

## Zonas de Disponibilidade

Na aula anterior vimos que as regiões são compostas por duas ou mais zonas de disponibilidade. Vamos ver agora o que são as zonas de disponibilidade.

Uma AZ é um grupo, um conjunto de data centers, altamente redundante, projetado para trabalhar de forma isolada de outras zonas de disponibilidade. Cada um desses data centers é projetado no estado da arte de redundância e segurança física, seguindo à risca as principais normas do mercado, como PCI, ISAE-3402 e certificações do Uptime Institute.

As zonas de disponibilidade ficam geograficamente distantes umas das outras para que, caso ocorra algum evento significativo em uma AZ, outra AZ dessa região não seja impactada. Quando falamos de evento estamos nos referindo, por exemplo, a desastres naturais como terremotos ou tornados. Caso alguma catástrofe aconteça, existirá pelo menos mais uma zona de disponibilidade naquela região.

Quanto à distância entre as AZs, elas ficam em uma determinada distância que possibilite uma interconexão de rede de baixa latência e, ao mesmo tempo, que fique longe o suficiente para que não seja impactada por um evento local, como um terremoto. Geralmente as AZs ficam a cerca de 150 km de distância umas das outras. No caso da América do Sul, existem data centers da AWS em São Paulo-Capital e no interior.

Apesar de muito raro, já houve casos em que uma AZ inteira ficou indisponível por alguns minutos.

Em 2012 uma AZ da região da virginia ficou indisponível por 10 minutos durante uma tempestade elétrica muito forte. E em 2013 uma das AZs em São Paulo também ficou completamente indisponível por 20 minutos devido a falhas nos geradores.

Isso reforça a importância de sempre pensar seu ambiente em multi-az. No slide coloquei os links para os anúncios dessas duas indisponibilidades que mencionei. Se tiverem curiosidade, vale a pena a leitura.

As zonas de disponibilidade são conectadas umas às outras com redes de fibra óptica de alta capacidade, que permitem a comunicação entre ambientes em diferentes AZs quase que como uma rede local, com baixíssima latência. Isso possibilita estratégias de replicação e operação em modo ativo/ativo entre múltiplas AZs. Essa é uma das grandes vantagens da AWS se compararmos com o esforço e custo necessários para implementarmos um ambiente altamente redundante em dois data centers no modelo tradicional. Muitas vezes somente o custo de uma interconexão entre data centers já inviabilizaria essa iniciativa.

Um ponto interessante sobre uma AZ é que, como já falamos ela é composta por mais de um data center e não compartilha data center com outra AZ, ou seja, um único data center não atende mais de uma zona de disponibilidade.



Isso é interessante de saber porque, mesmo que você tenha duas instâncias em uma mesma AZ, não significa que as instâncias estejam fisicamente em um mesmo data center.

Outra curiosidade é que a nomenclatura das AZs que vemos em nossa conta é aleatória. Quer dizer, no momento da criação da sua conta na AWS é definido aleatoriamente quais AZs terão a numenclatura A, B, C e assim por diante. Essa estratégia foi implementada para balancear o uso das AZs pelos clientes. Tenho certeza que a maioria de nós acaba sempre usando as AZs A, B e talvez a C, mesmo que a região que você estiver usando tenha 6 AZs, como é o caso da Virgínia. Sem essa rotação esses data centers seriam sempre os com capacidade física mais demandadas.

Isso implica que a AZ-A da minha conta não necessariamente é a mesma AZ-A da sua conta.

Atualmente a AWS possui 52 zonas de disponibilidade, divididas em 18 regiões espalhadas ao redor do mundo.

O conceito de zonas de disponibilidade é muito importante para projetarmos arquiteturas de ambientes altamente redundantes e tolerantes a falhas. É também um dos grandes diferenciais da AWS frente aos outros players de nuvem pública, porque geralmente eles podem ter até mais regiões, mas cada região não possibilita trabalhar com redundância controlada entre data centers.

