VPC – Virtual Private Cloud

Neste vídeo, você conhecerá a VPC (Amazon Virtual Private Cloud), a distribuição de endereçamento e recursos de conectividade com a internet.

Imagine os milhões de clientes que usam os serviços AWS e os milhões de recursos criados por eles, como as instâncias do Amazon EC2. Sem limites de acesso para todos esses recursos, o tráfego de rede fluiria sem restrições entre eles, permitindo que todos os recursos de todos os clientes tenham acesso uns ao outros.

Um serviço de rede que pode ser usado para definir limites para seus recursos AWS é o Amazon Virtual Private Cloud (Amazon VPC).

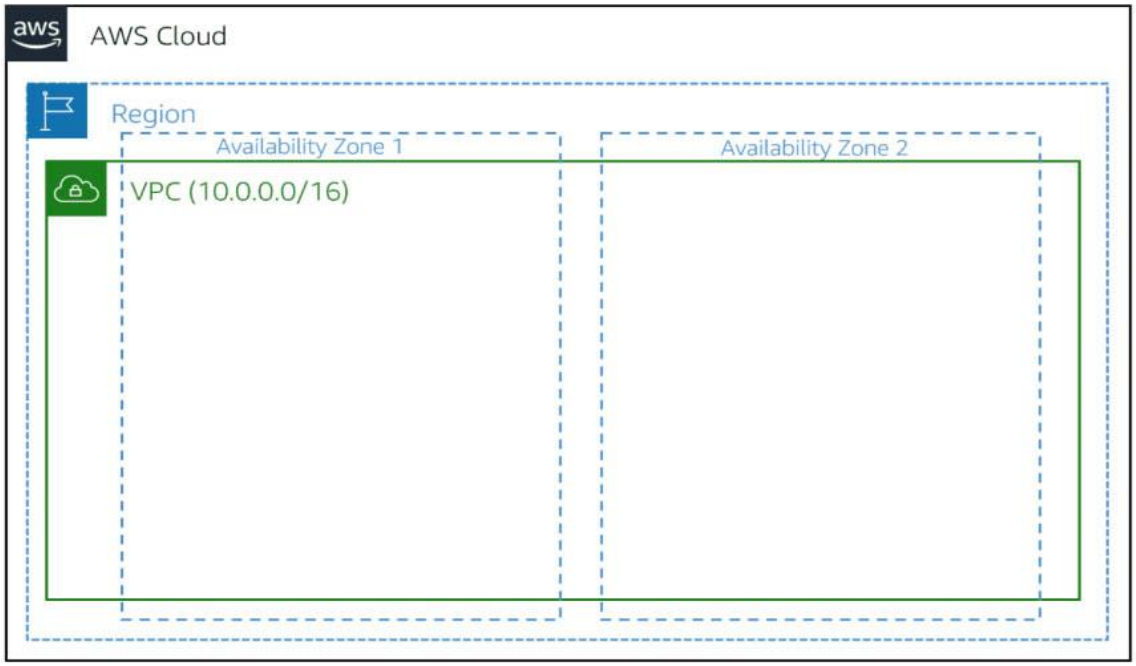
Esse serviço permite que você provisione uma seção isolada da nuvem AWS. Nessa seção, você pode executar os recursos em uma rede virtual que definir. Em uma Virtual Private Cloud (VPC), você pode organizar seus recursos em sub-redes. Uma sub-rede é uma seção de uma VPC que pode conter recursos como instâncias do Amazon EC2.

Ao criar uma VPC, você deve escolher três fatores principais:

1. O nome do VPC.
2. A região onde o VPC vai ser provisionado, já que cada VPC abrange várias zonas de disponibilidade na região selecionada.
3. O intervalo de IP para a VPC na notação CIDR. Isso determina o tamanho da sua rede. Cada VPC pode ter até quatro intervalos de IP /16.

Usando essas informações, a AWS provisionará uma rede e endereços IP para essa rede.

VPCs ficam dentro da região escolhida, mas podem se estender por múltiplas zonas de disponibilidade.



Zonas de disponibilidade das VPCs.

Sub-redes de VPC

Depois de criar sua VPC, você deve criar sub-redes dentro da rede. Pense nelas como redes menores dentro de sua rede base – ou redes locais virtuais (VLANs) em uma rede local tradicional. Em uma rede local, o caso de uso típico para sub-redes é isolar ou otimizar o tráfego de rede. Na AWS, essas sub-redes são usadas para fornecer alta disponibilidade e opções de conectividade para seus recursos.

Ao criar uma sub-rede, você deve especificar:

- A VPC na qual você deseja que sua sub-rede resida;.
- A zona de disponibilidade na qual você deseja que sua sub-rede resida.
- O bloco CIDR para sua sub-rede, que deve ser um subconjunto do bloco VPC CIDR.

Ao iniciar uma instância do EC2, você a inicia dentro de uma sub-rede, que estará localizada dentro da zona de disponibilidade que você escolher.

As sub-redes são basicamente de dois tipos: públicas e privadas.

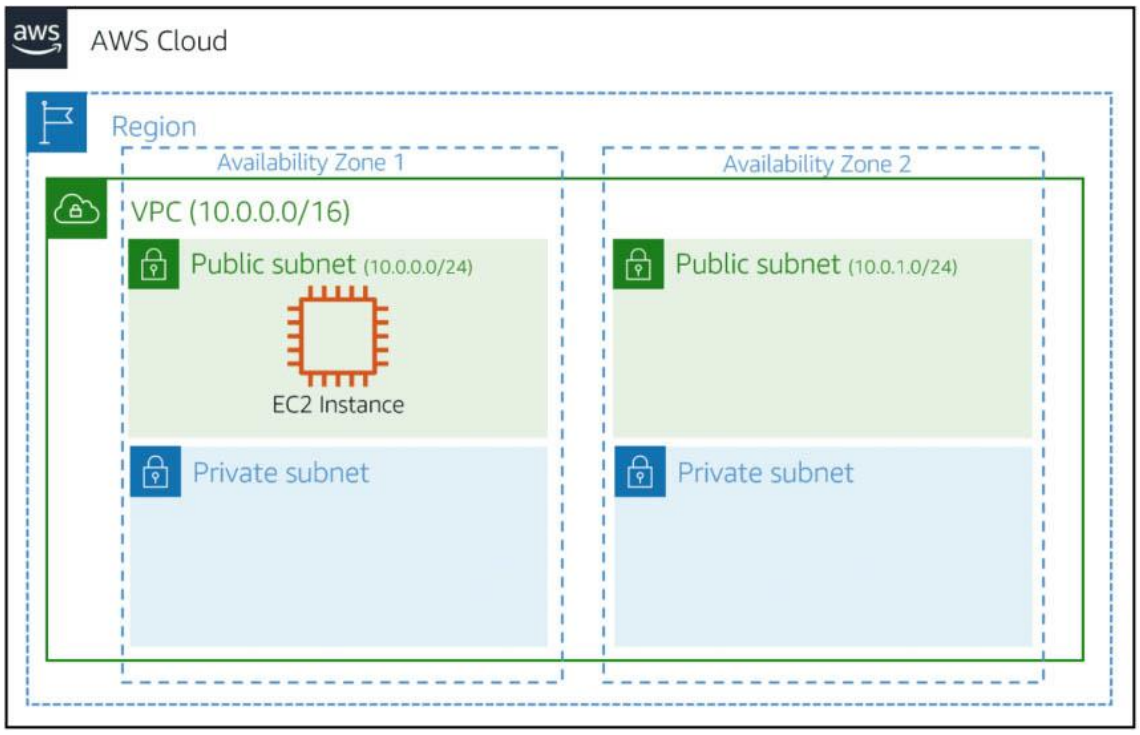
Sub-redes públicas

Possuem um gateway de internet (que veremos mais adiante) que permite o acesso à internet diretamente, de origem e destino, para as instâncias e recursos que foram provisionados nela.

Sub-redes privadas

Não possuem um gateway e o acesso à internet pode acontecer por intermédio de um gateway NAT ou uma instância NAT, provisionada em uma sub-rede pública.

A seguir, podemos ver as divisões em sub-rede dentro de uma mesma VPC e em diferentes zonas de disponibilidade.

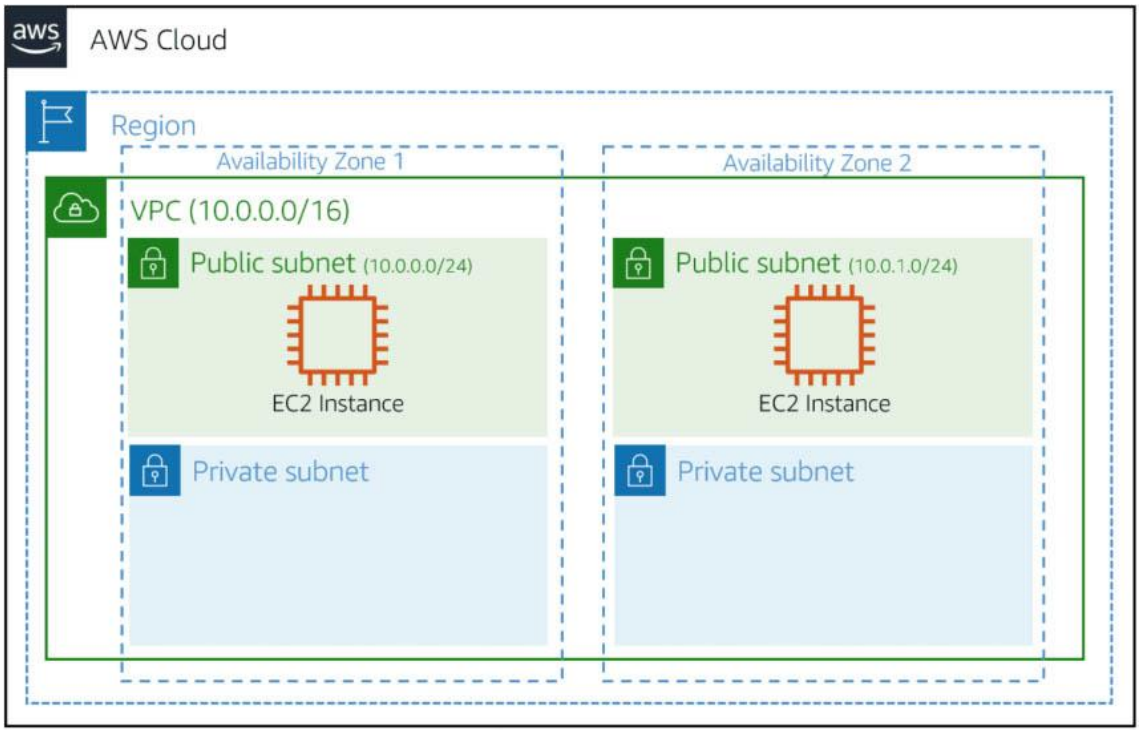


Divisões em sub-rede dentro de uma mesma VPC.

Alta disponibilidade em VPC

Ao criar suas sub-redes, tenha em mente a alta disponibilidade. Para manter a redundância e a tolerância a falhas crie, ao menos, duas sub-redes configuradas em duas zonas de disponibilidade. Lembre-se de que “tudo falha o tempo todo”. Com a rede de exemplo, se uma das AZs falhar, você ainda terá seus recursos disponíveis em outra AZ como backup/redundância. Recursos como instâncias EC2 não usufruem automaticamente de alta disponibilidade

com múltiplas AZs pois, ao lançar um novo servidor virtual, este estará localizado em apenas uma AZ determinada. Aplicações hospedadas em instâncias EC2 e que possuem capacidade de escala horizontal (múltiplos servidores, com a mesma aplicação, funcionando como um cluster) podem usufruir de alta disponibilidade se for provisionada mais de uma máquina virtual em AZs diferentes e essas máquinas utilizarem, como exemplo, um balanceador de carga para distribuição do tráfego.



Sub-redes distribuidas em duas zonas de disponibilidade.

Endereço IP elástico

Neste vídeo, você verá como se beneficiar dos IPs elásticos.

Um endereço IP elástico (Elastic IP) é um endereço IPv4 público e estático projetado para computação em nuvem dinâmica. Você pode associar um endereço IP elástico a qualquer instância ou interface de rede em qualquer VPC em sua conta. Com um endereço IP elástico, você pode mascarar a falha de uma instância remapeando rapidamente o endereço para outra instância em sua VPC. A seguir, algumas considerações sobre IPs elásticos:

1. Um endereço IP elástico pode ser associado a uma única instância ou interface de rede por vez.
2. Você pode mover um endereço IP elástico de uma instância ou interface de rede para outra.

3. Se você associar um endereço IP elástico à interface de rede primária de sua instância, seu endereço IPv4 público atual (se houver) será liberado para o pool de endereços IP públicos. Se você desassociar o endereço IP elástico, a interface de rede primária receberá automaticamente um novo endereço IPv4 público em alguns minutos. Isso não se aplica se você tiver anexado uma segunda interface de rede à sua instância.
4. Existe uma cobrança de 0,005 dólares por hora quando eles não estão associados a uma instância em execução ou quando estão associados a uma instância interrompida ou a uma interface de rede não conectada. Enquanto sua instância estiver em execução, você não será cobrado por um endereço IP elástico associado à instância, mas por quaisquer endereços IP elásticos adicionais (a partir do segundo na mesma instância) associados à instância.
5. Endereços IP elásticos para IPv6 não são suportados.

IPs reservados

Para que a AWS configure sua VPC adequadamente, ela reserva cinco endereços IP em cada sub-rede. Esses endereços IP são usados para roteamento, Domain Name System (DNS) e gerenciamento de rede.

Exemplo

Considere uma VPC com o intervalo de IP 10.0.0.0/22. A VPC inclui um total de 1.024 endereços IP. Isso é dividido em quatro sub-redes de tamanho igual, cada uma com um intervalo de IP /24 com 256 endereços IP. De cada um desses intervalos de IP, existem apenas 251 endereços IP que podem ser usados, já que a AWS reserva cinco. A primeira subnet pode ser 10.0.0.0/24.

Vejamos como são reservados pela AWS os cinco endereços IP:

Endereço IP	Propósito
10.0.0.0	Endereço de rede
10.0.0.1	Roteador da VPC
10.0.0.2	Servidor DNS

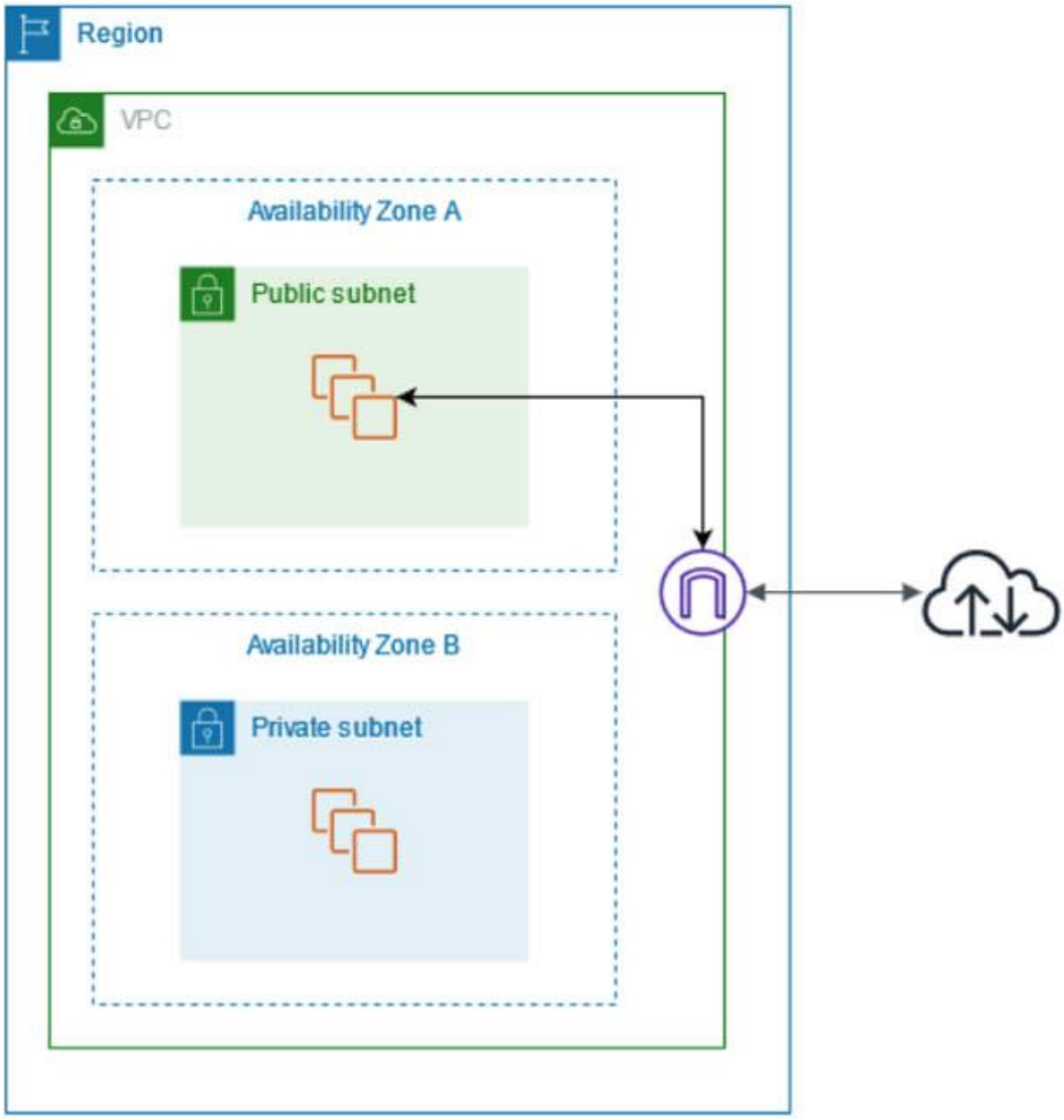
Endereço IP	Propósito
10.0.0.3	De uso futuro
10.0.0.255	Endereço de broadcast

Especificação de IPs e seus propósitos no AWS.
Gustavo Ribeiro

Os cinco endereços IP reservados podem afetar a forma como você projeta sua rede. Um ponto de partida comum para quem é novo na nuvem é criar uma VPC com um intervalo de IP de /16 e criar sub-redes com um intervalo de IP de /24. Isso fornece uma grande quantidade de endereços IP para trabalhar nos níveis de VPC e sub-rede.

Gateway de internet (internet gateway)

Para habilitar a conectividade para sua VPC, você deve criar um gateway de internet. Pense nele como algo semelhante a um modem. Da mesma forma que um modem conecta seu computador à internet, o gateway conecta sua VPC. Ao contrário do seu modem em casa, que às vezes fica inativo ou offline, o gateway de internet é altamente disponível e escalável, abrangendo todas as AZs disponíveis. Depois de criar um gateway da internet, você o anexa à sua VPC.



Gateway anexado à sua VPC.

Gateway NAT

Um gateway NAT é um serviço de Network Address Translation (NAT que pode ser usado para que as instâncias em uma sub-rede privada possam se conectar a serviços fora de sua VPC, mas os serviços externos não podem iniciar uma conexão com essas instâncias.

Ao criar um gateway NAT, você especifica um dos seguintes tipos de conectividade:]

Pública (padrão)

As instâncias em sub-redes privadas podem se conectar à internet por meio de um gateway NAT público, mas não podem receber conexões de entrada não solicitadas. Você cria um gateway NAT público em uma sub-rede pública e deve associar a ele, na criação, um endereço IP elástico. Você roteia o tráfego do gateway NAT para o gateway da internet para a VPC. Como alternativa, pode ser usado um gateway NAT público para se conectar a outras VPCs ou à sua rede local. Nesse caso, você roteia o tráfego do gateway NAT por meio de um gateway de trânsito ou de um gateway privado virtual.

Privada

As instâncias em sub-redes privadas podem se conectar a outras VPCs ou à sua rede local por meio de um gateway NAT privado. Serve basicamente para isolar, mas dar conectividade a outras redes privadas, como em outras VPCs ou com redes on-premise via VPN ou conectividade direta. Você não pode associar um endereço IP elástico a um gateway NAT privado. É possível anexar um gateway da internet a uma VPC usando um gateway NAT privado. Porém, se rotear o tráfego do gateway NAT para o gateway da internet, este último vai descartar o tráfego.

O gateway NAT substitui o endereço IP de origem das instâncias por endereço IP próprio. Para um gateway NAT público, este é o seu endereço IP elástico. Para um gateway NAT privado, este é o endereço IP privado do gateway NAT. Ao enviar tráfego de resposta para as instâncias, o dispositivo NAT converte os endereços de volta para o endereço IP original.

Os principais casos de uso do gateway NAT são:

Acesso à internet a partir de uma sub-rede privada

Você pode usar um gateway NAT público para permitir que instâncias em uma sub-rede privada enviem tráfego de saída para a internet, evitando sua conexão com as instâncias.

Acesso a uma rede usando endereços IP permitidos

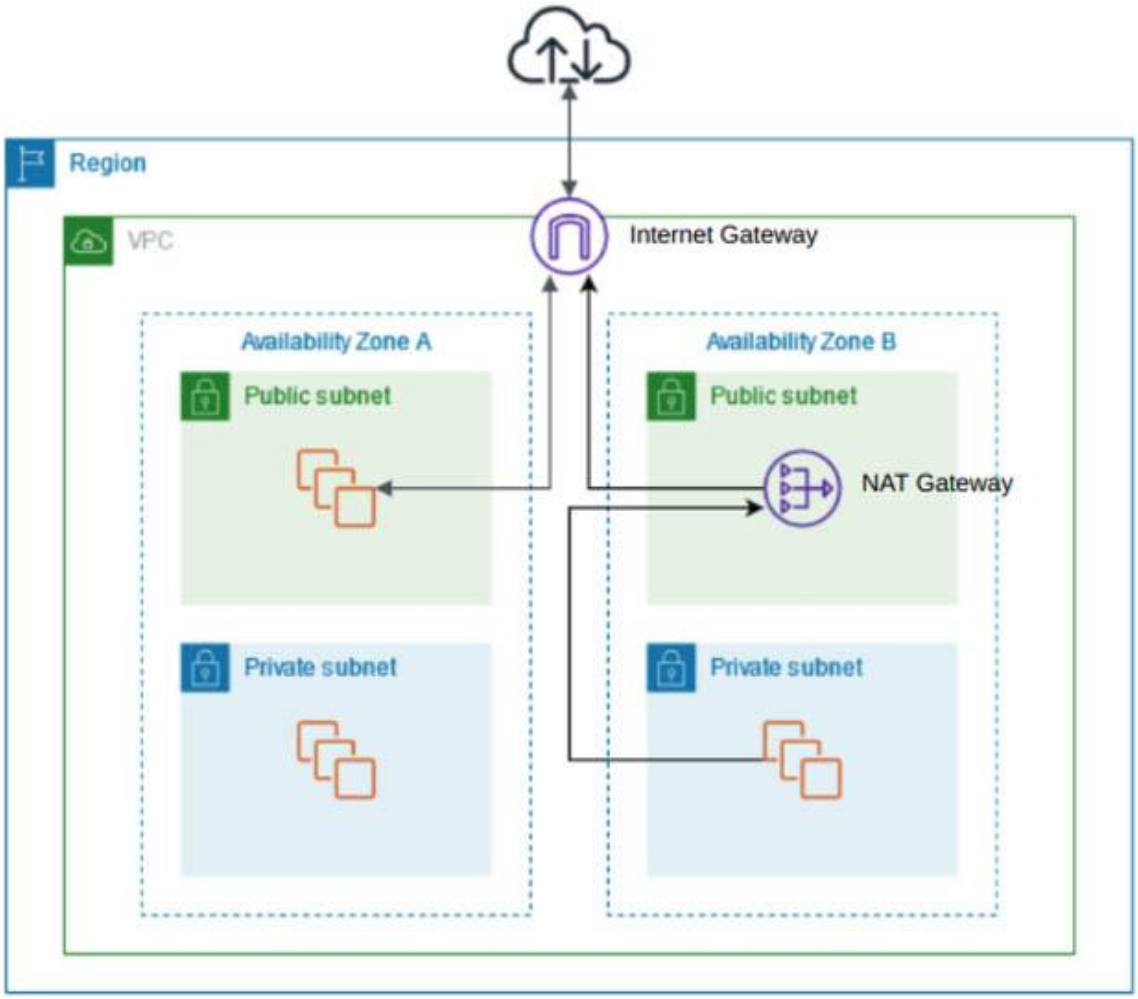
Se você precisa se comunicar com outra rede, pública ou privada, mas precisa restringir o acesso a determinados IPS, pode usar um gateway NAT, público ou privado, para representar um pool de IPs de instâncias, ao invés de criar regras de liberação para cada uma delas.

Ativar a comunicação entre redes sobrepostas

Você pode usar um gateway NAT privado para permitir a comunicação entre as redes, mesmo que tenham intervalos CIDR sobrepostos. Por exemplo,

suponhamos que as instâncias na VPC A precisem acessar os serviços fornecidos pelas instâncias na VPC B. Elas utilizarão o mesmo bloco de IPs 10.0.0.0/24.

A seguir, veremos um diagrama com gateway NAT público em conjunto com o gateway de Internet para garantir conectividade.



Tabelas de rota

Quando você cria uma VPC, a AWS cria tabelas de rota principal. As tabelas de rota contêm um conjunto de regras, chamadas rotas, que são usadas para determinar para onde é direcionado o tráfego de rede. A AWS supõe que, ao criar uma nova VPC com sub-redes, você deseja que o tráfego flua entre elas. Portanto, a configuração padrão da tabela de rota principal tem por objetivo permitir o tráfego entre todas as sub-redes da rede local. A seguir, apresentamos um exemplo de uma tabela de rota principal.

Vamos considerar que o destino e o alvo são duas partes principais dessa tabela de rotas.

- O destino (destination) é um intervalo de endereços IP para o qual você deseja que seu tráfego vá. No exemplo do envio de uma carta, você precisa de um destino para encaminhá-la ao local apropriado. O mesmo ocorre para o tráfego de roteamento. Nesse caso, o destino é o intervalo de IP da rede VPC.
- O alvo (target) é a conexão por meio da qual será enviado o tráfego. Nesse caso, o tráfego é roteado pela rede VPC local.

Destination	Target	Status	Propagated
0.0.0.0/0	igw-63082	Active	No
172.31.0.0/16	local	Active	No

Tabela de rota indicando destino (destination) e alvo (target).

VPC padrão

Neste vídeo, você conhecerá a VPC que a AWS já pré-provisiona na região.

Ao criar uma conta na AWS, você já encontra uma VPC padrão (default) em cada região (a própria AWS já deixou isso pré-provisionado). Uma VPC padrão vem com uma sub-rede pública em cada zona de disponibilidade, um gateway de internet e configurações para habilitar a resolução de DNS, permitindo que você possa iniciar imediatamente as instâncias do Amazon EC2.

Uma VPC padrão é adequada para começar rapidamente e iniciar instâncias públicas, como um blog ou site simples. Você pode modificar os componentes de sua VPC padrão conforme necessário para adequar às suas necessidades.

A VPC padrão possui as seguintes características:

- VPC com um bloco CIDR IPv4 de tamanho /16 (172.31.0.0/16), o que permite fornecer até 65.536 endereços IPv4 privados.
- Sub-redes de tamanho /20 em cada zona de disponibilidade. Isso fornece até 4.096 endereços por sub-rede, lembrando que alguns são de uso reservado.
- Um gateway de internet já conectado à VPC padrão.
- Uma tabela de rotas principal que aponta todo o tráfego (0.0.0.0/0) para o gateway da internet.
- Grupo de segurança padrão.
- Opções de DHCP padrão definidas para sua conta da AWS com sua VPC padrão.