

Infraestrutura global da AWS

Este artigo aborda o treinamento de infraestrutura global da AWS, que é uma área de tecnologia importante coberta no modelo do exame Cloud Practitioner. A infraestrutura da AWS é construída em torno de regiões e zonas de disponibilidade (AZs).

Uma região da AWS é um local físico no mundo onde a AWS tem várias AZs.

As AZs consistem em um ou mais data centers discretos, cada um com energia, rede e conectividade redundantes, hospedados em instalações separadas.

Cada região é completamente independente. Cada zona de disponibilidade é isolada, mas as **zonas de disponibilidade em uma região são conectadas por meio de links de baixa latência.**

Regiões ¶

Uma região é uma área geográfica.

Cada região consiste em 2 ou mais zonas de disponibilidade.

Cada Região é projetada para ser completamente isolada das demais Regiões.

Cada região da AWS tem várias zonas de disponibilidade e data centers.

Você pode replicar dados dentro de uma região e entre regiões usando conexões de Internet públicas ou privadas.

Você mantém controle e propriedade completos sobre a região em que seus dados estão fisicamente localizados, facilitando o cumprimento dos requisitos regionais de conformidade e residência de dados.

Observe que há uma cobrança pela transferência de dados entre regiões.

Ao executar uma instância do EC2, você deve selecionar uma AMI que esteja na mesma região. Se a AMI estiver em outra região, você poderá copiar a AMI para a região que está usando.

Regiões e endpoints:

Ao trabalhar com uma instância usando a interface de linha de comando ou ações de API, você deve especificar seu endpoint regional. Para reduzir a latência de dados em seus aplicativos, a maioria dos Amazon Web Services oferece um endpoint regional para fazer suas solicitações. Um endpoint é uma URL que é o ponto de entrada para um serviço da web. Por exemplo, <https://dynamodb.us-west-2.amazonaws.com> (<https://dynamodb.us-west-2.amazonaws.com>) é um ponto de entrada para o serviço Amazon DynamoDB.

Zonas de disponibilidade

As zonas de disponibilidade são fisicamente separadas e isoladas umas das outras.

As AZs abrangem um ou mais data centers e têm conexões de rede diretas, de baixa latência, alta taxa de transferência e redundantes entre si.

Cada AZ é projetada como uma zona de falha independente.

Ao executar uma instância, você pode selecionar uma zona de disponibilidade ou deixar a AWS escolher uma para você.

Se você distribuir suas instâncias do EC2 em várias zonas de disponibilidade e uma instância falhar, você poderá projetar seu aplicativo para que uma instância em outra zona de disponibilidade possa lidar com solicitações.

Você também pode usar endereços IP elásticos para mascarar a falha de uma instância em uma zona de disponibilidade remapeando rapidamente o endereço para uma instância em outra zona de disponibilidade.

Uma zona de disponibilidade é representada por um código de região seguido por uma letra identificadora; por exemplo, us-east-1a.

Para garantir que os recursos sejam distribuídos nas zonas de disponibilidade de uma região, a AWS mapeia independentemente as zonas de disponibilidade para nomes de cada conta da AWS.

Por exemplo, a zona de disponibilidade us-east-1a para sua conta da AWS pode não ser o mesmo local que us-east-1a para outra conta da AWS.

Para coordenar as zonas de disponibilidade entre contas, você deve usar o AZ ID , que é um identificador exclusivo e consistente para uma zona de disponibilidade.

As AZs são fisicamente separadas dentro de uma região metropolitana típica e estão localizadas em planícies de inundação de menor risco.

As AZs usam UPS discretas e instalações de geração de backup no local e são alimentadas por diferentes redes de instalações independentes.

As AZs são todas conectadas de forma redundante a vários provedores de trânsito de nível 1.

Zonas locais

As zonas locais da AWS colocam computação, armazenamento, banco de dados e outros serviços selecionados da AWS mais próximos dos usuários finais.

Com as zonas locais da AWS, você pode executar facilmente aplicativos altamente exigentes que exigem latências de milissegundos de um dígito para seus usuários finais.

Cada local da zona local da AWS é uma extensão de uma região da AWS onde você pode executar seus aplicativos sensíveis à latência usando serviços da AWS, como Amazon Elastic Compute Cloud, Amazon Virtual Private Cloud, Amazon Elastic Block Store, Amazon File Storage e Amazon Elastic Load Balancing em proximidade geográfica com os usuários finais.

As zonas locais da AWS fornecem uma conexão segura e de alta largura de banda entre cargas de trabalho locais e aquelas executadas na região da AWS, permitindo que você se conecte perfeitamente a toda a gama de serviços na região por meio das mesmas APIs e conjuntos de ferramentas.

Comprimento de onda da AWS

O AWS Wavelength permite que os desenvolvedores criem aplicativos que fornecem latências de milissegundos de um dígito para dispositivos móveis e usuários finais.

Os desenvolvedores da AWS podem implantar seus aplicativos em zonas de comprimento de onda, implantações de infraestrutura da AWS que incorporam serviços de computação e armazenamento da AWS nos datacenters dos provedores de telecomunicações na borda das redes 5G e acessar facilmente a variedade de serviços da AWS na região.

O AWS Wavelength leva os serviços da AWS para a borda da rede 5G, minimizando a latência para se conectar a um aplicativo a partir de um dispositivo móvel.

Postos avançados da AWS

Os AWS Outposts trazem serviços, infraestrutura e modelos operacionais nativos da AWS para praticamente qualquer data center, espaço de co-localização ou instalação local.

Você pode usar as mesmas APIs, ferramentas e infraestrutura da AWS no local e na nuvem da AWS para oferecer uma experiência híbrida verdadeiramente consistente.

O AWS Outposts foi projetado para ambientes conectados e pode ser usado para dar suporte a cargas de trabalho que precisam permanecer no local devido à baixa latência ou às necessidades de processamento de dados locais.

Locais de borda e caches de borda regionais

Os pontos de presença são endpoints da Content Delivery Network (CDN) para o CloudFront.

Existem muito mais pontos de presença do que regiões.

Atualmente, existem mais de 200 pontos de presença.

Os caches de borda regionais ficam entre os servidores de origem do CloudFront e os pontos de presença.

Um cache de borda regional tem uma largura de cache maior do que cada um dos pontos de presença individuais.