Beatriz Dassan 3º Info

1. O que é EC2?

O Amazon EC2 (Elastic Compute Cloud) é basicamente o serviço de servidores virtuais da AWS. Ele te permite criar servidores na nuvem para rodar tudo o que você precisar, sem se preocupar com a parte física dos servidores, já que tudo acontece de forma virtual. EC2 é super flexível, então você pode escalar de acordo com a demanda, seja para rodar um site, processar dados pesados ou até rodar grandes aplicações. Ele faz parte da infraestrutura de IaaS (Infraestrutura como Serviço) da AWS, o que significa que você só paga pelo que usa, e pode aumentar ou diminuir a capacidade conforme for necessário.

2. Tipos de Instâncias EC2

A AWS oferece diversos tipos de instâncias EC2, e cada uma é feita para um tipo específico de tarefa. Vou explicar os principais tipos:

Instâncias de Uso Geral (t2, t3, etc.)

Essas são as instâncias mais versáteis e são ótimas para a maioria dos aplicativos. Se você precisar de um servidor para rodar um site, banco de dados pequeno ou uma aplicação mais simples, essas instâncias são ideais. Elas são equilibradas, ou seja, têm um bom desempenho de CPU, memória e rede. As t2 e t3, por exemplo, têm um sistema de créditos de CPU que permite você "acumular" capacidade de processamento quando precisar de um pico de desempenho.

Instâncias de Computação Otimizadas (c4, c5, etc.)

Se você precisa de muito poder de processamento, as instâncias c4 e c5 são as melhores. Elas são feitas para tarefas que exigem muita CPU – como processamento em grande escala, simulações complexas, renderização de vídeos ou até servidores de jogos. Ou seja, se a sua aplicação demanda muita potência de cálculo, essas instâncias são ideais.

Instâncias de Armazenamento Otimizado (i3, etc.)

Essas instâncias são ótimas para quem precisa de armazenamento de alto desempenho, como quando você está lidando com grandes volumes de dados ou precisa de muita velocidade para leitura e gravação. As instâncias i3 são equipadas com SSD NVMe, garantindo uma performance excelente para quem precisa de alta taxa de I/O (entrada/saída).

Instâncias de Memória Otimizada (r4, r5, etc.)

Essas instâncias são feitas para tarefas que exigem muita memória, como bancos de dados em memória ou processamento de grandes volumes de dados. Se você está rodando algo que exige muito desempenho de memória, como Big Data, as instâncias r4 ou r5 são as mais indicadas.

3. Configuração de Instâncias EC2

Configurar uma instância EC2 é bem simples e envolve alguns passos. O principal é escolher o sistema operacional que você vai rodar. A AWS oferece Linux (várias distribuições como Ubuntu, CentOS, Amazon Linux) e Windows Server.

Depois de escolher o sistema operacional, você vai configurar o tipo de instância que atende às suas necessidades (se vai precisar de mais CPU, mais memória, etc.), e também definir questões como rede, segurança e armazenamento. A AWS tem várias opções para te ajudar a garantir que sua instância vai estar configurada de acordo com o que você precisa.

4. Custo e Otimização de Instâncias EC2

O custo das instâncias EC2 pode variar bastante, mas existem algumas maneiras de otimizar e pagar menos:

Instâncias Spot

As instâncias spot são um jeito de aproveitar a capacidade ociosa da AWS, e você paga bem mais barato por isso. A grande vantagem é o desconto de até 90% comparado com as instâncias sob demanda. O problema é que essas instâncias podem ser interrompidas a qualquer momento, então elas são melhores para tarefas que podem ser pausadas ou que não exigem continuidade, como processamento em segundo plano.

Instâncias Reservadas

Se você sabe que vai precisar de uma instância por um longo período (1 ou 3 anos), pode optar pelas instâncias reservadas. Ao fazer um compromisso de longo prazo, você consegue um desconto bem significativo – de até 75% – em relação ao preço das instâncias sob demanda. Isso é perfeito para quem tem uma carga de trabalho constante e previsível.

Autoescala

O Auto Scaling é uma funcionalidade que ajusta automaticamente o número de instâncias que você está usando com base na demanda. Por exemplo, se o tráfego do seu site aumentar, o Auto Scaling pode lançar novas instâncias para acompanhar esse aumento, e quando o tráfego diminuir, ele vai reduzir o número de instâncias, o que ajuda a economizar recursos. Isso é ótimo para empresas que têm picos de tráfego e não querem gastar mais do que o necessário.

Essas estratégias podem ajudar muito a otimizar o custo das suas instâncias EC2, garantindo que você só pague pelo que realmente usa.