

AMAZON ELASTIC COMPUTE CLOUD

O Amazon EC2 (Elastic Compute Cloud) oferece uma impressionante variedade de mais de 500 tipos de instâncias, permitindo aos usuários escolher a configuração exata de processador, armazenamento, rede e sistema operacional que melhor atenda às suas necessidades específicas de carga de trabalho. Este nível de customização torna o EC2 uma ferramenta extremamente versátil e poderosa para hospedar uma ampla gama de aplicações na nuvem. Vamos explorar como o EC2 está organizado e como você pode selecionar o tipo ideal de instância para suas necessidades.

O Que é uma Instância EC2?

Uma instância EC2 é basicamente uma máquina virtual (VM) que roda na infraestrutura de cloud da AWS. Quando você lança uma instância, está alugando capacidade virtual de computação da AWS por um período determinado. Essa instância age como um servidor virtual dedicado capaz de rodar aplicações diversas.

Escolhendo o Tipo de Instância

Ao executar uma instância, a primeira decisão que você deve tomar é a configuração de hardware. Isso é feito escolhendo um tipo de instância específico. Cada tipo de instância oferece uma combinação diferente de capacidade de computação, memória, armazenamento e capacidades de rede, permitindo que você ajuste suas escolhas ao perfil específico de desempenho necessário para sua aplicação.

Famílias e Sub-Famílias de Instâncias

As instâncias são categorizadas em famílias e sub-famílias, cada uma otimizada para diferentes tipos de cargas de trabalho:

Famílias de Uso Geral: São equilibradas em termos de CPU, memória e recursos de rede. Exemplos incluem as instâncias t3 e m5.

Famílias Otimizadas para Computação: Oferecem alto desempenho de CPU em relação à memória e são ideais para cargas de trabalho de computação científica e jogos. Exemplos são as instâncias c5 e c6g.

Famílias Otimizadas para Memória: Projetadas para aplicações que processam grandes conjuntos de dados em memória. Exemplos incluem as instâncias r4 e r5.

Famílias Otimizadas para Armazenamento: Fornecem I/O de disco rápido, ideais para bancos de dados de alta performance. Exemplos são as instâncias i3 e d2.

Famílias Aceleradas por GPU: São adequadas para tarefas que requerem processamento gráfico intenso, como visualização de dados e aprendizado de máquina. Exemplos incluem p3 e g4.

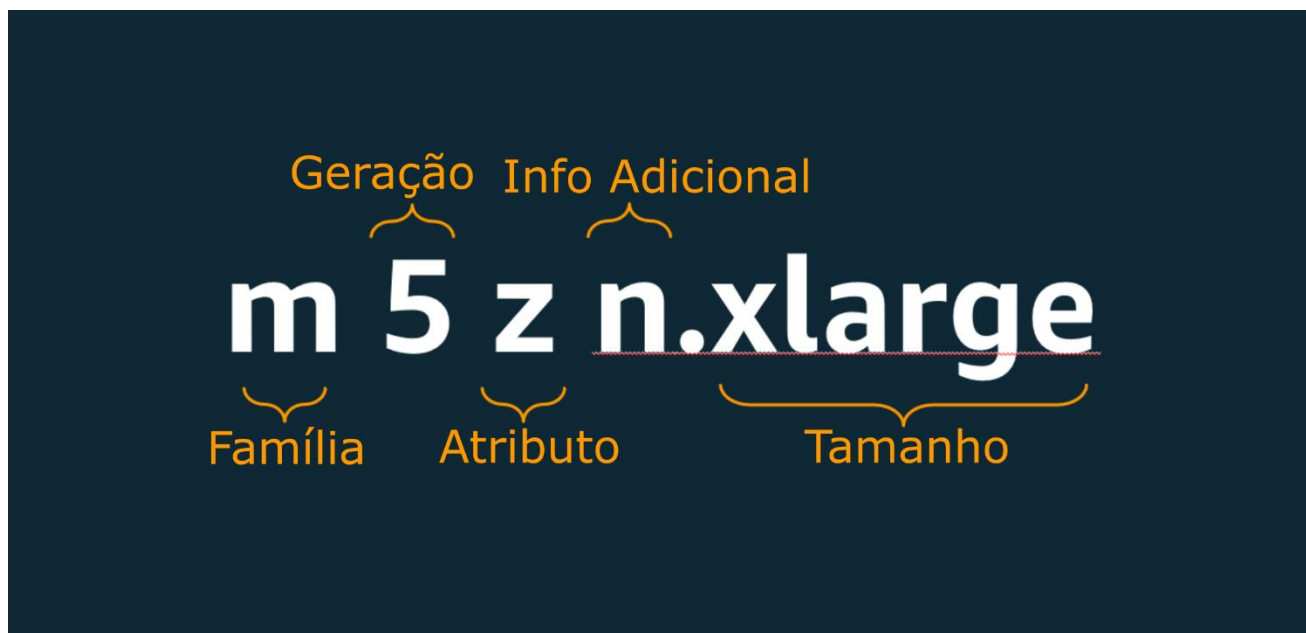
vCPUs

Uma unidade de processamento central virtual (vCPU) é uma medida de capacidade de processamento. Para a maioria dos tipos de instância, uma vCPU representa um thread do núcleo físico do CPU subjacente. Por exemplo, se um tipo de instância possui dois núcleos de CPU e dois threads por núcleo, ele terá quatro vCPUs.

Como Nomear uma Instância Indica Suas Funções

O nome de cada tipo de instância não é arbitrário; ele faz referência às funções e aos recursos de hardware disponíveis para aquela instância. Por exemplo, a letra inicial (como m de uso geral, c de computação, r de memória, i de I/O intensivo) indica a família de otimização, enquanto os números e letras subsequentes fornecem informações sobre a geração e características específicas.

Decodificando os nomes das instâncias



Adaptado de: AWS

Família da Instância

A letra na primeira posição representa a família da instância. Em muitos casos, você pode lembrar disso pelo que a letra representa. M é uma subfamília de instâncias de propósito geral. M pode ser lembrada como a instância a ser usada para propósitos principais, muitos, maioria, e assim por diante. Existem vários mnemônicos usados para lembrar as famílias de instâncias, e qualquer pesquisa no navegador lhe dará uma variedade de opções para escolher.

Geração da Instância

Este número representa a geração do processador, sendo que números mais altos indicam a geração mais recente. Gerações mais antigas são suportadas para aplicações que foram otimizadas para uma geração específica. No entanto, é uma boa prática atualizar suas instâncias à medida que novas são lançadas. A geração mais recente geralmente proporciona uma melhoria de desempenho, uma economia de custos ou ambos. Números mais altos neste campo representam gerações mais novas.

Atributo da Instâncias

Nem todas as instâncias têm uma opção de atributo. Este atributo fornece informações adicionais sobre as capacidades da instância. No caso da instância m5z, o "z" indica alta frequência e nos diz que essa instância é ideal para desempenho de thread único de alta frequência.

Informação Adicional da Instância

Nem todas as instâncias têm um atributo adicional que detalha para o que a instância é otimizada.

No caso da instância m5zn, aprendemos que "m" é a família de instâncias, "5" é a geração do processador, "z" significa que é otimizada para alto desempenho, e agora o "n" representa rede e indica que essa instância é otimizada para alta taxa de transferência e baixa latência de rede. Combinada com o atributo "z" para alta computação, essa instância é de alto desempenho para cargas de trabalho como jogos, modelagem e HPC.

Tamanho

Isso representa o tamanho da instância. Quanto maior o tamanho, mais recursos, como CPU e memória. Os tamanhos são baseados na combinação de recursos de hardware usados para essa instância.

Tamanho das Instâncias

As instâncias EC2 são dimensionadas com base nos recursos de hardware combinados consumidos por esse tipo de instância. Isso significa que o tamanho é a capacidade total configurada de vCPU, memória, armazenamento e rede. Os tamanhos variam de nano até mais de 32xlarge, com uma instância de tamanho nano usando a menor quantidade de recursos de hardware e a instância 32xlarge usando a maior quantidade de recursos de hardware (128 vCPU e 1.024 GiB de memória). Vamos dar uma olhada rápida em um gráfico de comparação de tamanhos para ajudá-lo a entender como o hardware alocado corresponde ao tamanho da instância. No caso da família de instâncias de propósito geral T, a alocação de vCPU permanece a mesma, mas a memória dobra a cada tamanho maior.

Família da Instância	Tamanho da Instância	vCPU	Memória (GiB)
Propósito Geral	t4g.nano	2	0,5
Propósito Geral	t4g.micro	2	1
Propósito Geral	t4g.small	2	2
Propósito Geral	t4g.medium	2	4
Propósito Geral	t4g.large	2	8

Um gibibyte (GiB) é uma unidade de medida que representa a capacidade de tamanho. Às vezes, a medida gigabyte (GB) e GiB são usadas de forma intercambiável, mas isso não é preciso. Um GiB é definido como 1024³ bytes, enquanto um GB é definido como 1000³ bytes.

Por que importa se você usa GiB ou GB? Para pequenas quantidades de capacidade, os números de GiB versus GB variam apenas ligeiramente, mas à medida que você escala para a capacidade massiva de uma solução em nuvem, a variação de tamanho entre os números cresce exponencialmente. Para obter mais informações, pesquise GB vs GiB em seu navegador favorito e leia mais sobre como os números diferem.

Tipos de instâncias EC2

Para obter informações sobre cada família de instâncias e a divisão individual dos componentes, como CPU, memória e armazenamento, visite o link: [Amazon EC2 tipos de instâncias - AWS](#)