OS TIPOS DE INSTÂNCIA DO AMAZON EC2

Os tipos de instância do Amazon EC2, são otimizados para tarefas diferentes.

1. **Instâncias de uso geral**

**Instâncias de uso geral** equilibram os recursos de computação, memória e rede. Você pode usá-las para diversas cargas de trabalho, como:

* servidores de aplicativos
* servidores de jogos
* servidores de back-end para aplicativos empresariais
* bancos de dados pequenos e médios

Suponha que você tenha um aplicativo no qual as necessidades de recursos para computação, memória e rede sejam praticamente equivalentes. Você pode executar esse aplicativo em uma instância de uso geral porque ele não precisa de otimização em nenhuma área de recurso único.

1. **Instâncias otimizadas para computação**

**Instâncias otimizadas para computação** são ideais para aplicativos vinculados à computação que se beneficiam de processadores de alto desempenho. Assim como instâncias de uso geral, você pode usar instâncias otimizadas para computação para cargas de trabalho, como servidores web, de aplicativos e de jogos.  
  
No entanto, a diferença é que aplicativos otimizados para computação são ideais para servidores web de alto desempenho, servidores de aplicativos de computação intensiva e servidores de jogos dedicados. Você também pode usar instâncias otimizadas para computação para cargas de trabalho de processamento em lote, com o processamento de muitas transações em um único grupo.

1. **Instâncias otimizadas para memória**

**Instâncias otimizadas para memória**

são projetadas para fornecer desempenho rápido para cargas de trabalho que processam grandes conjuntos de dados na memória. Na computação, a memória é uma área de armazenamento temporário. Ela contém todos os dados e instruções de que uma unidade central de processamento (CPU) precisa para conseguir realizar ações. Antes que um programa de computador ou aplicativo possa ser executado, ele é carregado do armazenamento para a memória. Esse processo de pré-carregamento dá à CPU acesso direto ao programa de computador.  
  
Suponha que você tenha uma carga de trabalho que precise que grandes quantidades de dados sejam pré-carregados antes de executar um aplicativo. Esse cenário pode ser de um banco de dados de alto desempenho ou uma carga de trabalho que envolva a execução de processamento em tempo real de uma grande quantidade de dados não estruturados. Nesses tipos de casos de uso, considere usar uma instância otimizada para memória. As instâncias otimizadas para memória permitem que você execute cargas de trabalho com altas necessidades de memória e tenha um ótimo desempenho.

1. **Instâncias de computação acelerada**

**Instâncias de computação acelerada** usam aceleradores de hardware, ou coprocessadores, para executar algumas funções de forma mais eficiente do que é possível em um software executado em CPUs. Exemplos dessas funções são cálculos de números com vírgula flutuante, processamento de gráficos e correspondência de padrões de dados.  
  
Na computação, um acelerador de hardware é um componente que pode agilizar o processamento de dados. As instâncias de computação acelerada são ideais para cargas de trabalho, como aplicativos gráficos e streaming de jogos e de aplicativos.

1. **Instâncias otimizadas para armazenamento**

**As instâncias otimizadas para armazenamento** são projetadas para cargas de trabalho que exigem alto acesso sequencial de leitura e gravação a grandes conjuntos de dados no armazenamento local. Exemplos de cargas de trabalho adequadas para instâncias otimizadas para armazenamento são sistemas de arquivos distribuídos, aplicativos de data warehouse e sistemas de processamento de transações on-line de alta frequência (OLTP).  
  
Na computação, o termo operações de entrada/saída por segundo (IOPS) é uma métrica que mensura o desempenho de um dispositivo de armazenamento. Ela indica quantas operações diferentes de entrada ou saída um dispositivo pode executar em um segundo. As instâncias otimizadas para armazenamento foram projetadas para fornecer dezenas de milhares de IOPS aleatórias e de baixa latência para aplicativos.  
  
Imagine as operações de entrada como dados colocados em um sistema, como registros inseridos em um banco de dados. Uma operação de saída são dados gerados por um servidor. Um exemplo de saída pode ser a análise realizada nos registros em um banco de dados. Se você tiver um aplicativo com alto requisito de IOPS, uma instância otimizada para armazenamento poderá fornecer melhor desempenho em relação a outros tipos de instâncias não otimizados para esse tipo de caso de uso.