de recuperação.
* Escalar horizontalmente para aumentar a disponibilidade agregada do sistema.
* Recuperar-se automaticamente de falhas.

**Eficiência de performance -**  capacidade de usar recursos computacionais com eficiência para cumprir requisitos do sistema e manter essa eficiência à medida que a demanda muda e as tecnologias evoluem. A avaliação de eficiência de performance de sua arquitetura inclui experimentar com maior frequência, usar arquiteturas sem servidor e projetar sistemas para ter alcance global em minutos.

O AWS Well-Architected Tool é uma ferramenta gratuita que ajuda a analisar os estados das cargas de trabalho. Ele compara as cargas de trabalho com as melhores práticas de arquitetura mais recentes da AWS. O usuário define a carga de trabalho e responde a uma série de perguntas nas áreas de excelência operacional, segurança, confiabilidade, eficiência de performance e otimização de custos. Em seguida, ele oferece um plano para definir a arquitetura com base nas melhores práticas estabelecidas.

Custo total de propriedade

É uma métrica financeira usada para estimar e comparar custos diretos e indiretos de um produto ou serviço. Normalmente, ele inclui os custos reais de aquisição, gerenciamento, manutenção e desativação de recursos de hardware, em geral durante um período de três ou cinco anos. Tendo em vista a infinidade de diferentes configurações de hardware e serviços disponíveis atualmente, às vezes é difícil saber quais os custos reais. É um desafio criar um modelo de TCO preciso que represente o verdadeiro custo de execução da aplicação.

Com a calculadora de preços da AWS, até mesmo alguém com um nível básico de conhecimento sobre infraestrutura pode gerar uma comparação de TCO baseada em fatos para uma infraestrutura da AWS ou local. Em três etapas simples, ela fornece um resumo e um relatório detalhado e abrangente com uma seção de perguntas frequentes que explica as suposições e os métodos usados.

Para usar a calculadora de preços da AWS, pesquise e adicione os produtos da AWS dos quais o usuário necessita. Em seguida, configure o serviço inserindo os detalhes necessários para a configuração desejada. Em seguida, o usuário poderá ver os custos estimados por serviço, grupo de serviço e totais.

Modelos de preços da AWS

Na AWS existem três modelos principais de preço aos clientes.

Pagamento conforme o uso - O usuário pode usar o pagamento conforme o uso para adaptar facilmente a mudanças nas necessidades dos negócios sem sobrecarregar os orçamentos e melhorar sua capacidade de respostas à mudanças. Com o modelo conforme o uso, o usuário pode adaptar sua empresa de acordo com a necessidade e não com base em previsões, o que reduz o risco de provisionamento em excesso ou perda de capacidade.

Economize ao reservar - com relação à computação em nuvem, ao machine learning da AWS e aos serviços de bancos de dados da AWS, os saving plans oferecem economia em relação ao acesso sob demanda em troca de compromisso de usar um valor específico. Esse valor é medido em dólares por hora de uso de um produto da AWS ou de uma categoria de serviços, por um período de um ou três anos. Geralmente chamado de compra de instância reserva.

Pague menos usando mais - com a AWS, o usuário pode obter descontos baseados em volume e economias importantes à medida que o uso aumenta. Para serviços como o Amazon S3, o preço é definido em camadas, o que significa que, quanto mais se utiliza, menos é pago por GB. A AWS também oferece compra de serviços que ajudam a atender às necessidades de um negócio.

Painel de Faturamento AWS

O usuário pode usar o painel do AWS Billing and Cost Management para pagar a conta da AWS., monitorar o uso, analisar e controlar os custos. Na seção de Spend Sumary, o usuário pode comparar o saldo atual acumulado do mês com o mês anterior. O usuário recebe uma previsão do mês atual com base no uso atual.

Na seção Month-to-Date Spend by Service, o usuário pode ver quanto cada serviço que está sendo utilizado está custando no momento da referência do mês.

Na seção Top Free Tier Services by Usage, é possível ver quanto o nível gratuito o usuário utilizou no mês atual. É possível ver quanto ainda falta para atingir o limite do nível gratuito.

Exemplos de faturamento

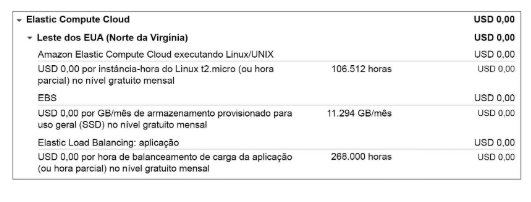
Com a Amazon EC2, o usuário paga apenas pelo tempo de computação usado enquanto suas instâncias estão em execução. Para algumas cargas de trabalho, se o usuário estiver executando um trabalho de processamento em lote que consiga suportar interrupções, poderá reduzir significativamente os custos do Amazon EC2 usando instâncias Spot. O uso de uma Instância Spot pode oferecer uma redução de custo de até 90%.

O usuário pode obter uma redução de custo adicional para o Amazon EC2 se optar pelos saving plans e instâncias reservadas.

As cobranças de serviços nesse exemplo incluem detalhes para os seguintes itens:

* Cada tipo de instância do Amazon EC2 que foi usado.
* A quantidade de espaço de armazenamento do Amazon Elastic Block Storage (Amazon EBS) provisionada.
* O período em que o Elastic Load Balancing foi usado.

Neste exemplo, todos os valores de uso estão abaixo dos limites no AWS Free Tier. Portanto, o proprietário da conta não precisaria pagar por nenhum uso da Amazon EC2 no mês em questão.



Quanto ao preço do Amazon S3, considere os seguintes componentes de custo:

Armazenamento: você paga apenas pelo armazenamento usado e será cobrado de acordo com as taxas aplicadas para armazenar objetos, nas classes de armazenamento e na quantidade de tempo durante o qual o usuário armazena cada objeto ao longo do mês.

Solicitações e recuperação de dados: somente é pago pelas solicitações feitas para seus objetos e buckets do Amazon S3.

Transferências de dados: paga pelos dados que transfere do/ para o Amazon S3, mas algumas exceções.

Gerenciamento e replicação: paga pelos recursos de gerenciamento de armazenamento habilitados nos buckets do Amazon S3 da sua conta. Neste exemplo, a conta AWS usou o Amazon S3 em duas regiões: Norte da Virgínia e Ohio. Para cada região, os encargos detalhados em itens baseiam-se nos seguintes fatores:

* Número de solicitações para adicionar ou copiar objetos em um bucket.
* Número de solicitações para recuperar objetos em um bucket.
* Quantidade de espaço de armazenamento usada.



Quanto ao AWS Lambda, o usuário é cobrado com base no número de solicitações de suas funções e no tempo que leva para elas serem executadas. Com a AWS Lambda, o usuário pode fazer 1 milhão de solicitações gratuitas e usar até 3,2 milhões de segundos de tempo de computação por mês.

O usuário pode economizar nos custos do AWS Lambda cadastrando-se em um Compute Saving Plan. Um compute Saving Plan oferece custos de computação mais baixos em troca do compromisso com uma quantidade consistente de uso durante um período de um ou três anos.

Se estiver usando o AWS Lambda em várias Regiões AWS, poderá visualizar as cobranças discriminadas por Região na fatura.

