

Instâncias EC2 da AWS

Amazon Elastic Compute Cloud (EC2) é um serviço que fornece capacidade de computação escalável na nuvem. Com o EC2, você pode configurar, iniciar e gerenciar servidores virtuais chamados instâncias.

- **Primeiros serviços ofertados pela AWS:** O EC2 foi um dos primeiros serviços lançados pela AWS, e continua a ser amplamente utilizado para hospedagem e execução de aplicativos.
- **Free Tier:** Para novos usuários, a AWS oferece uma camada gratuita com 750 horas de uso de instância EC2 (micro instâncias) por mês durante o primeiro ano, permitindo experimentar o serviço sem custo inicial.
- **Suporte a múltiplos sistemas operacionais:** EC2 permite escolher entre diversos sistemas operacionais, incluindo distribuições Linux (como Ubuntu, Amazon Linux, CentOS, Red Hat) e Windows Server, além de suportar AMIs (Amazon Machine Images) personalizadas para necessidades específicas.

Principais Características do EC2

1. **Escalabilidade e Elasticidade:** EC2 permite redimensionar a capacidade de computação com facilidade, o que é útil para lidar com variações de carga.
2. **Vários tipos de instância:** EC2 oferece tipos de instâncias otimizados para diferentes cargas de trabalho, como instâncias otimizadas para CPU, memória, GPU e armazenamento.
3. **Opções de Preço:**
 - **Sob demanda:** Pague apenas pelo tempo usado, sem compromissos de longo prazo.
 - **Reservadas:** Compromisso de uso de 1 ou 3 anos, com descontos significativos.
 - **Spot Instances:** Oferecem capacidade de computação a preços reduzidos, ideais para tarefas não críticas.

Componentes do Amazon EC2

Para configurar e gerenciar instâncias EC2 com segurança e eficiência, alguns componentes principais são essenciais:

1. **VPC (Virtual Private Cloud):**
 - Cria uma rede virtual isolada na nuvem AWS, onde você pode executar suas instâncias EC2.
 - Permite definir faixas de IP, sub-redes, tabelas de roteamento e gateways para controlar o tráfego de entrada e saída da sua rede.
2. **Subnet:**
 - Segmento dentro da VPC, usado para dividir a rede em sub-redes lógicas.
 - Pode ser pública (acessível pela internet) ou privada (isolada da internet), dependendo dos requisitos de segurança.
3. **Elastic IP Address:**
 - Endereço IPv4 fixo para sua instância EC2, permitindo o reassociamento entre instâncias.
 - Útil para manter o mesmo IP público, mesmo que a instância seja reiniciada ou substituída.
4. **IAM (Identity and Access Management):**
 - Permite gerenciar usuários e permissões para acesso aos recursos EC2 e outros serviços da AWS.
 - **Roles** podem ser atribuídas às instâncias para conceder permissões temporárias, sem a necessidade de armazenar credenciais.
5. **Security Group:**
 - Atua como um firewall virtual para controlar o tráfego de entrada e saída da instância.
 - Define regras para permitir ou negar acessos em portas e protocolos específicos.
6. **Key Pair:**
 - Par de chaves pública e privada utilizado para autenticar o acesso SSH às instâncias.
 - A chave pública é armazenada na AWS, enquanto a chave privada fica com o usuário para acesso seguro.

Esses componentes combinados ajudam a configurar, proteger e controlar o acesso às suas instâncias EC2, garantindo isolamento e segurança adequados para suas cargas de trabalho.

Modelos de Preços para Amazon ECS (Elastic Container Service)

1. On Demand:

- Pague apenas pelo que usar, sem contratos a longo prazo.
- Ideal para cargas de trabalho imprevisíveis ou de curto prazo.
- É a opção mais cara entre os modelos de preços.

2. Instâncias Reservadas:

- Contratação por períodos de 1 ou 3 anos, com descontos de até 75%.
- Ideal para cargas de trabalho estáveis e previsíveis, onde o uso prolongado justifica o custo reduzido.

3. Instâncias Spot:

- Oferecem descontos de até 90% em relação ao preço On Demand.
- Ideais para cargas de trabalho que podem ser interrompidas, pois a AWS pode revogar o recurso quando precisar.
- Excelente para tarefas em lotes e cargas de trabalho tolerantes a falhas.

4. Instâncias Dedicadas:

- Servidor físico dedicado ao cliente, sem compartilhamento de hardware com outros usuários.
- Útil para atender a requisitos de conformidade ou para obter um isolamento físico completo.

5. Hosts Dedicados:

- Similar às instâncias dedicadas, mas oferecem controle total sobre a colocação e o uso de instâncias no host físico.
- Ideal para licenciamento de software por CPU ou para necessidades específicas de conformidade.

6. Saving Plans:

- Planos flexíveis de economia para uso de longo prazo de serviços como Lambda, EC2 e Fargate.
- Com pagamento por hora e opção de contrato por 1 a 3 anos, esses planos oferecem custos reduzidos.
- Permite comprometimento de uso em termos de horas por ano, garantindo previsibilidade e economia.

Essas opções oferecem flexibilidade para alinhar os custos ao perfil de utilização e garantir economia para workloads de diferentes características no ECS.

Na AWS, a **cobrança das instâncias EC2** varia conforme o sistema operacional e o tipo de armazenamento:

Cobrança por Instância:

1. **Licenças Comerciais (Windows, Red Hat):**
 - Cobradas por hora, com o mínimo de 1 hora, o que significa que se você iniciar e parar uma instância com uma dessas licenças, será cobrado pela hora completa.
2. **Sistemas Livres (Linux e similares):**
 - Cobrados por segundo, com o mínimo de 1 minuto. Assim, você paga apenas pelo tempo exato de uso.
3. **Elastic Block Store (EBS):**
 - Cobrados por segundo, com o mínimo de 1 minuto.
 - O **armazenamento EBS continua sendo cobrado mesmo com a instância desligada**, pois os dados permanecem no disco e ocupam espaço de armazenamento.

Essa flexibilidade permite economizar em custos, principalmente para instâncias Linux e armazenamento EBS, pois só são cobrados pelos recursos que estão realmente em uso.

Tipos de Instâncias Reservadas

1. Standard Reserved Instances:

- **Compromisso de 1 ou 3 anos:** Contratar essas instâncias por um longo período permite economia de até 75% em comparação com as instâncias On-Demand.
- **Flexibilidade de Modificação:**
 - Pode alterar a instância de uma **Availability Zone (AZ)** para outra dentro da mesma região.
 - Permite **mudar o tamanho** da instância (dentro da mesma família de instâncias).
 - **Rede:** Pode alterar o tipo de rede.
- **Família e Sistema Operacional:** Não é possível mudar a família da instância nem o sistema operacional.

2. Convertible Reserved Instances:

- **Compromisso de 1 ou 3 anos:** Também oferece economia, mas com maior flexibilidade em relação às Standard Reserved Instances.
- **Flexibilidade de Modificação:**
 - Permite mudar a instância entre diferentes **Availability Zones**.
 - Possibilita **alterar o tamanho e tipo de rede** da instância.
 - **Alteração de Família e Sistema Operacional:** Oferece mais flexibilidade, pois você pode **mudar a família da instância** (por exemplo, de **t3** para **m5**) e **alterar o sistema operacional**.
- **Benefício Principal:** Mais adaptável para cenários onde as demandas podem mudar, proporcionando maior versatilidade ao longo do período do contrato.

Comparação Rápida

Característica	Standard Reserved Instances	Convertible Reserved Instances
Período de Contrato	1 a 3 anos	1 a 3 anos
Desconto	Até 75%	Um pouco menor que Standard
Mudança de Tamanho da Instância	Sim	Sim
Alteração de Família	Não	Sim
Alteração do Sistema Operacional	Não	Sim
Alteração de AZ	Sim	Sim
Alteração do Tipo de Rede	Sim	Sim

Essas opções ajudam a otimizar custos para empresas que têm cargas de trabalho estáveis e previsíveis, com a Convertible sendo indicada para casos em que é preciso ajustar a configuração ao longo do tempo.

Os **Saving Plans** da AWS são uma alternativa econômica ao modelo On-Demand, oferecendo descontos em troca de um compromisso de uso por um período de 1 a 3 anos. Eles ajudam a reduzir os custos para cargas de trabalho flexíveis e previsíveis, com opções voltadas para serviços específicos ou com maior flexibilidade de aplicação.

Tipos de Saving Plans

1. Compute Savings Plans:

- **Cobertura:** Aplicável ao **AWS Lambda, EC2 e AWS Fargate**.
- **Flexibilidade:** Pode ser usado em **qualquer região, qualquer família de instâncias, qualquer tamanho e qualquer sistema operacional**.
- **Compromisso:** 1 a 3 anos.
- **Benefício:** Maior flexibilidade para modificar configurações ou regiões sem perder os benefícios do plano.

2. EC2 Savings Plans:

- **Cobertura:** Aplicável **somente ao EC2**.
- **Especificação de Região e Família:** Exige definir a **região** e a **família da instância** (como **t3** ou **m5**).
- **Flexibilidade:** Funciona com qualquer **tamanho de instância** e **qualquer sistema operacional** dentro da família e região escolhidas.
- **Compromisso:** 1 a 3 anos.
- **Benefício:** Maior desconto em comparação ao Compute Savings Plans, mas com menos flexibilidade.

Comparação Rápida

Característica	Compute Savings Plans	EC2 Savings Plans
Cobertura de Serviços	Lambda, EC2, Fargate	Somente EC2
Flexibilidade de Região	Qualquer região	Deve especificar a região
Flexibilidade de Família	Qualquer família	Deve especificar a família
Tamanho da Instância	Qualquer tamanho	Qualquer tamanho dentro da família

Sistema Operacional	Qualquer	Qualquer
Compromisso	1 a 3 anos	1 a 3 anos
Desconto	Menor que EC2 Savings Plans	Maior que Compute Savings Plans

Esses planos são ideais para empresas que querem economizar com um comprometimento mínimo e ainda manter flexibilidade de uso, especialmente quando usam várias instâncias ou regiões.

O ciclo de vida de uma Instância EC2

O ciclo de vida de uma **Instância EC2** envolve vários estados pelos quais a instância pode passar, dependendo das ações do usuário e do sistema. Aqui está uma visão geral do processo:

1. **Pending:**
 - Ocorre quando a instância é iniciada. A AWS configura a instância, realiza verificações (como permissões de IAM) e provisiona os recursos.
 - Esse é o estado de inicialização antes de a instância entrar em operação.
2. **Running:**
 - A instância está ativa e pronta para uso. Durante o estado "Running", é possível:
 - **Reboot:** reinicia a instância sem alterar o seu estado.
 - **Stopping:** interrompe a instância, o que leva ao estado "Stopped".
 - **Shutdown:** encerra a instância de forma definitiva, movendo-a para o estado "Terminated".
3. **Stopping:**
 - A instância está em processo de interrupção e logo mudará para o estado "Stopped". No estado "Stopping", o sistema fecha os processos antes de interromper completamente.
4. **Stopped:**
 - A instância está interrompida, mas seus dados são preservados.
 - Nesse estado, é possível reiniciar a instância, o que a leva de volta para o estado "Pending" e depois para "Running".
 - Também é possível optar por finalizar a instância, movendo-a para o estado "Terminated".
5. **Terminated:**
 - A instância foi encerrada permanentemente, e seus dados são apagados (a menos que volumes EBS persistentes sejam configurados). Uma vez em "Terminated", a instância não pode mais ser reiniciada.

Esse ciclo de vida permite que os usuários tenham flexibilidade para gerenciar o estado das instâncias EC2 de acordo com suas necessidades, reduzindo custos quando as instâncias não estão em uso (estado "Stopped") e encerrando-as de forma definitiva (estado "Terminated") quando não forem mais necessárias.

Posicionamento de EC2

O AWS EC2 permite configurar **Placement Groups** para otimizar o posicionamento de instâncias, melhorando a conectividade, a resiliência ou a redundância, de acordo com a necessidade. Existem três tipos principais de Placement Groups:

1. Cluster Placement Group:

- **Posicionamento:** As instâncias são colocadas próximas uma da outra, geralmente no mesmo rack dentro de uma única Zona de Disponibilidade (AZ).
- **Uso:** Ideal para aplicações que exigem baixa latência e alta taxa de transferência de rede entre instâncias.
- **Limitações:** Limita a resiliência, pois qualquer problema que afete o rack impacta todas as instâncias.

2. Spread Placement Group (Disseminar):

- **Posicionamento:** Cada instância é posicionada em um rack diferente, distribuída em até sete Zonas de Disponibilidade na mesma região.
- **Uso:** Indicado para aplicações críticas onde cada instância deve estar isolada para reduzir o risco de falha simultânea.
- **Limitações:** Máximo de sete instâncias por grupo por Zona de Disponibilidade.

3. Partition Placement Group:

- **Posicionamento:** As instâncias são distribuídas em diferentes partições dentro da mesma Zona de Disponibilidade ou em várias Zonas.
- **Uso:** Adequado para grandes clusters de dados que precisam ser particionados para alta resiliência, como bancos de dados distribuídos.
- **Partições:** As instâncias de cada partição não compartilham o mesmo rack, melhorando a tolerância a falhas.

Essas opções permitem uma configuração flexível para atender a requisitos específicos de desempenho e resiliência, otimizando os recursos de acordo com a carga de trabalho de cada aplicação.