AWS Networking and Content Delivery

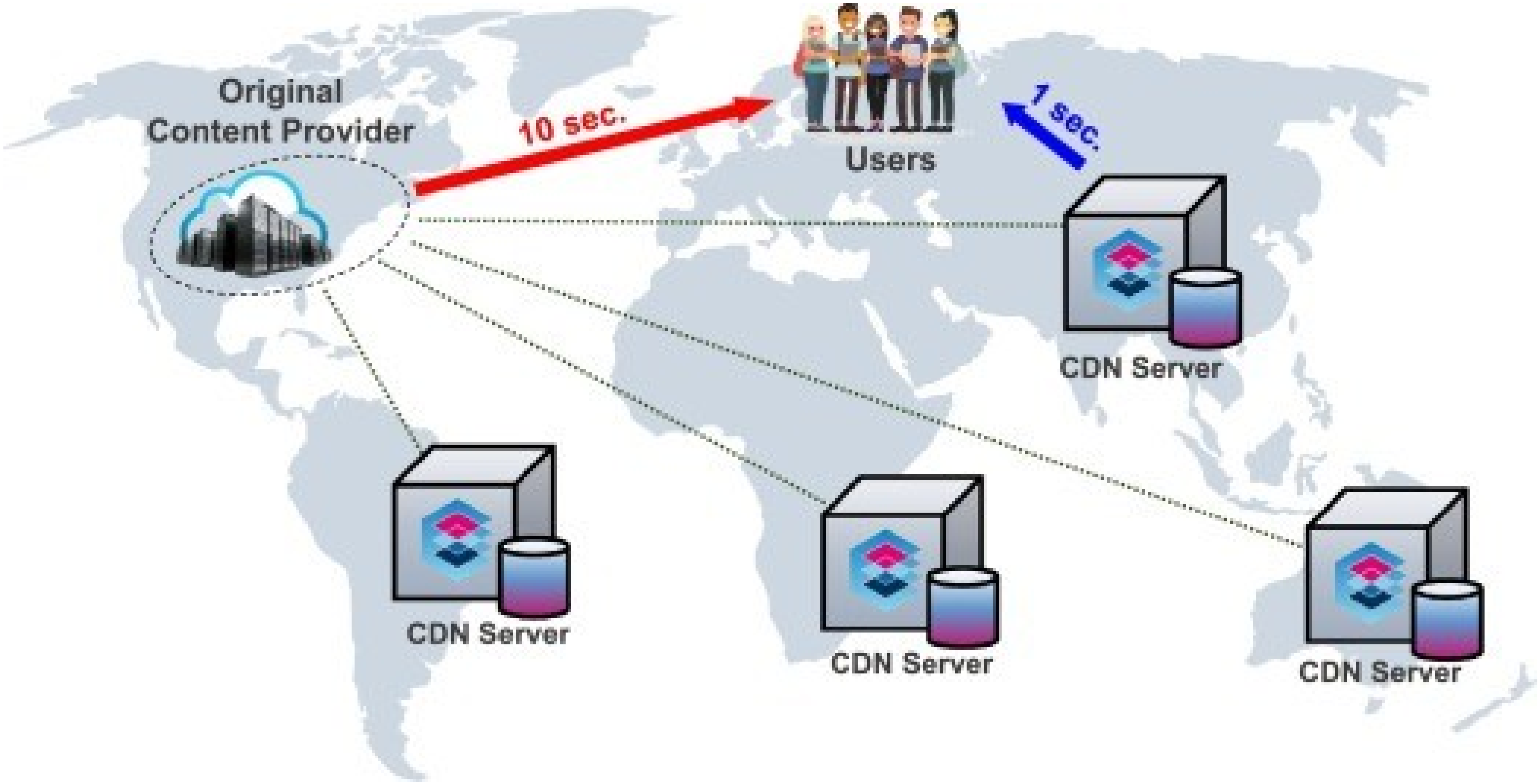
Module 3

What is a CDN?

A content delivery network (CDN) is a network of interconnected servers that speeds up webpage loading for data-heavy applications. CDN can stand for content delivery network or content distribution network.

When a user visits a website, data from that website's server has to travel across the internet to reach the user's computer.

If the user is located far from that server, it will take a long time to load a large file, such as a video or website image. Instead, the website content is stored on CDN servers geographically closer to the users and reaches their computers much faster.


Original
Content Provider
10 sec.
Users
1 sec.
CDN Server
CDN Server
CDN Server
CDN Server

Why is a CDN important?

The primary purpose of a content delivery network (CDN) is to reduce latency, or reduce the delay in communication created by a network's design.

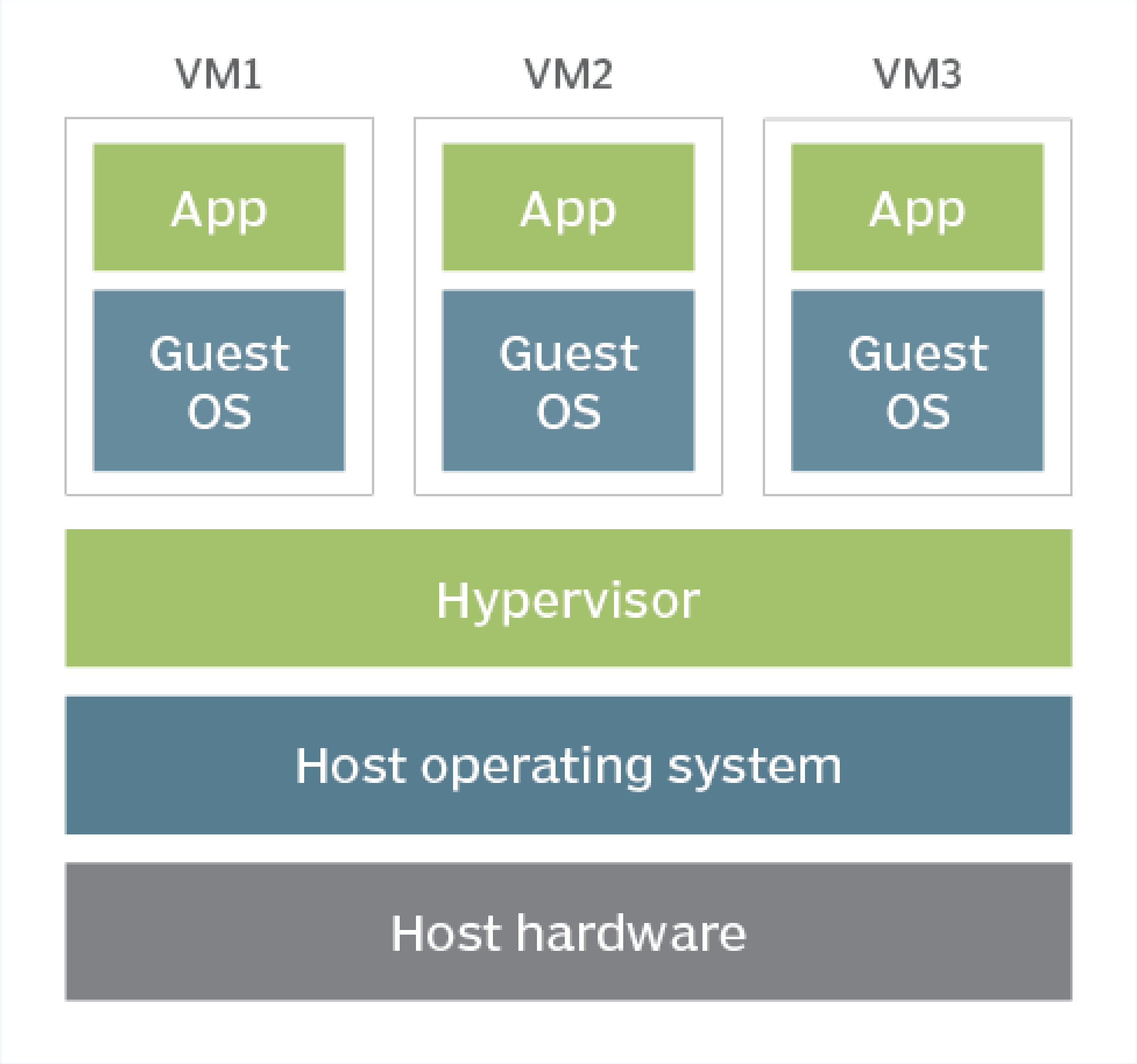
Because of the global and complex nature of the internet, communication traffic between websites (servers) and their users (clients) has to move over large physical distances.

A CDN improves efficiency by introducing intermediary servers between the client and the website server. These CDN servers manage some of the client-server communications. They decrease web traffic to the web server, reduce bandwidth consumption, and improve the user experience of your applications.

A **virtual network** connects virtual machines and devices, no matter their location, using software. In a physical network, layer 2 and 3 functions of the OSI model happen within physical switches and routers.

In its simplest form, a **virtual machine**, or VM, is a digitized version of a physical computer. Virtual machines can run programs and operating systems, store data, connect to networks, and do other computing functions. However, a VM uses entirely virtual resources instead of physical components.

Virtual machines



Amazon VPC (Virtual Private Cloud)

Amazon Virtual Private Cloud (Amazon VPC) lets you provision a logically isolated section of the AWS Cloud where you can launch AWS resources in a virtual network that you define.

Amazon VPC or Amazon Virtual Private Cloud is a service that allows its users to launch their virtual machines in a protected as well as isolated virtual environment defined by them.

Amazon VPC can be referred to as the private cloud inside the cloud. It is a logical grouping of servers in a specified network. The servers that you are going to deploy in the Virtual Private Cloud(VPC) will be completely isolated from the other servers that are deployed in the Amazon Web Services.

You have complete control over your virtual networking environment.

Amazon VPC (Virtual Private Cloud)

You can easily customize the network configuration for your VPC. For example, you can create a public-facing subnet for your web servers that has access to the Internet, and place your backend systems, such as databases or application servers, in a private-facing subnet with no Internet access.

You can leverage multiple layers of security (including security groups and network access control lists) to help control access to EC2 instances in each subnet.

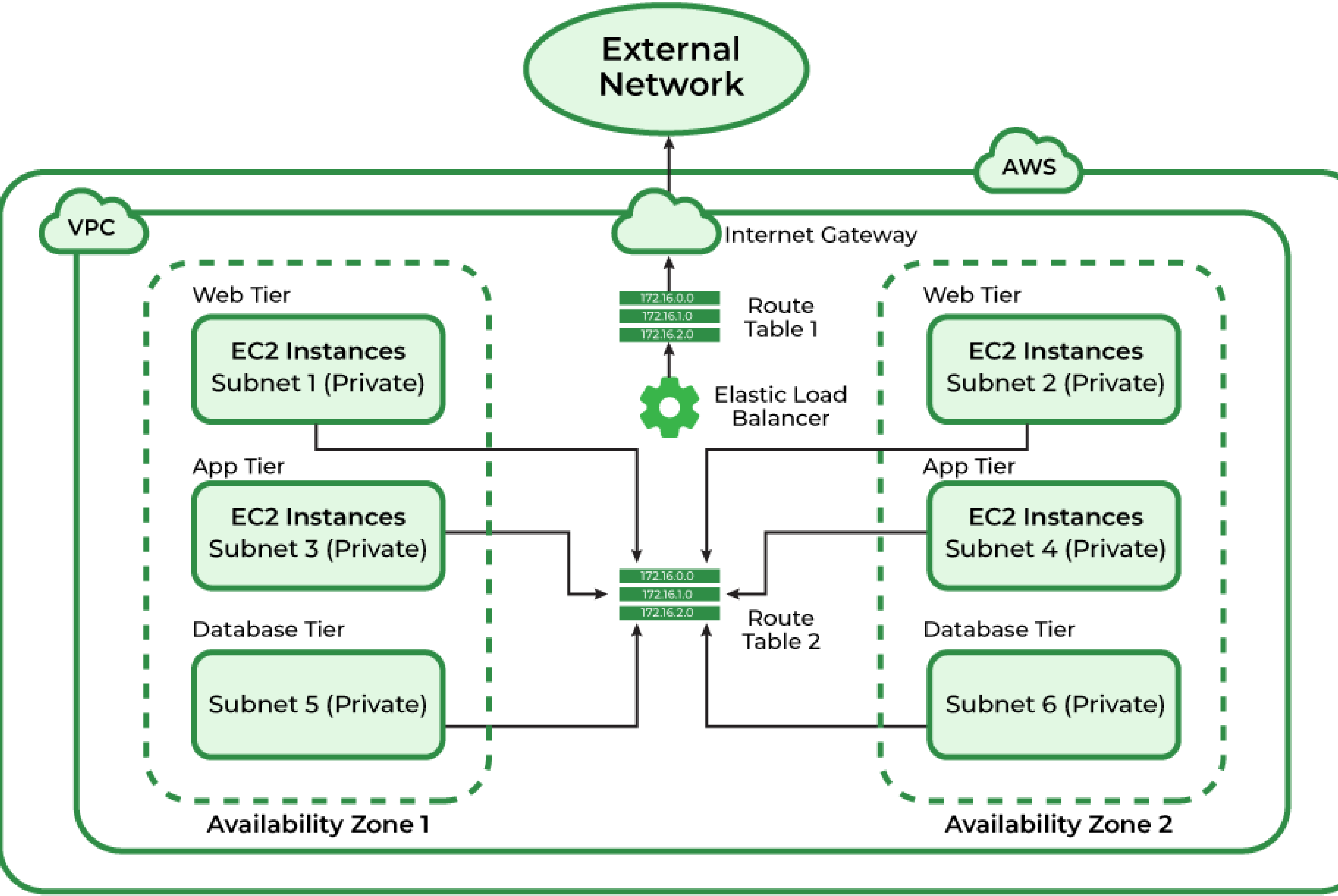
Additionally, you can create a hardware virtual private network (VPN) connection between your corporate data center and your VPC and leverage the AWS Cloud as an extension of your corporate data center.

Amazon VPC (Virtual Private Cloud) Architecture

The basic architecture of a properly functioning VPC consists of many distinct services such as Gateway, Load Balancer, Subnets, etc.

Altogether, these resources are clubbed under a VPC to create an isolated virtual environment. Along with these services, there are also security checks on multiple levels.

It is initially divided into subnets, connected with each other via route tables along with a load balancer.


External Network
AWS
VPC
Internet Gateway
172.16.0.0
172.16.1.0
172.16.2.0
Route Table 1
Elastic Load Balancer
Web Tier
EC2 Instances Subnet 1 (Private)
App Tier
EC2 Instances Subnet 3 (Private)
Database Tier
Subnet 5 (Private)
Web Tier
EC2 Instances Subnet 2 (Private)
App Tier
EC2 Instances Subnet 4 (Private)
Database Tier
Subnet 6 (Private)
172.16.0.0
172.16.1.0
172.16.2.0
Route Table 2
Availability Zone 1
Availability Zone 2

Amazon VPC Components

Subnets

To reduce traffic, the subnet will divide the big network into smaller, connected networks. Up to /16, 200 user-defined subnets.

Route Tables

Route Tables are mainly used to Define the protocol for traffic routing between the subnets.

Network Access Control Lists

Network Access Control Lists (NACL) for VPC serve as a firewall by managing both inbound and outbound rules. There will be a default NACL for each VPC that cannot be deleted.

Amazon VPC Components

Internet Gateway(IGW)

The Internet Gateway (IGW) will make it possible to link the resources in the VPC to the Internet.

Network Address Translation (NAT)

Network Address Translation (NAT) will enable the connection between the private subnet and the internet.

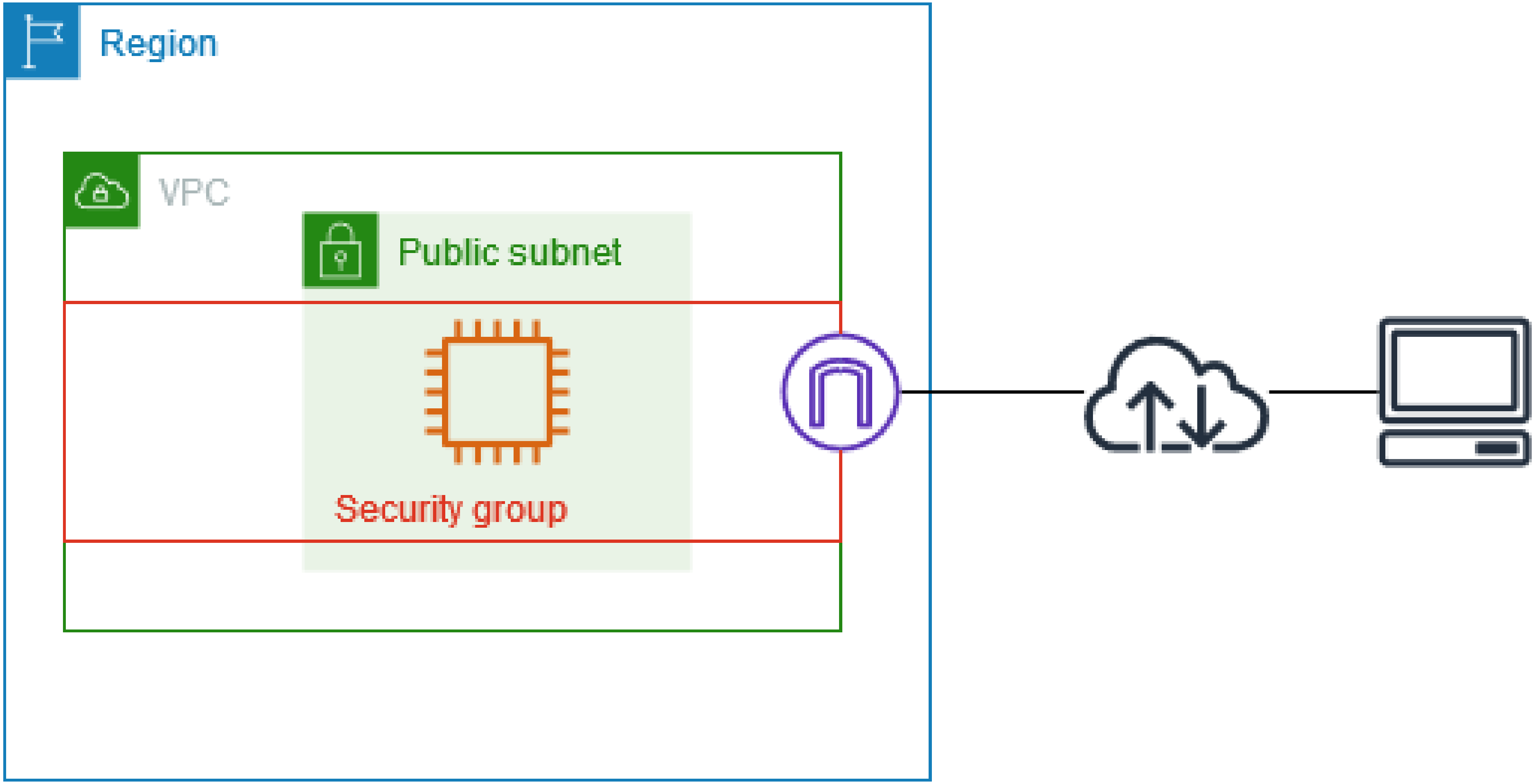
Security Groups

A *security group* controls the traffic that is allowed to reach and leave the resources that it is associated with.

For example, after you associate a security group with an EC2 instance, it controls the inbound and outbound traffic for the instance.

When you create a VPC, it comes with a default security group. You can create additional security groups for a VPC, each with their own inbound and outbound rules.

You can specify the source, port range, and protocol for each inbound rule. You can specify the destination, port range, and protocol for each outbound rule.


Region
VPC
Public subnet
Security group

Security Groups

The diagram shows a VPC with a subnet, an internet gateway, and a security group. The subnet contains an EC2 instance. The security group is assigned to the instance. The security group acts as a virtual firewall. The only traffic that reaches the instance is the traffic allowed by the security group rules.

For example, if the security group contains a rule that allows ICMP traffic to the instance from your network, then you could ping the instance from your computer.

If the security group does not contain a rule that allows SSH traffic, then you could not connect to your instance using SSH.

Security group basics

- You can assign a security group only to resources created in the same VPC as the security group. You can assign multiple security groups to a resource.
- When you create a security group, you must provide it with a name and a description. The following rules apply:
    - A security group name must be unique within the VPC.
    - Names and descriptions can be up to 255 characters in length.
    - Names and descriptions are limited to the following characters: a-z, A-Z, 0-9, spaces, and ._-:/()#,@[]+=&;{}!$*.
    - When the name contains trailing spaces, we trim the space at the end of the name. For example, if you enter "Test Security Group " for the name, we store it as "Test Security Group".
    - A security group name cannot start with sg-.

Security group basics

Security groups are stateful. For example, if you send a request from an instance, the response traffic for that request is allowed to reach the instance regardless of the inbound security group rules. Responses to allowed inbound traffic are allowed to leave the instance, regardless of the outbound rules.

Security groups do not filter traffic destined to and from the following:

Amazon Domain Name Services (DNS)

Amazon Dynamic Host Configuration Protocol (DHCP)

Amazon EC2 instance metadata

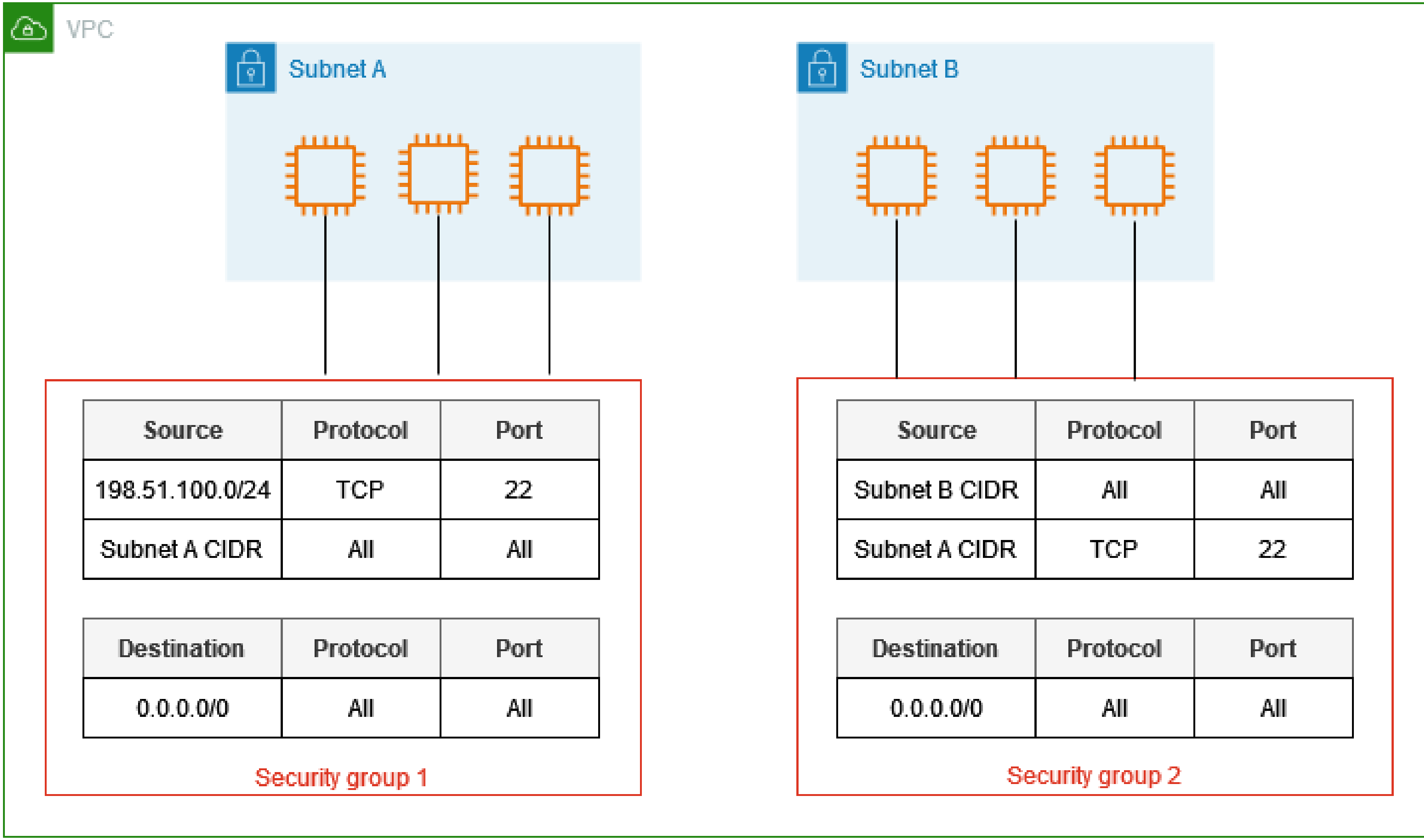
Amazon ECS task metadata endpoints

License activation for Windows instances

Amazon Time Sync Service

Reserved IP addresses used by the default VPC router

Analyze the Security Group process in the diagram



Source	Protocol	Port
198.51.100.0/24	TCP	22
Subnet A CIDR	All	All

Destination	Protocol	Port
0.0.0.0/0	All	All

Source	Protocol	Port
Subnet B CIDR	All	All
Subnet A CIDR	TCP	22

Destination	Protocol	Port
0.0.0.0/0	All	All

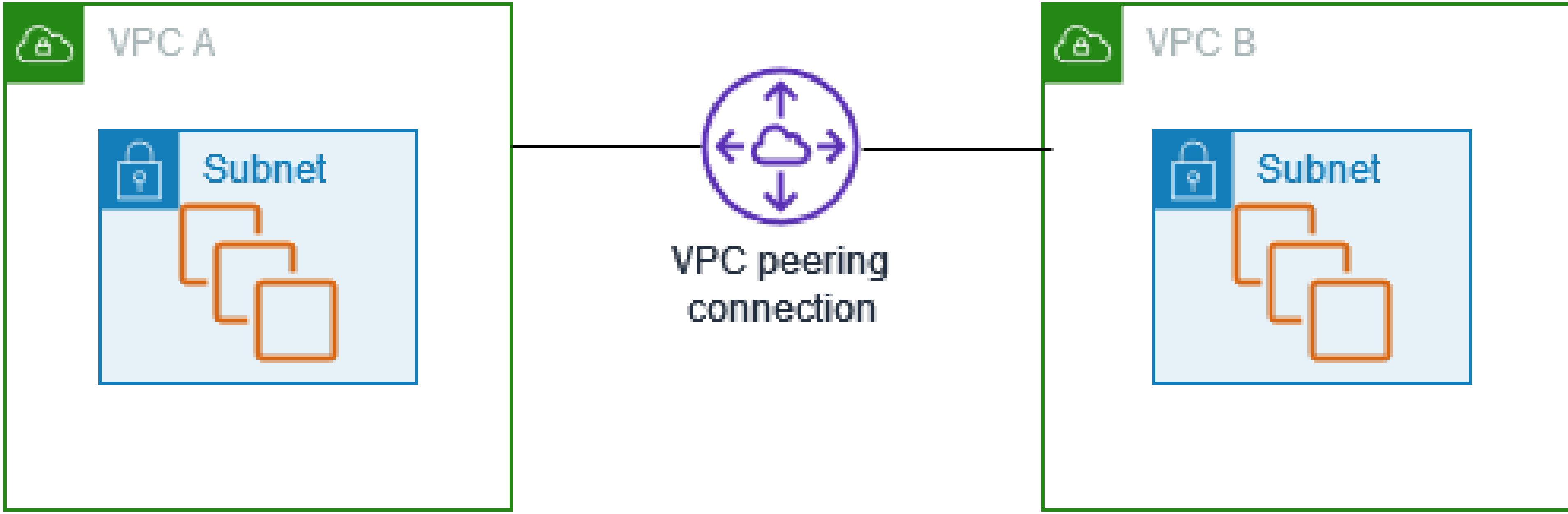
VPC Peering

A VPC peering connection is a networking connection between two VPCs that enables you to route traffic between them using private IPv4 addresses or IPv6 addresses.

Instances in either VPC can communicate with each other as if they are within the same network.

You can create a VPC peering connection between your own VPCs, or with a VPC in another AWS account.

The VPCs can be in different Regions (***also known as an inter-Region VPC peering connection***).


VPC A
Subnet
VPC peering connection
VPC B
Subnet

VPC Peering

AWS uses the existing infrastructure of a VPC to create a VPC peering connection; it is neither a gateway nor a VPN connection, and does not rely on a separate piece of physical hardware.

There is no single point of failure for communication or a bandwidth bottleneck.

A VPC peering connection helps you to facilitate the transfer of data.

For example, if you have more than one AWS account, you can peer the VPCs across those accounts to create a file sharing network.

You can also use a VPC peering connection to allow other VPCs to access resources you have in one of your VPCs.

VPC Peering

When you establish peering relationships between VPCs across different AWS Regions, resources in the VPCs (for example, EC2 instances and Lambda functions) in different AWS Regions can communicate with each other using private IP addresses, without using a gateway, VPN connection, or network appliance.

The traffic remains in the private IP address space. All inter-Region traffic is encrypted with no single point of failure, or bandwidth bottleneck.

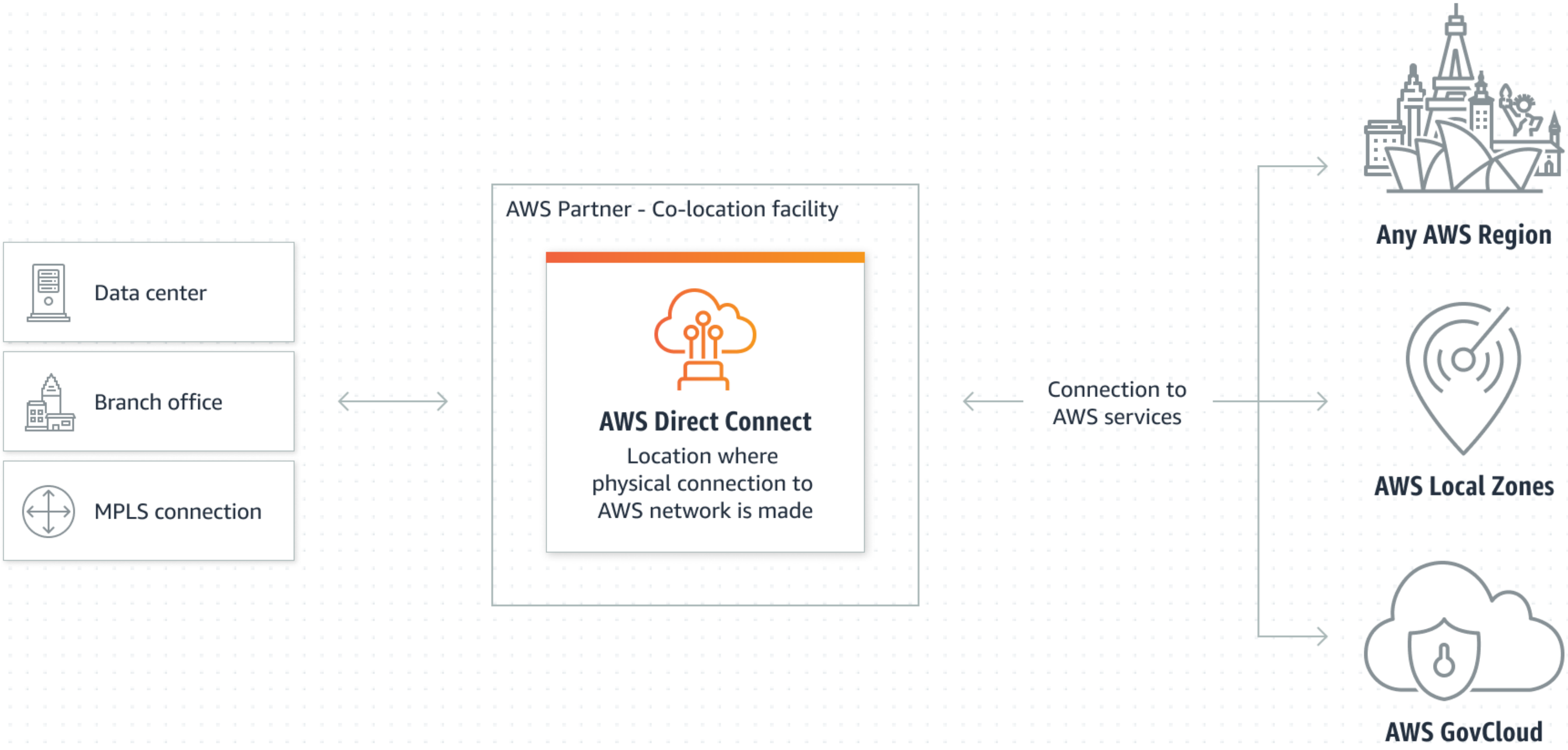
Traffic always stays on the global AWS backbone, and never traverses the public internet, which reduces threats, such as common exploits, and DDoS attacks. Inter-Region VPC peering provides a simple and cost-effective way to share resources between Regions or replicate data for geographic redundancy.

AWS Direct Connect

AWS Direct Connect makes it easy to establish a dedicated connection from an on-premises network to one or more VPCs.

AWS Direct Connect can reduce network costs, increase bandwidth throughput, and provide a more consistent network experience than internet-based connections.

It uses industry-standard 802.1Q VLANs to connect to Amazon VPC using private IP addresses.


Data center
Branch office
MPLS connection
AWS Partner - Co-location facility
AWS Direct Connect
Location where physical connection to AWS network is made
Connection to AWS services
Any AWS Region
AWS Local Zones
AWS GovCloud

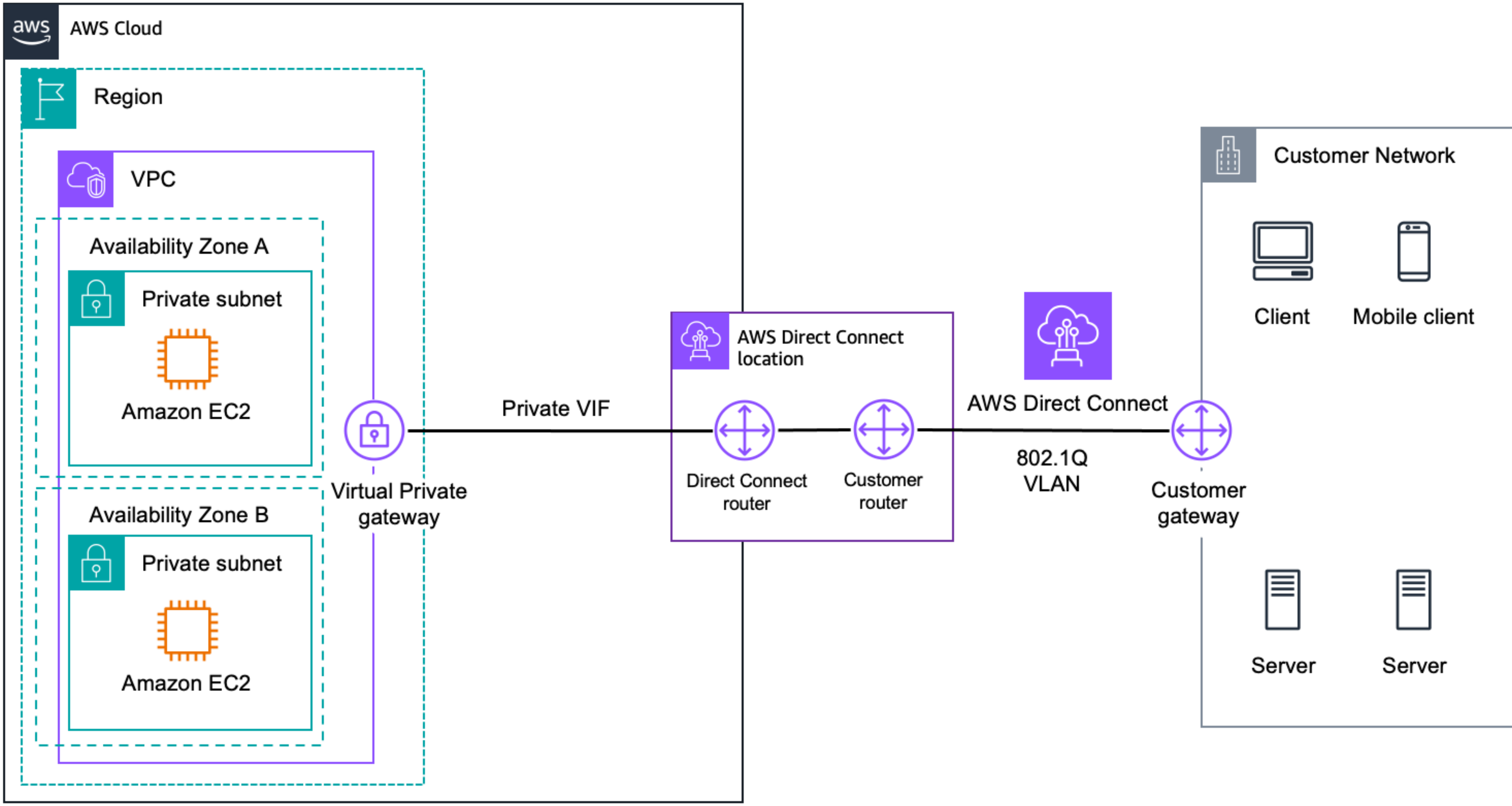
AWS Direct Connect

The VLANs are configured using virtual interfaces (VIFs), and you can configure three different types of VIFs:

Public virtual interface - Establish connectivity between AWS public endpoints and your data center, office, or colocation environment.

Transit virtual interface - Establish private connectivity between AWS Transit Gateway and your data center, office, or colocation environment. This connectivity option is covered in the section AWS Direct Connect + AWS Transit Gateway.

Private virtual interface - Establish private connectivity between Amazon VPC resources and your data center, office, or colocation environment.


AWS Cloud
Region
VPC
Availability Zone A
Private subnet
Amazon EC2
Availability Zone B
Private subnet
Amazon EC2
Virtual Private gateway
Private VIF
AWS Direct Connect location
Direct Connect router
Customer router
AWS Direct Connect
802.1Q VLAN
Customer gateway
Customer Network
Client
Mobile client
Server
Server

AWS Direct Connect

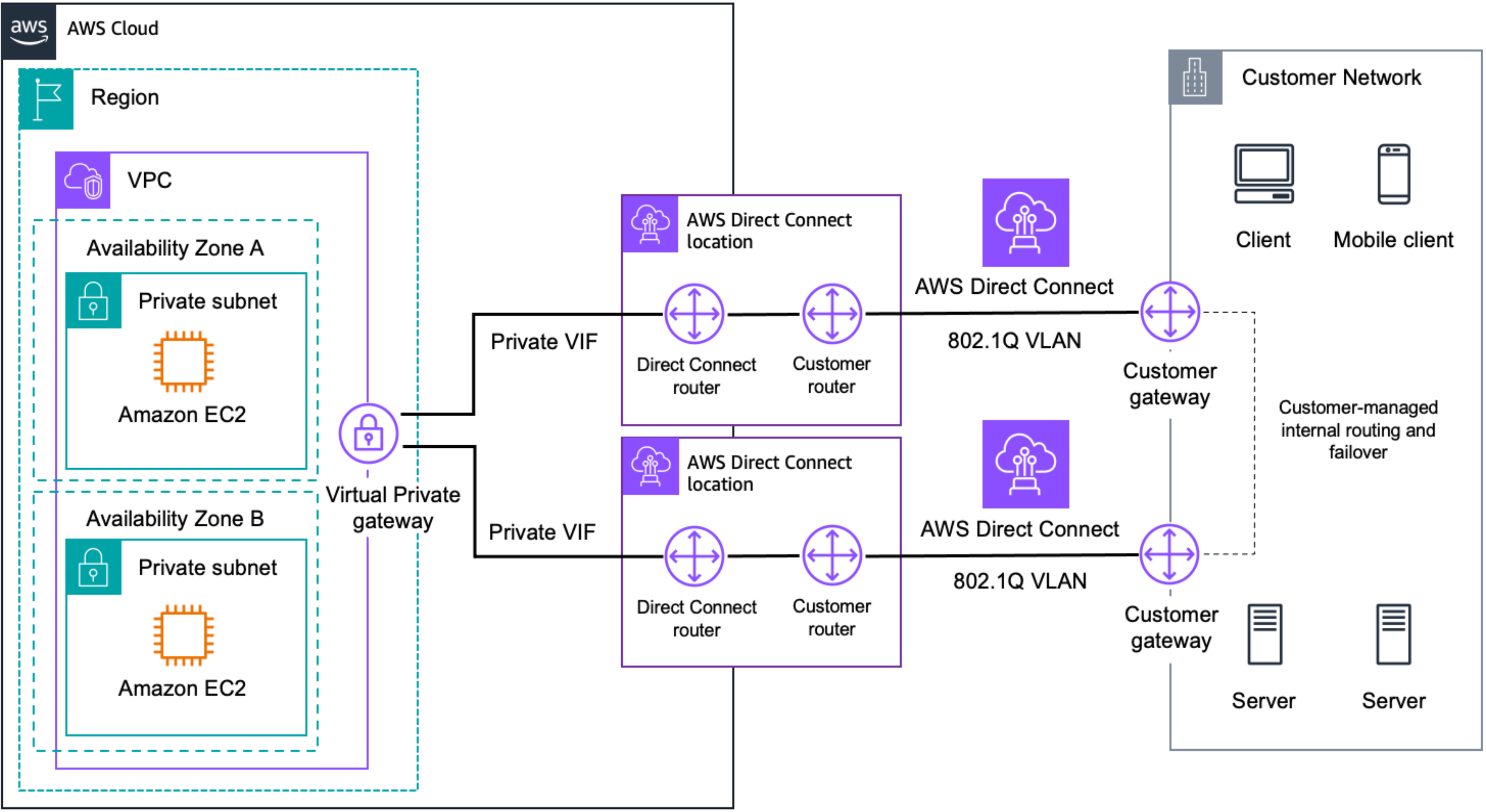
You can establish connectivity to the AWS backbone using AWS Direct Connect by establishing a cross-connect to AWS devices in a Direct Connect location.

With AWS Direct Connect, you have two types of connection:

Dedicated connections, where a physical ethernet connection is associated with a single customer. You can order port speeds of 1, 10, or 100 Gbps.

Hosted connections, where a physical ethernet connection is provisioned by an AWS Direct Connect Partner and shared with you. You can order port speeds between 50 Mbps and 10 Gbps

Analyze the highly resilient network connections based on the diagram


AWS Cloud
Region
VPC
Availability Zone A
Private subnet
Amazon EC2
Availability Zone B
Private subnet
Amazon EC2
Virtual Private gateway
Private VIF
Private VIF
AWS Direct Connect location
Direct Connect router
Customer router
AWS Direct Connect location
Direct Connect router
Customer router
AWS Direct Connect
802.1Q VLAN
AWS Direct Connect
802.1Q VLAN
Customer Network
Client
Mobile client
Customer gateway
Customer-managed internal routing and failover
Customer gateway
Server
Server

Amazon CloudFront

There was a time when having an own Content Delivery Network was extremely rare for companies due to its high costs and complicated IT infrastructure.

But now, AWS CloudFront has helped users to request data resulting in low latency, low network traffic, and quick data access with minimal cost. Thus, making it a very popular network.

AWS CloudFront is a globally-distributed network offered by Amazon Web Services, which securely transfers content such as software, SDKs, videos, etc., to the clients, with high transfer speed.

Benefits of AWS CloudFront

- It will cache your content in edge locations and decrease the workload, thus resulting in high availability of applications.

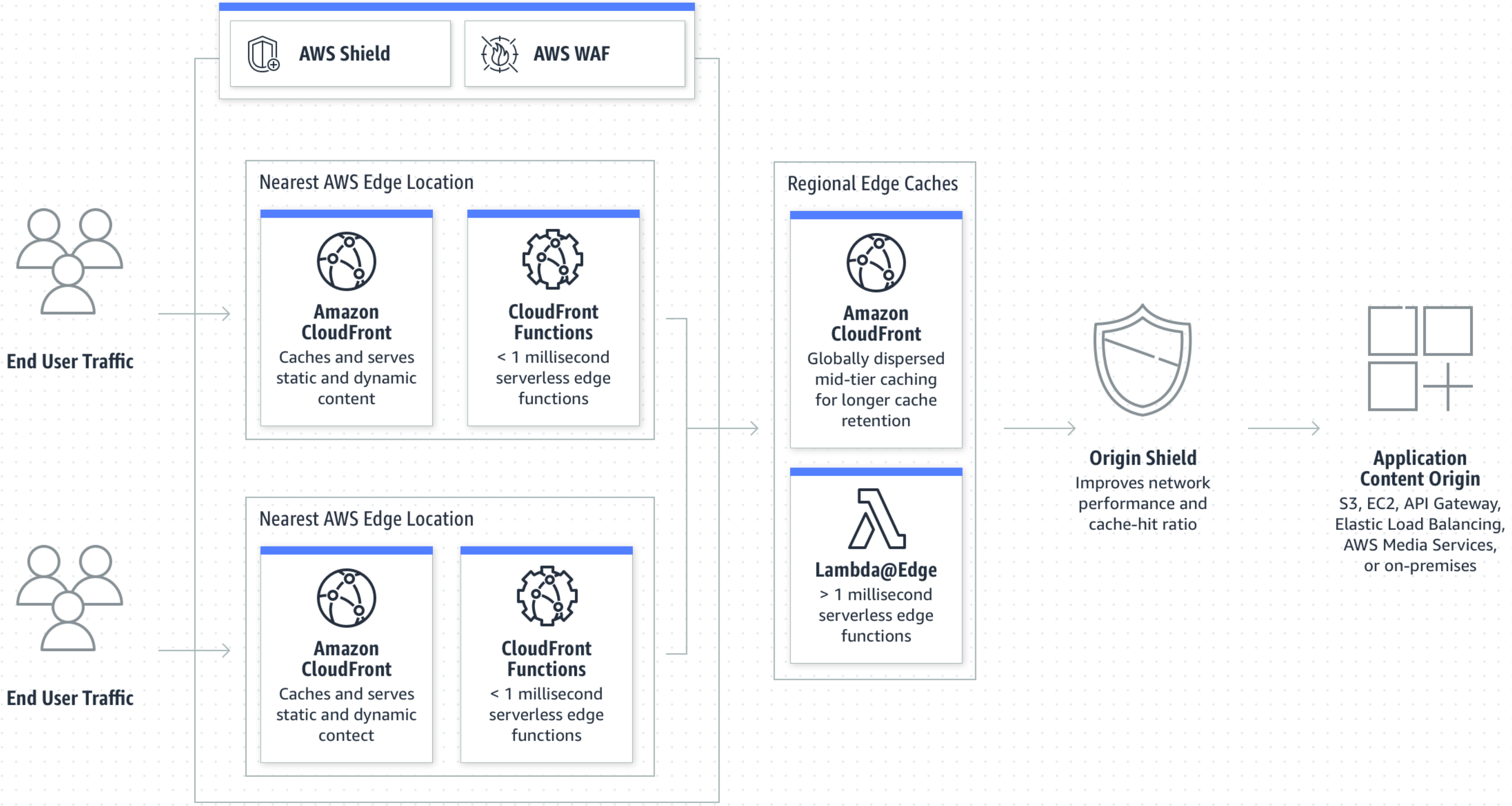
- It is simple to use and ensures productivity enhancement.

- It provides high security with the ‘Content Privacy’ feature.

- It facilitates GEO targeting service for content delivery to specific end-users.

- It uses HTTP or HTTPS protocols for quick delivery of content.

- It is less expensive, as it only charges for the data transfer.



AWS Shield

AWS WAF

Nearest AWS Edge Location

Amazon CloudFront
Caches and serves static and dynamic content

CloudFront Functions
< 1 millisecond serverless edge functions

Nearest AWS Edge Location

Amazon CloudFront
Caches and serves static and dynamic contect

CloudFront Functions
< 1 millisecond serverless edge functions

End User Traffic

End User Traffic

Regional Edge Caches

Amazon CloudFront
Globally dispersed mid-tier caching for longer cache retention

Lambda@Edge
> 1 millisecond serverless edge functions

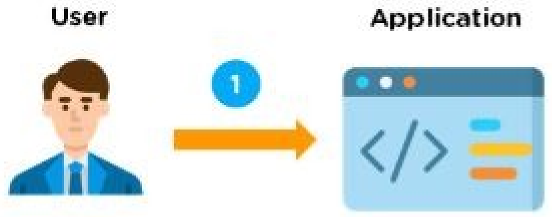
Origin Shield
Improves network performance and cache-hit ratio

Application Content Origin
S3, EC2, API Gateway, Elastic Load Balancing, AWS Media Services, or on-premises

How Does AWS CloudFront Work?

Step 1

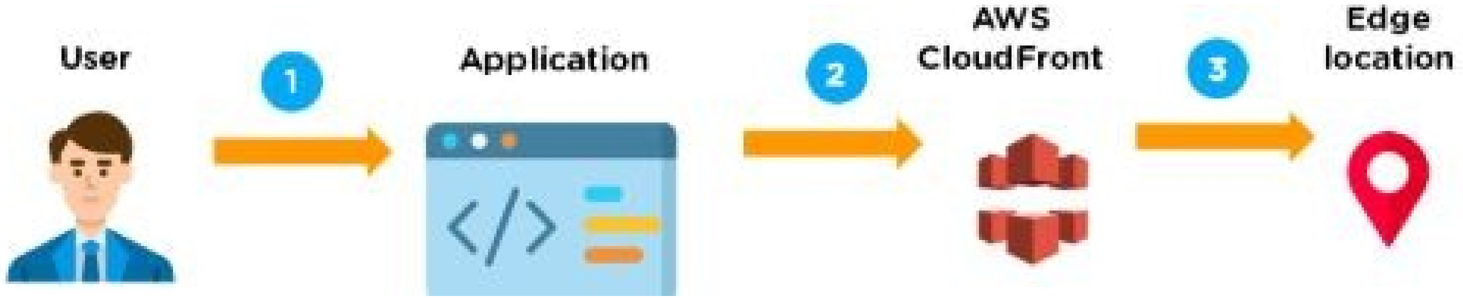
The client accesses a website and requests to download a file (like image file).



Client access

Step 2

Now, the DNS routes the client request to the nearest edge location through CloudFront to serve the user request.

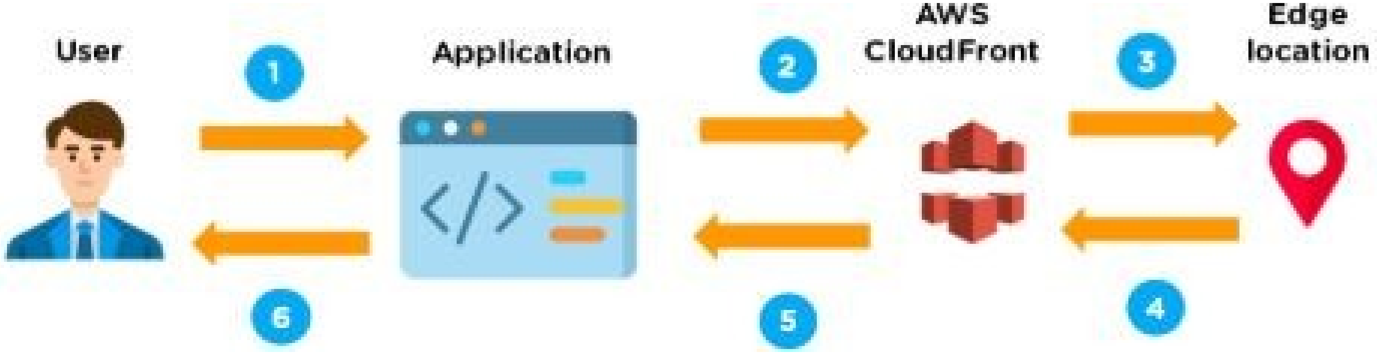


Serve user request

How Does AWS CloudFront Work?

Step 3

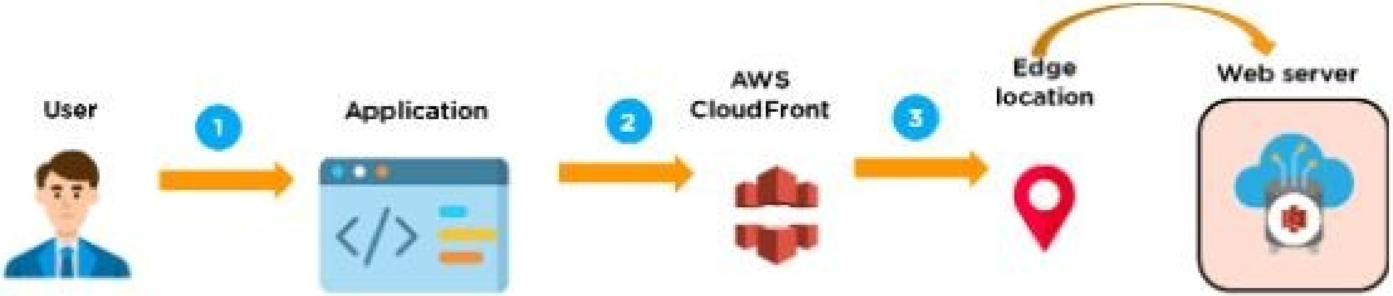
At edge location, CloudFront looks for its requested cache file. Once the file is found, CloudFront sends the file to the user.



Cache file - CloudFront

Step 4

But, if the file is not found then CloudFront compares the requirements with the specifications and shares it with the respective server.

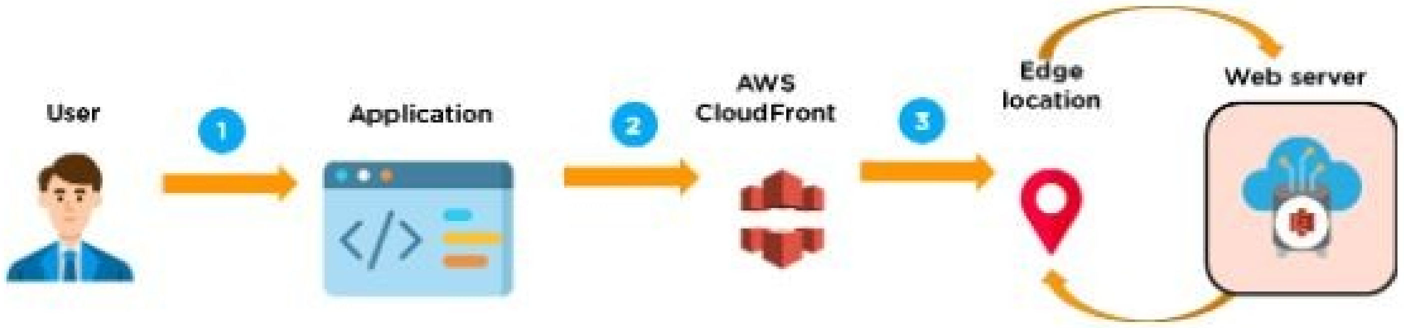


Web server - CloudFront

How Does AWS CloudFront Work?

Step 5

The web server responds to the request by sending the files back to the CloudFront edge location.



Share File

Step 6:

As soon CloudFront receives the file, it shares it with the client and adds the file to the edge location.

Create a Distribution Network for CloudFront