WHAT IS AMAZON VPC? AMAZON VIRTUAL PRIVATE CLOUD EXPLAINED

Amazon's Virtual Private Cloud (VPC) is a foundational AWS service in both the Compute and Network AWS categories. Being foundational means that other AWS services, such as Elastic Compute Cloud (EC2), cannot be accessed without an underlying VPC network.

Creating a VPC is critical to running in the AWS cloud. Let's take a look at:

- How VPCs work
- Where they live
- VPC management
- Elements of a VPC
- Shared responsibility

(This tutorial is part of our AWS Guide. Use the right-hand menu to navigate.)

How VPCs work: virtual networking environments

Each VPC creates an isolated virtual network environment in the AWS cloud, dedicated to your AWS account. Other AWS resources and services operate inside of VPC networks to provide cloud services.

AWS VPC will look familiar to anyone used to running a physical Data Center (DC). A VPC behaves like a traditional TCP/IP network that can be expanded and scaled as needed. However, the DC components you are used to dealing with—such as routers, switches, VLANS, etc.—do not explicitly exist in a VPC. They have been abstracted and re-engineered into cloud software.

Using VPC, you can quickly spin up a virtual network infrastructure that AWS instances can be launched into. Each VPC defines what your AWS resources need, including:

- IP addresses
- Subnets
- Routing
- Security
- Networking functionality

Where VPCs live

All VPCs are created and exist in one—and only one—AWS region. AWS regions are geographic locations around the world where Amazon clusters its cloud data centers.

The advantage of regionalization is that a regional VPC provides network services originating from that geographical area. If you need to provide closer access for customers in another region, you can set up another VPC in that region.

This aligns nicely with the theory of AWS cloud computing where IT applications and resources are delivered through the internet on-demand and with pay-as-you-go pricing. Limiting VPC configurations to specific regions allows you to selectively provide network services where they are needed, as they are needed.

Each Amazon account can host multiple VPCs. Because VPCs are isolated from each other, you can duplicate private subnets among VPCs the same way you could use the same subnet in two different physical data centers. You can also add public IP addresses that can be used to reach VPC-launched instances from the internet.

Amazon creates one default VPC for each account, complete with:

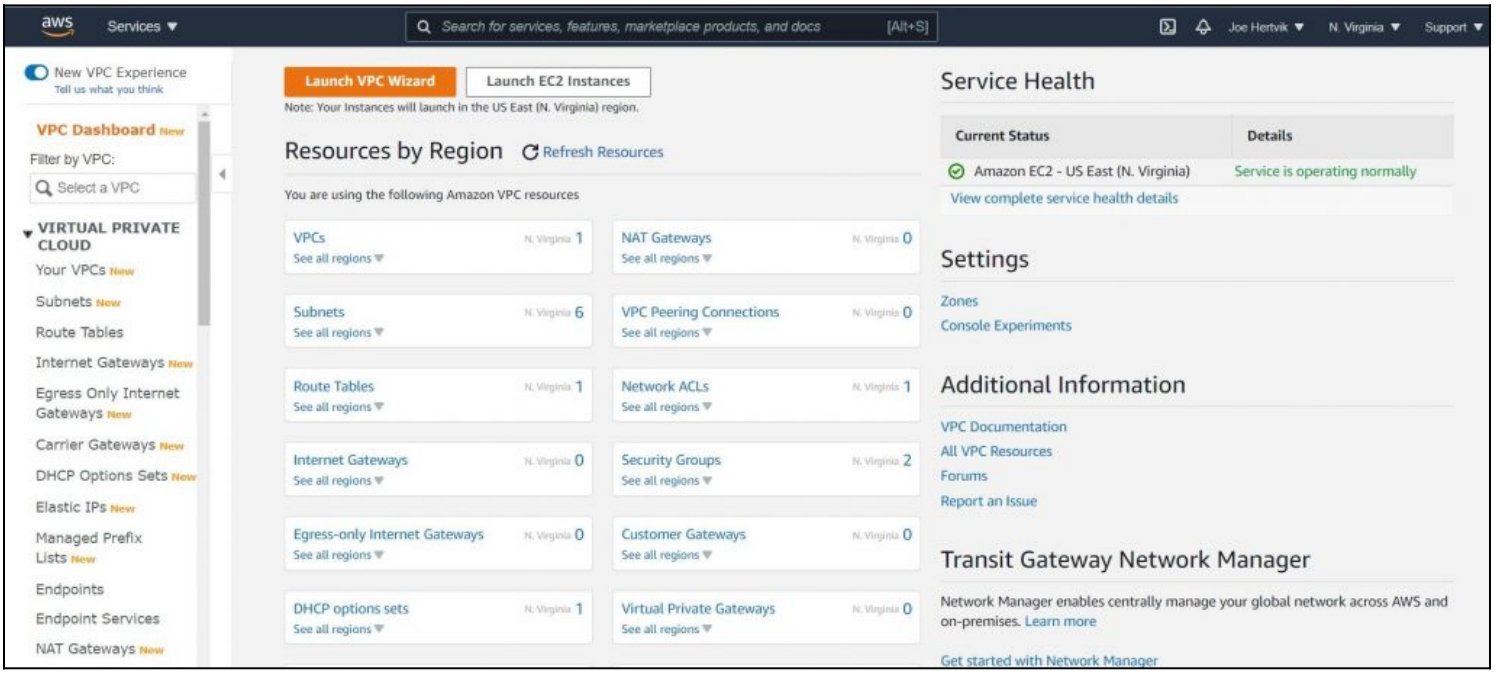
- Default subnets
- Routing tables
- Security groups
- Network access control list

You can modify or use that VPC for your cloud configurations or you can build a new VPC and supporting services from scratch.

Managing your VPCs

VPC administration is handled through these AWS management interfaces:

- **AWS Management Console** is the web interface for managing all AWS functions (image below).
- **AWS Command Line Interface (CLI)** provides Windows, Linux, and Mac commands for many AWS services. AWS frequently provides configuration instructions as CLI commands.
- **AWS Software Development Kit (SDK)** provides language-specific APIs for AWS services, including VPCs.
- **Query APIs.** Low-level API actions can be submitted through HTTP or HTTPS requests. Check AWS's EC2 API Reference for more information.



The AWS Management Console manages your VPCs and other AWS services

(Learn about more AWS management tools.)

Elements of a VPC

The web-based AWS management console, show above, shows most of the VPC resources you can create and manage. VPC network services include:

- IPv4 and IPv6 address blocks
- Subnet creation
- Route tables
- Internet connectivity
- Elastic IP addresses (EIPs)
- Network/subnet security
- Additional networking services

Let's look briefly at each.

IPv4 and IPv6 address blocks

VPC IP address ranges are defined using Classless interdomain routing (CIDR) IPv4 and IPv6 blocks. You can add primary and secondary CIDR blocks to your VPC, if the secondary CIDR block comes from the same address range as the primary block.

AWS recommends that you specify CIDR blocks from the private address ranges specified in RFC 1918, shown in the table below. See the AWS VPCs and Subnets page for restrictions on which CIDR blocks can be used.

RFC 1918 Private IP Range
10.0.0.0 – 10.255.255.255
172.16.0.0 – 172.31.255.255
192.168.0.0 – 192.168.255.255

Recommended private IPv4 address ranges for AWS VPC CIDR blocks

Subnet creation

Launched EC2 instances run inside a designated VPC subnet (sometimes referred to as launching an instance into a subnet).

For IP addressing, each subnet's CIDR contains a subset of the VPC CIDR block. Each subnet isolates its individual traffic from all other VPC subnet traffic. A subnet can only contain one CIDR block. You can designate different subnets to handle different types of traffic.

For example, file server instances can be launched into one subnet, web and mobile applications can be launched into a different subnet, printing services into another, and so on.

Route tables

Route tables contains the rules (routes) that determine how network traffic is directed inside your VPC and subnets. VPC creates a default route table called the *main route table*. The main route table is automatically associated with all VPC subnets. Here, you have two options:

- Update and use the main route table to direct network traffic.
- Create your own route table to be used for individual subnet traffic.

Internet connectivity

For Internet access, each VPC configuration can host one Internet Gateway and provide network address translation (NAT) services using the Internet Gateway, NAT instances, or a NAT gateway.

Elastic IP addresses (EIPs)

EIPs are static public IPv4 addresses that are permanently allocated to your AWS account (EIP is not offered for IPv6). EIPs are used for public Internet access to:

- An instance
- An AWS elastic network interface (ENI)
- Other services needing a public IP address

You allocate EIPs for long-term permanent network usage.

Network/subnet security

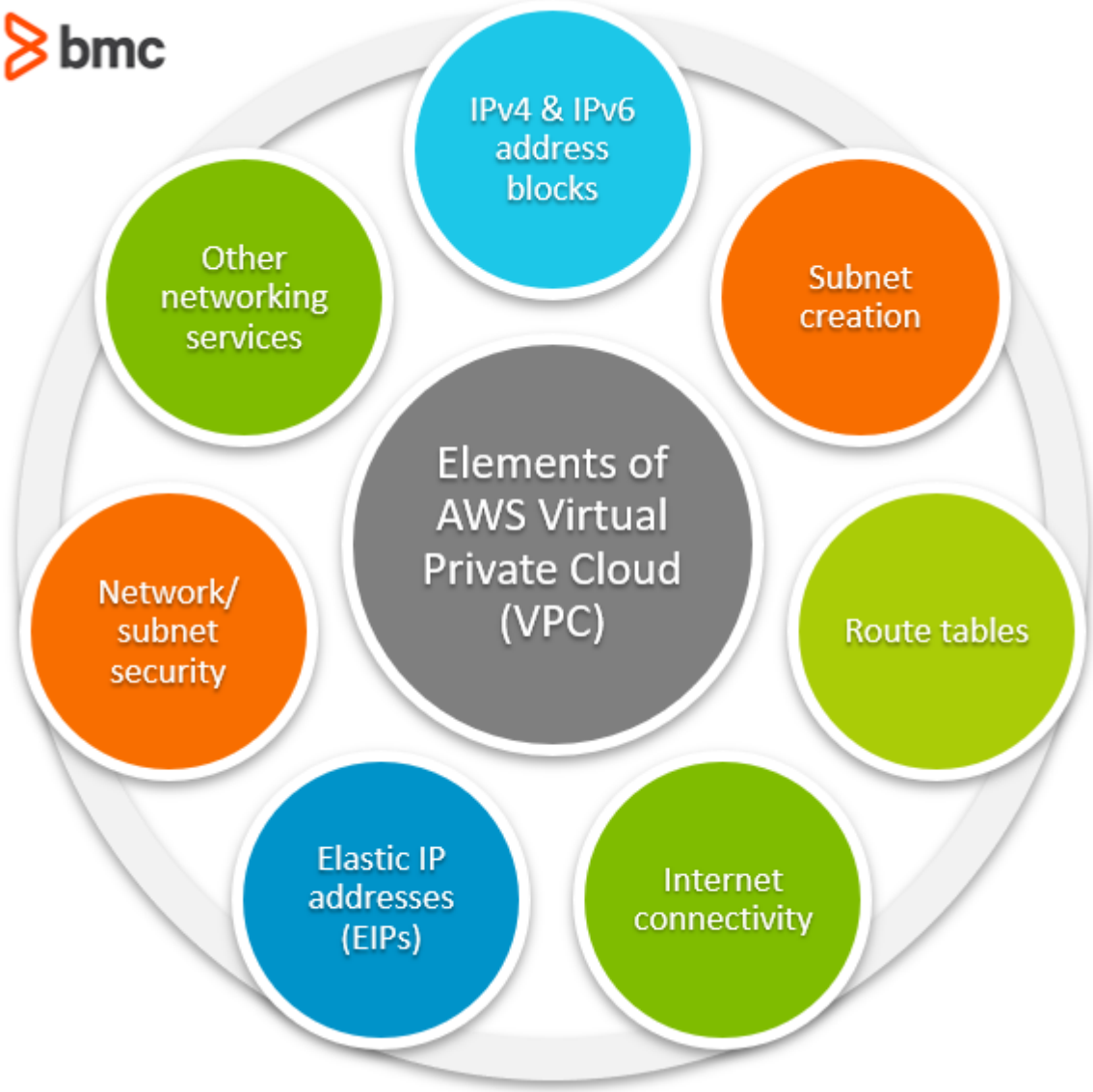
VPCs use security groups to provide stateful protection (the state of the connection session is maintained) for instances. AWS describes security groups as virtual firewalls.

VPCs also provide network access control lists (NACLs) to stateless VPC subnets—that is, the state of the connection is not maintained.

Additional networking services

Of course, these are not the only AWS services a VPC provides. You can use VPC to configure other common networking services such as:

- Virtual Private Networks (VPNs)
- Direct connectivity between VPCs (VPC peering)
- Gateways
- Mirror sessions



VPCs & shared responsibility

Before you start configuring VPCs, check out Amazon's Shared Responsibility model. Per Amazon, security and compliance is a shared responsibility between AWS and its customers.

For your AWS account and configurations, AWS is responsible for the "Security of the Cloud" while customers are responsible for "Security in the Cloud." Generally:

- AWS is responsible for the AWS cloud infrastructure (hardware, cloud software, networking, facilities) that run AWS services.
- Customers are responsible for what they run in the cloud, such as servers, data, encryption, applications, security, access, operating systems, etc.

The shared responsibility model lays out who is responsible for specific issues when you experience AWS downtime, security breaches, or loss of business. It is important to understand these limits as you set up your VPC configuration. Consult the shared responsibility model for more information.

Related reading

- BMC Multi-Cloud Blog
- The AWS Well-Architected Framework: 5 Pillars & Best Practices
- Public vs Private vs Hybrid: Cloud Differences Explained
- Rise of Data Centers & Private Clouds in Response to Amazon's Hegemony
- Cloud Growth, Trends & Outlook