Unit-6
Cloud Network and Security

-Dr. Radhika V. Kulkarni

Associate Professor, Dept. of Computer Engineering,

Vishwakarma Institute of Technology, Pune.

DISCLAIMER

This presentation is created as a reference material for the students of TY-Comp. Engg., VIT (AY 2022-23 Sem-2).

It is restricted only for the internal use and any circulation is strictly prohibited.

Syllabus

Unit-VI Cloud Network and Security

[CO6: PO1, PO2, PO3, PO4, PO5, PO10, PO12 - Strength 2,2,1,3,1,3]

Introduction to networking in the cloud, defining a Virtual Private Cloud, Public and private IP address basics, Google’s network architecture, Routes and firewall rules in the cloud, Multiple VPC networks, building hybrid clouds using VPNs, interconnecting, and direct peering, Different options for load balancing. Introduction to security in the cloud, the shared security model, Encryption options, Authentication and authorization with Cloud IAM, Identify Best Practices for Authorization using Cloud IAM. [6 Hrs]

Basics of Network Addressing

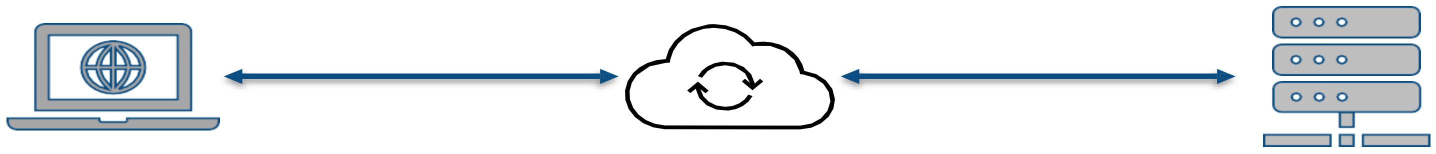
Public IP, Private IP, CIDR

For more details refer to: Google Cloud Documentation.
AWS documentation

Basics of IP Addressing

- An IP address (Internet Protocol address) is a numerical identifier, such as 192.0.2.1, that is associated with a computer network that communicates using the Internet Protocol.
- IP address is used for two purposes:
  1. Identifying a host or network interface and
  2. Addressing a specific location.

- Public IP Address: Your internet service provider assigns a public IP address to your network router so that it may be accessed directly over the internet (ISP).
- Public IP address is an **external addresses that uniquely identifies a device on the internet.**
- Devices that communicate directly over the internet require a Public IP address.
- Public IP address is globally unique.

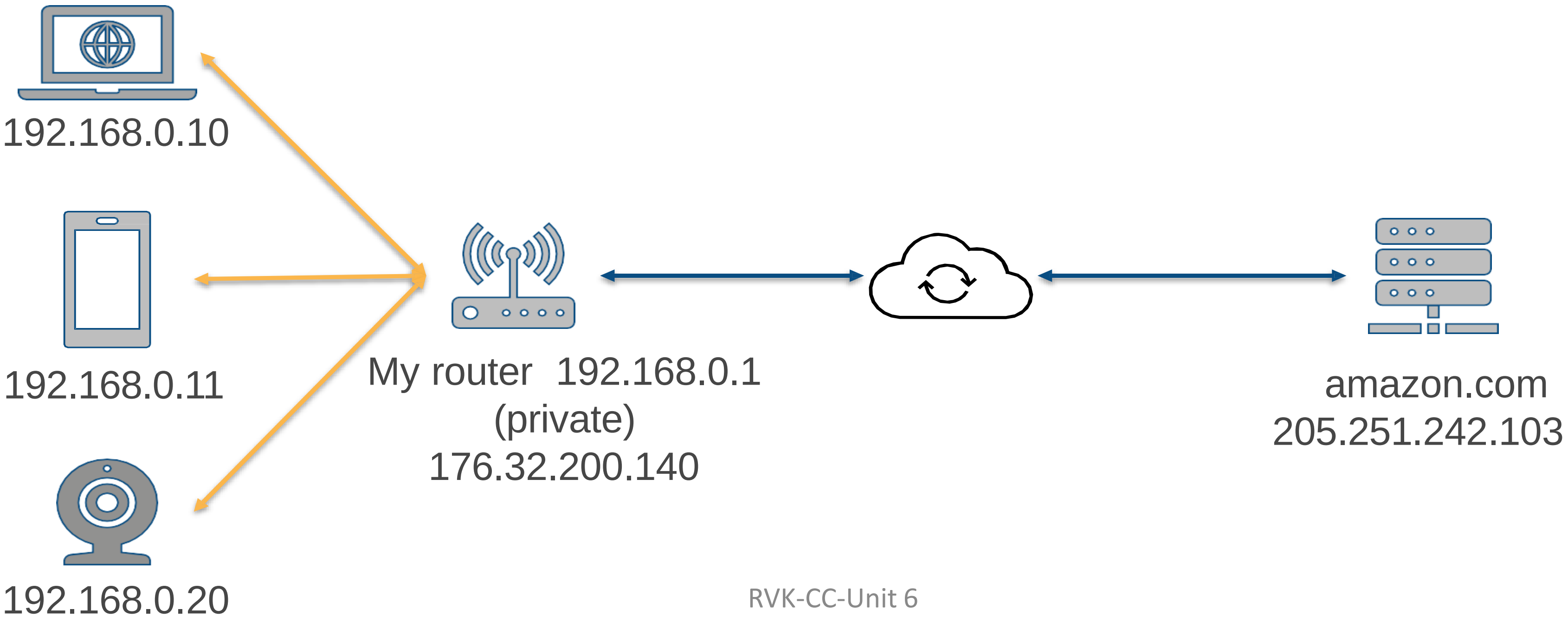


My laptop
176.32.200.140

amazon.com
205.251.242.103

Basics of IP Addressing (cont..)

- Private IP Address: The address that your network router provides to your device is known as a private IP address. Each device on the same internal network is given a unique private IP address (also known as a private network address) that allows them to communicate with one another.
- Private IP addresses are **internal addresses which enable devices on the same network to interact without needing to connect to the internet.**
- Private IP address is reusable – unique only within the private network.
- Router talks to both internet and private network – so has two IP addresses.



Private IP Address	Public IP Address
Private IP Address is used to communicate within the network and hence the scope is local.	Public IP Address is used to communicate outside the network and hence the scope is global.
Private IP Addresses differ in a uniform manner.	Public IP Addresses differ in varying range.
Local Network Operator creates private IP addresses using network operating system.	Internet Service Provider (ISP) controls the public IP address.
Private IP Addresses are free of cost.	Public IP Address comes with a cost.
Private IP Address can be located using “ipconfig” command.	Public IP Address needs to be searched as “what is my ip” on search engine like google.
Private IP Address range: Class A: 10.0.0.0 – 10.255.255.255, Class B: 172.16.0.0 – 172.31.255.255, Class C: 192.168.0.0 – 192.168.255.255	Except private IP Addresses, rest IP addresses are public.
Private IP address is not unique and can be reused. It is unique within the network.	Public IP uses a numeric code that is unique and cannot be used by other
Private IP addresses require Network Address Translation (NAT) to communicate with devices	Public IP does not require a network translation
Private IP address is an internal address and hence more secure.	Public IP address is an external address and hence has no security.

CIDR

- The Classless Inter-Domain Routing (CIDR) is a method to calculate the no. of IP addresses and an efficient way of allocating IPs in the network.
- This addressing scheme was introduced in the year 1993 by the Internet Engineering Task Force (IETF) to prevent the wastage of IPv4 addresses and to prevent the complexity of routing tables. It replaces the outdated way of a classful addressing system.

- Attributes of the CIDR addressing:
  - The IP addresses in a CIDR block are to be continuous, as the ISP will provide them in a sequence of numbers, to minimize the wastage of IP addresses.
  - The size of the CIDR Block should be of power 2, and to identify the number of addresses assigned, check the subnet mask of the IP address.
  - Syntax of CIDR Block is **IP address/Subnet mask**.
  - E.g. **10.0.0.0/24 is a CIDR Block**. Here, **24 is subnet mask**.

    The **number of IP addresses in the CIDR block = 2^n ;** where **n = 32 – subnet mask**

    Here, n = 32 – 24 = 8. So, the number of IP addresses in the block 10.0.0.0/24 are **2^8 = 256**

    i.e. the IP range of this CIDR block is **10.0.0.0 to 10.0.0.255**

    The **number of usable IP's is (2^n) – 2**.

    So, usable IP address range is **256-2 = 254** because the first IP of the range will be reserved as a Network address, and the last IP will be reserved as a Broadcast address by the system, and they cannot be used for devices.

Subnet

- Network can be sub-divided into subnets inside an organization.
- Subnetting aids in manageability, security, isolation and so forth.

Q. Divide a network CIDR: 193.239.32.0/20 into four subnets.

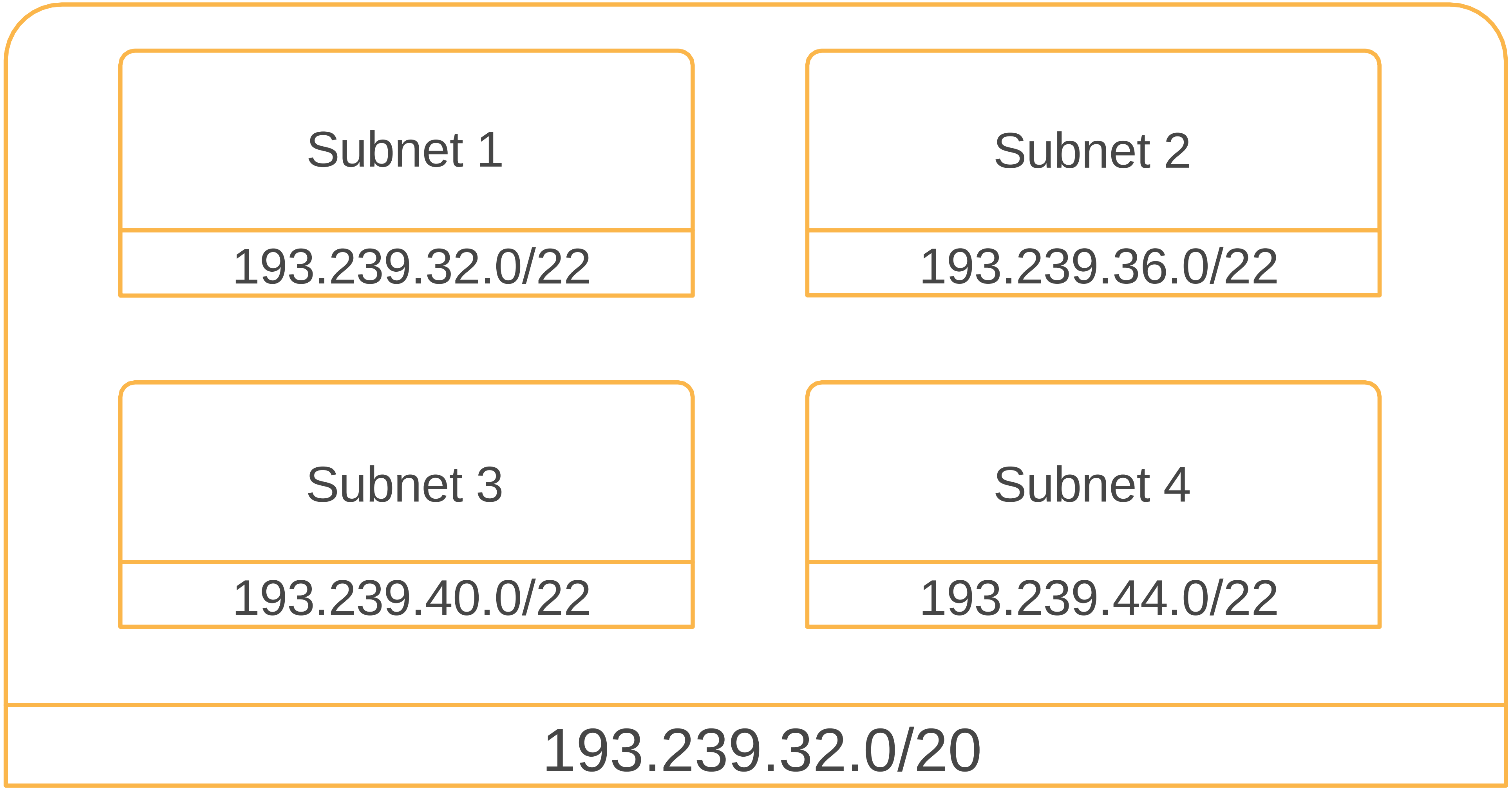
Ans.

Additional two-bits are needed to indicate the subnets.

Subnet CIDR: 193.239.32.0/22

193.239.32.0/22	11000001.11101111.0010**00**00.00000000
193.239.36.0/22	11000001.11101111.0010**01**00.00000000
193.239.40.0/22	11000001.11101111.0010**10**00.00000000
193.239.44.0/22	11000001.11101111.0010**11**00.00000000

Subnet Example



Subnet 1

193.239.32.0/22

Subnet 2

193.239.36.0/22

Subnet 3

193.239.40.0/22

Subnet 4

193.239.44.0/22

193.239.32.0/20

1,019 hosts in each subnet (1024 – 5)

Networking in the Cloud

Cloud Networking

- It is a type of IT infrastructure in which some or all of an organization’s network capabilities and resources are hosted in a public or private cloud platform, managed in-house or by a service provider, and available on demand.

- Companies can either use on-premises cloud networking resources to build a private cloud network or use cloud-based networking resources in the public cloud, or a hybrid cloud combination of both.

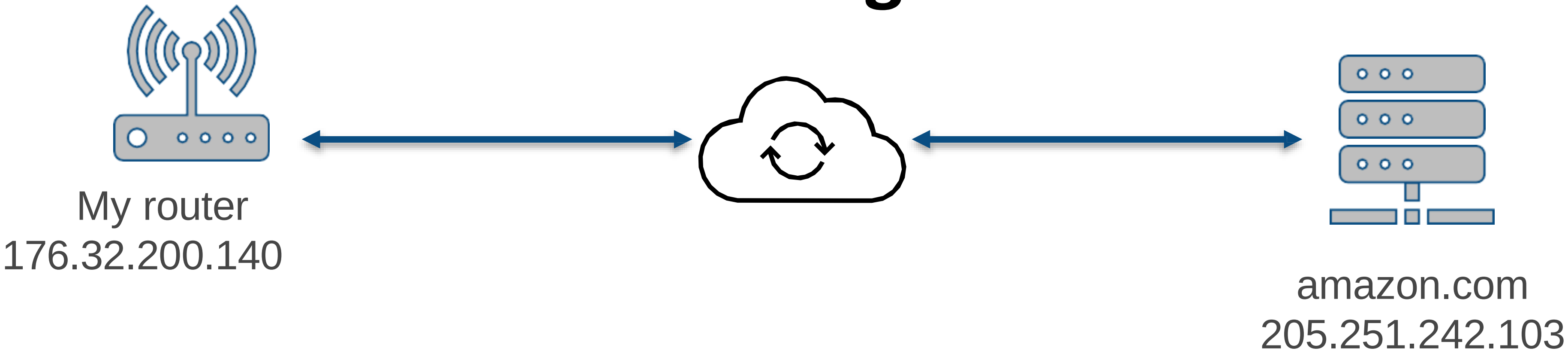
- These network resources can include virtual routers, firewalls, and bandwidth and network management software, with other tools and functions available as required.

Network Addressing in Cloud

- Managing IP addresses is one of the most challenging aspects of cloud migrations and management.
  - **In Azure usable IP Range is (2^n) – 5** instead of (2^n) – 2. These 5 addresses are reserved for the below purposes:
    - Zero IP – Network Address
    - First IP – DHCP Address
    - Second IP – Router Address
    - Third IP – Microsoft reserves it for future purposes.
    - Last (255) IP – Broadcast Address.
    - E.g. In Azure the usable IPs in the **10.0.0.0/24 CIDR Block** would be from **10.0.0.4 to 10.0.0.254.**

- Aside from their obvious role in network reachability, IP addresses are also used to identify resources, isolate organizations and services, and to apply policies.
- The ideals of cloud, which promise portability, programmability, and on-demand scalability are often at odds with those traditional uses.
- For most enterprises, IP address exhaustion, overlapping subnets, and policy constraints can create risks or even cause projects to grind to a halt.

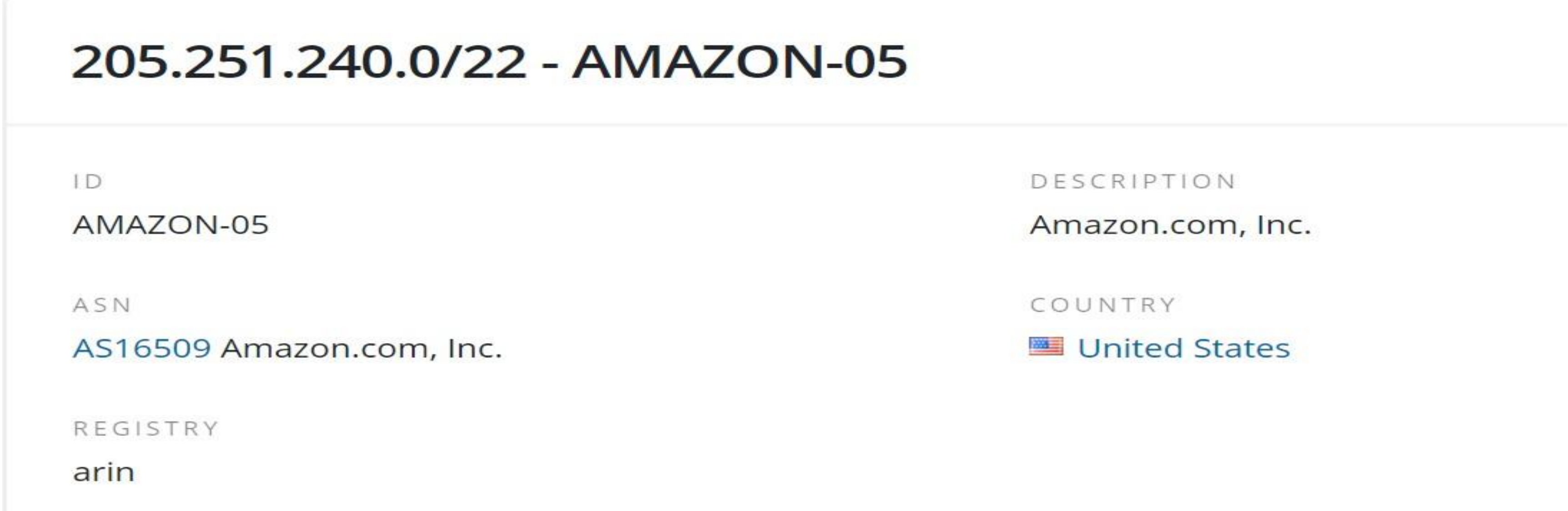
Routing



My router
176.32.200.140

amazon.com
205.251.242.103

- IP address identifies the country, organization and device



205.251.240.0/22 - AMAZON-05

ID
AMAZON-05

DESCRIPTION
Amazon.com, Inc.

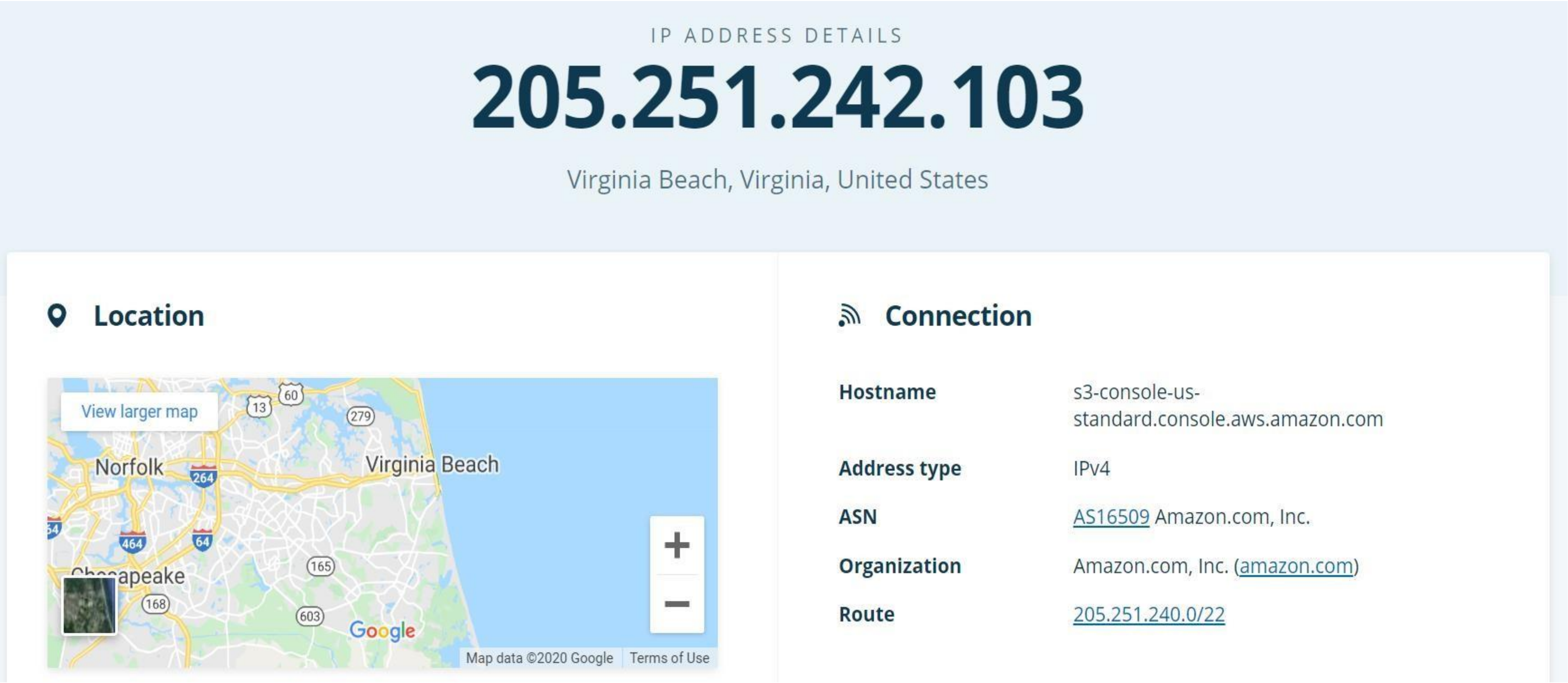
ASN
AS16509 Amazon.com, Inc.

COUNTRY
United States

REGISTRY
arin

Source https://ipinfo.io/AS16509/205.251.240.0/22

IP Country, Location and Organization Lookup



IP ADDRESS DETAILS

205.251.242.103

Virginia Beach, Virginia, United States

Location

Connection

Hostname	s3-console-us-standard.console.aws.amazon.com
Address type	IPv4
ASN	AS16509 Amazon.com, Inc.
Organization	Amazon.com, Inc. (amazon.com)
Route	205.251.240.0/22

Source https://ipinfo.io/205.251.242.103

CIDR Example

Q. Identify the network of IP address 205.251.242.103.

Ans.

IP address 205.251.242.103 is part of the 205.251.240.0/22.

/22 indicates that the first 22 bits of IPv4 is used as the network identifier.

CIDR Block 205.251.240.0/22 was assigned to Amazon

11001101.11111011.11110000.00000000

IP 205.251.242.103 belongs to Amazon

11001101.11111011.111100**10.01100111**

Virtual Private Cloud

For more details refer to: AWS VPC
Google Cloud VPC

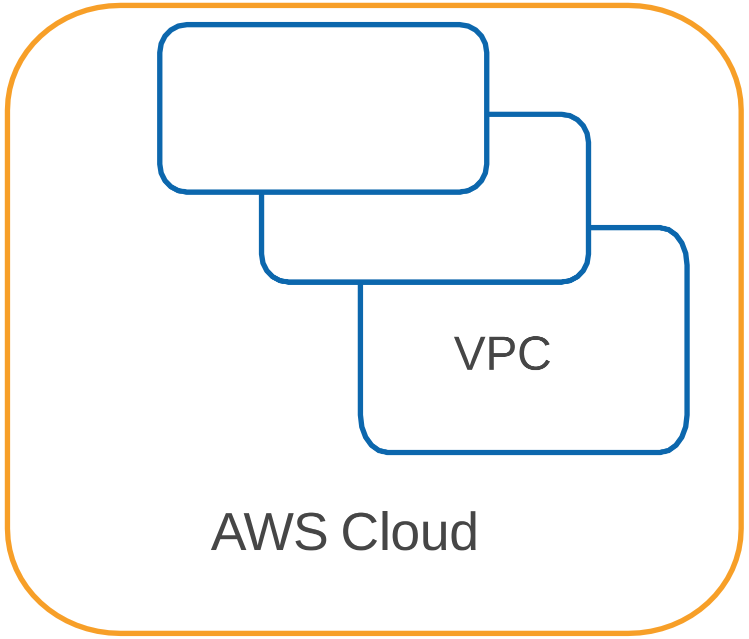
Public, Private and Virtual Private Cloud

- A public cloud is shared cloud infrastructure. Multiple customers of the cloud vendor access that same infrastructure, although their data is not shared. It is known as “multitenancy”.

- A private cloud is single-tenant. It is a cloud service that is exclusively offered to one organization.

- A virtual private cloud (VPC) is a private cloud within a public cloud; no one else shares the VPC with the VPC customer. A VPC isolates computing resources from the other computing resources available in the public cloud.

- AWS Virtual Private Cloud uses Private IP address.
- Public IP address is assigned to servers that need to communicate over internet.



Virtual Private Cloud

- The key technologies for isolating a VPC from the rest of the public cloud are:
  - Subnets: A subnet is a range of IP addresses within a network that are reserved so that they're not available to everyone within the network, essentially dividing part of the network for private use. In a VPC these are private IP addresses that are not accessible via the public Internet, unlike typical IP addresses, which are publicly visible.
  - VLAN: A virtual LAN (VLAN) is a type of subnetwork group geographically separate devices together. Like a subnet, VLAN is a way of partitioning a network, but the partitioning takes place at a different layer within the OSI model (layer 2 instead of layer 3).
  - VPN: The virtual private network (VPN) technology is a service which uses encryption to create a private network over the top of a public network. VPN traffic passes through publicly shared Internet infrastructure – routers, switches, etc. – but the traffic is scrambled and not visible to anyone.
- A VPC will have a dedicated subnet and VLAN that are only accessible by the VPC customer. This prevents anyone else within the public cloud from accessing computing resources within the VPC – effectively placing the “Reserved” sign on the table.
- The VPC customer connects via VPN to their VPC, so that data passing into and out of the VPC is not visible to other public cloud users.

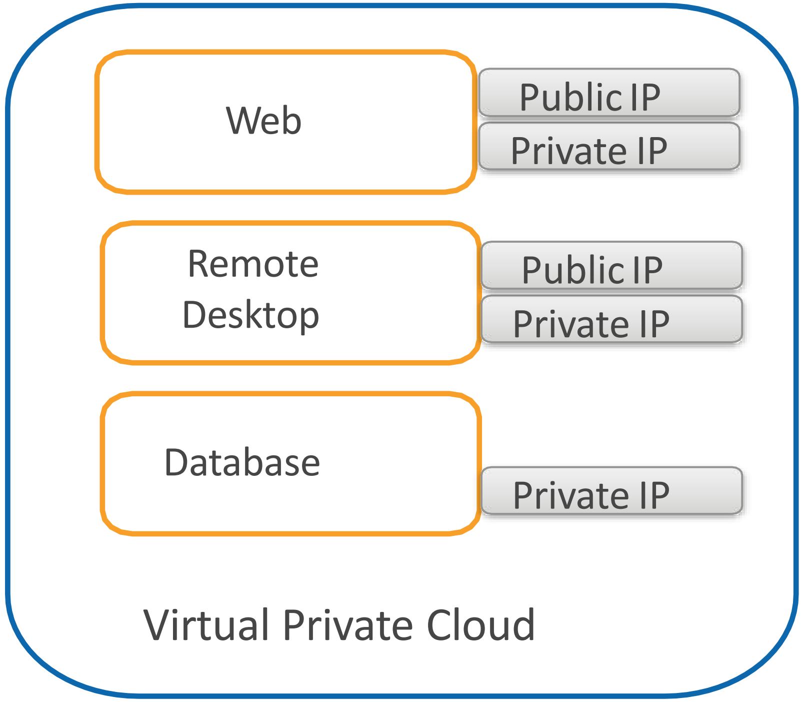
Virtual Private Cloud (cont..)

- Some VPC providers offer additional customization with:
  - Network Address Translation (NAT): This feature matches private IP addresses to a public IP address for connections with the public Internet. With NAT, a public-facing website or application could run in a VPC.
  - BGP (Border Gateway Protocol) route configuration: Some providers allow customers to customize BGP routing tables for connecting their VPC with their other infrastructure.

- **Advantages of using a VPC instead of a private cloud:**
  - Scalability: Because a VPC is hosted by a public cloud provider, customers can add more computing resources on demand.
  - Easy hybrid cloud deployment: It's relatively simple to connect a VPC to a public cloud or to on-premises infrastructure via the VPN.
  - Better performance: Cloud-hosted websites and applications typically perform better than those hosted on local on-premises servers.
  - Better security: The public cloud providers that offer VPCs often have more resources for updating and maintaining the infrastructure, especially for small and mid-market businesses.

Virtual Private Cloud (cont..)

- Web and Remote Desktop connected to the internet.
- Database is accessible only in the private network.



Web

Public IP

Private IP

Remote Desktop

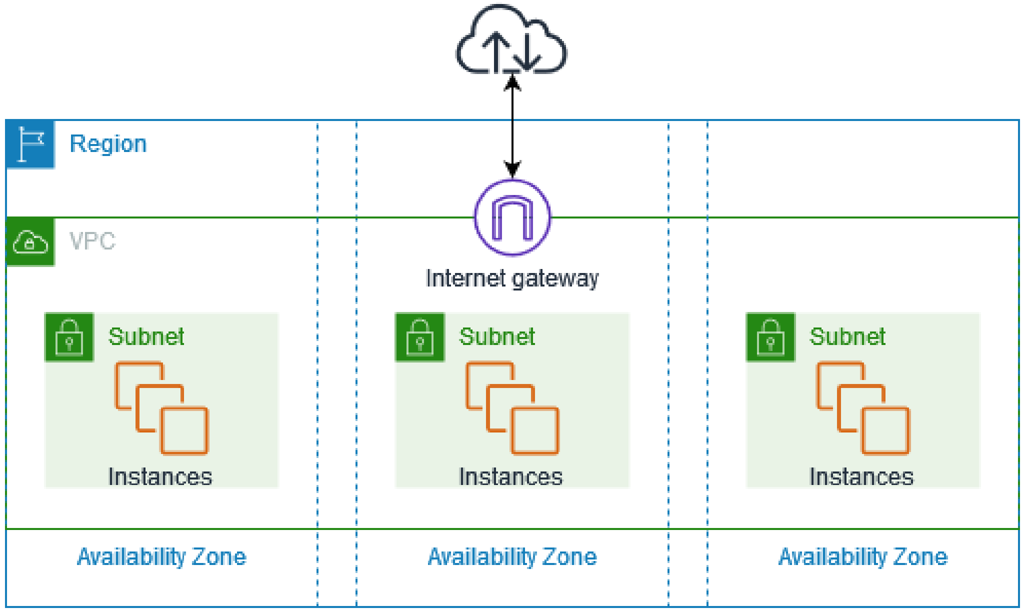
Public IP

Private IP

Database

Private IP

Virtual Private Cloud



- Amazon VPC shown in this figure has one subnet in each of the Availability Zones in the Region, EC2 instances in each subnet, and an internet gateway.

Amazon VPC Configuration Features

- Virtual private clouds (VPC): A VPC is a virtual network that closely resembles a traditional network that you'd operate in your own data center. After you create a VPC, you can add subnets.

- Subnets: A subnet is a range of IP addresses in your VPC. A subnet must reside in a single Availability Zone. After you add subnets, you can deploy AWS resources in your VPC.

- IP addressing: You can assign IP addresses, both IPv4 and IPv6, to your VPCs and subnets.

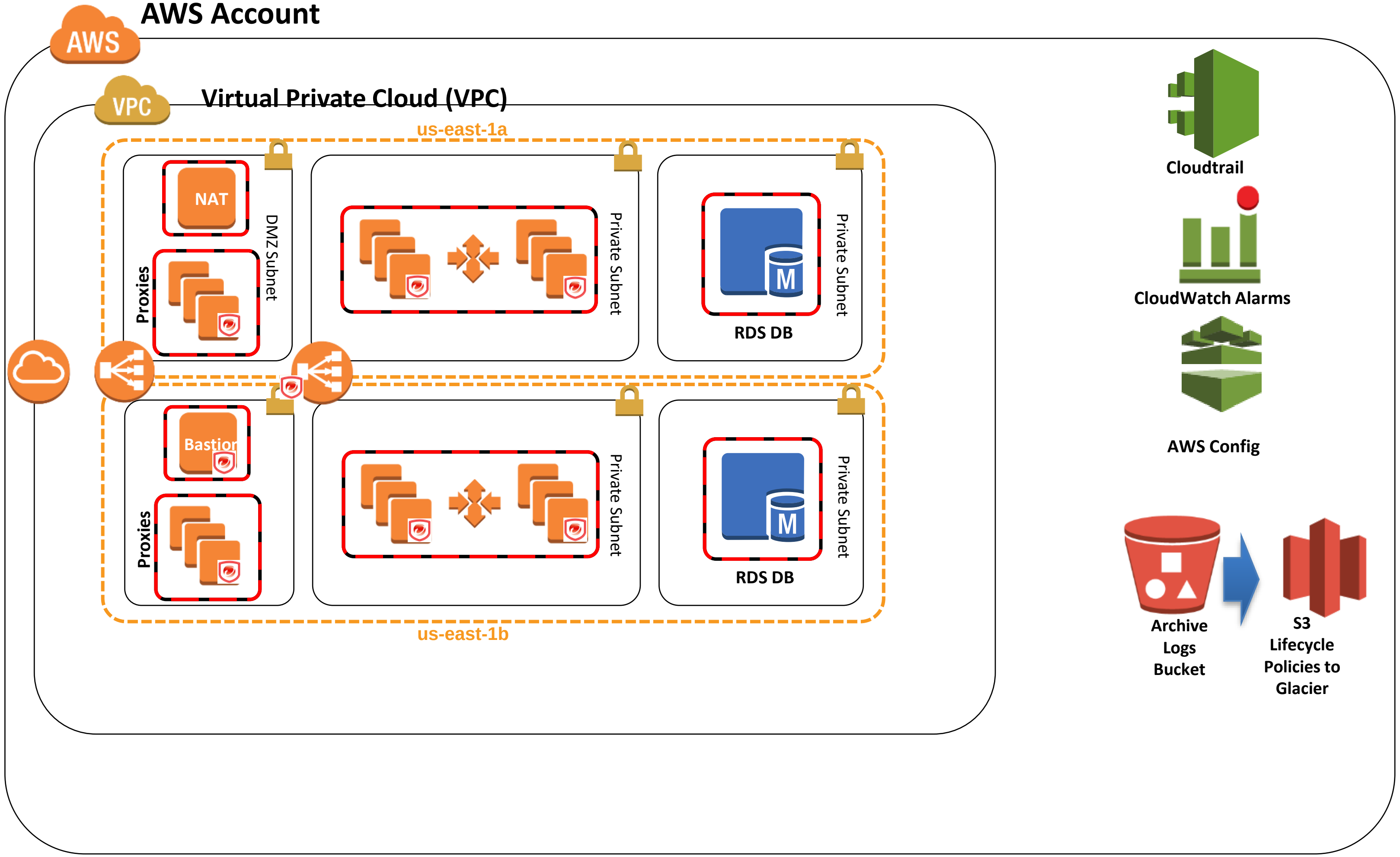
- Routing: Use route tables to determine where network traffic from your subnet or gateway is directed.

- Gateways and endpoints: A gateway connects your VPC to another network. For example, use an internet gateway to connect your VPC to the internet. Use a VPC endpoint to connect to AWS services privately, without the use of an internet gateway or NAT device.

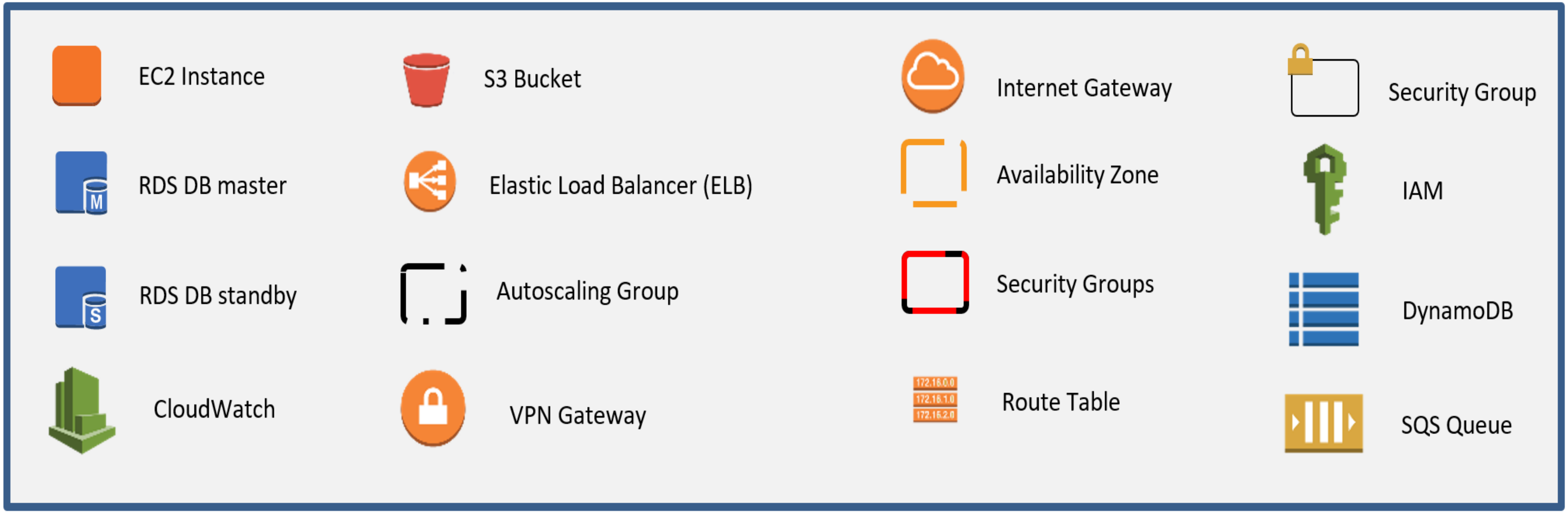
Amazon VPC Configuration Features(cont..)

- Peering connections: Use a VPC peering connection to route traffic between the resources in two VPCs.
- Traffic Mirroring: Copy network traffic from network interfaces and send it to security and monitoring appliances for deep packet inspection.
- Transit gateways: Use a transit gateway, which acts as a central hub, to route traffic between your VPCs, VPN connections, and AWS Direct Connect connections.
- VPC Flow Logs: A flow log captures information about the IP traffic going to and from network interfaces in your VPC.
- VPN connections: Connect your VPCs to your on-premises networks using AWS Virtual Private Network (AWS VPN).

Standard Architecture Deployed by AWS QuickStart



https://aws.amazon.com/architecture/icons/



EC2 Instance

RDS DB master

RDS DB standby

CloudWatch

S3 Bucket

Elastic Load Balancer (ELB)

Autoscaling Group

VPN Gateway

Internet Gateway

Availability Zone

Security Groups

Route Table

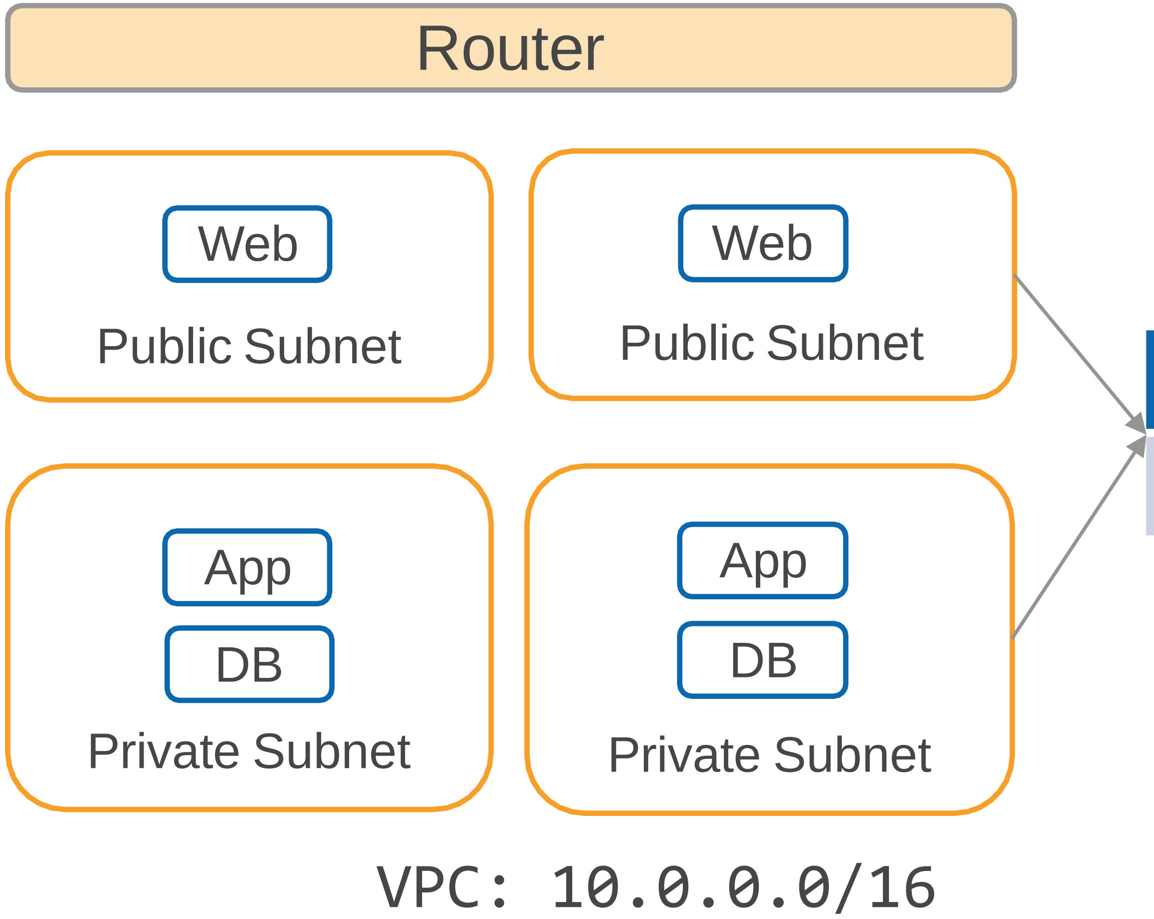
Security Group

IAM

DynamoDB

SQS Queue

VPC Router

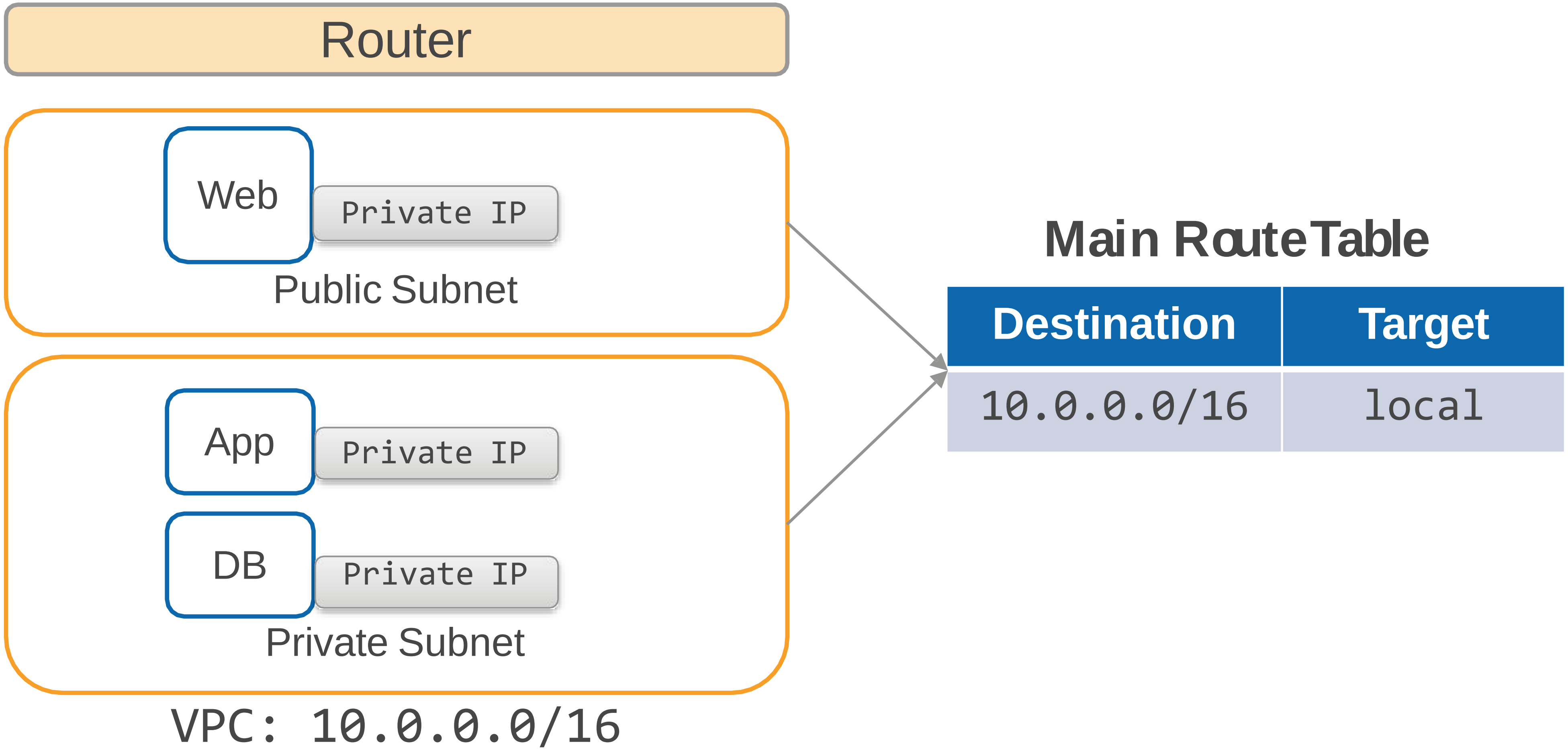


VPC: 10.0.0.0/16

Main Route Table

Destination	Target
10.0.0.0/16	local

VPC IP



Router

Web

Private IP

Public Subnet

App

Private IP

DB

Private IP

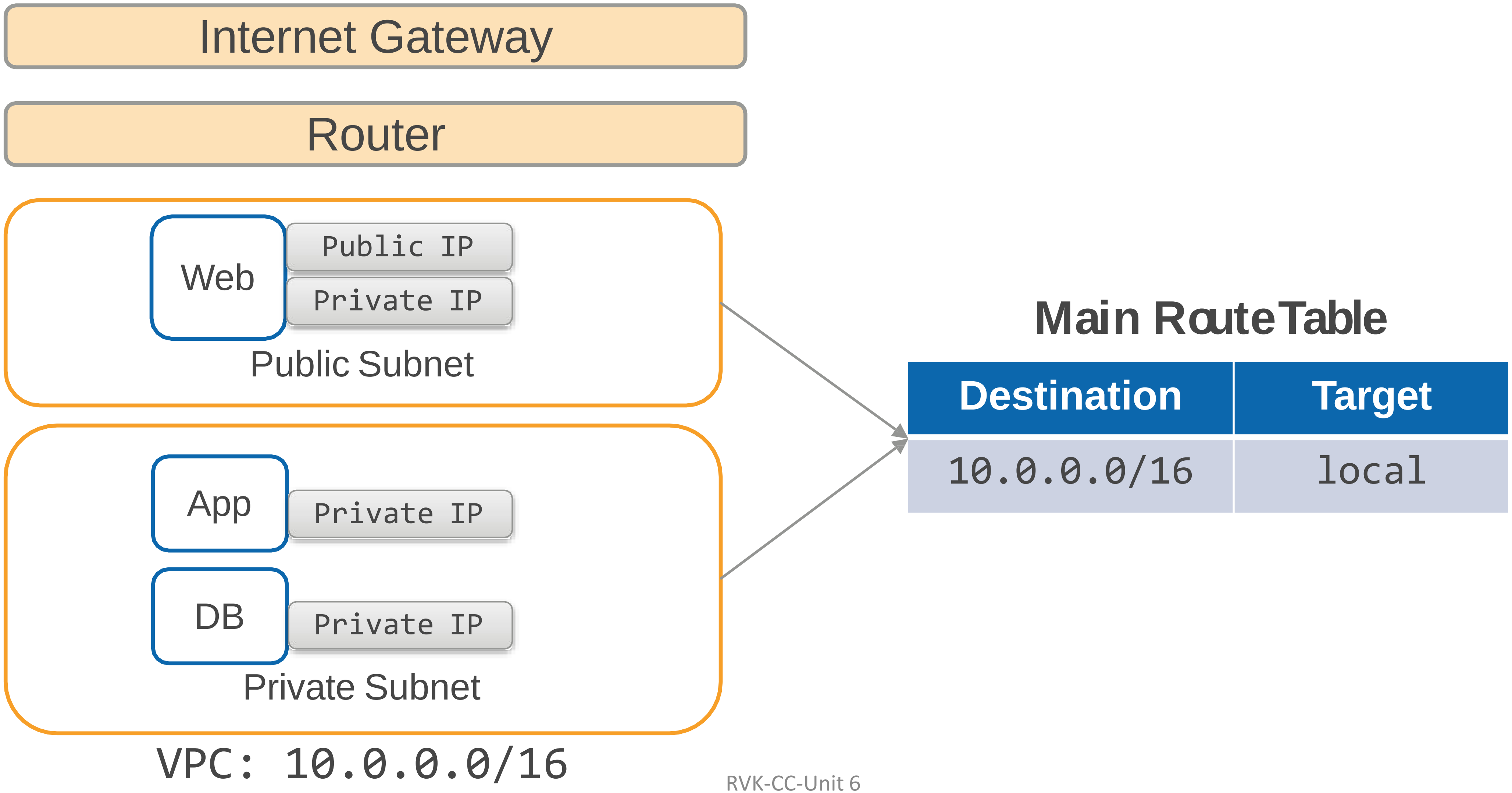
Private Subnet

VPC: 10.0.0.0/16

Main Route Table

Destination	Target
10.0.0.0/16	local

VPC Internet Gateway



Internet Gateway

Router

Web

Public IP

Private IP

Public Subnet

App

Private IP

DB

Private IP

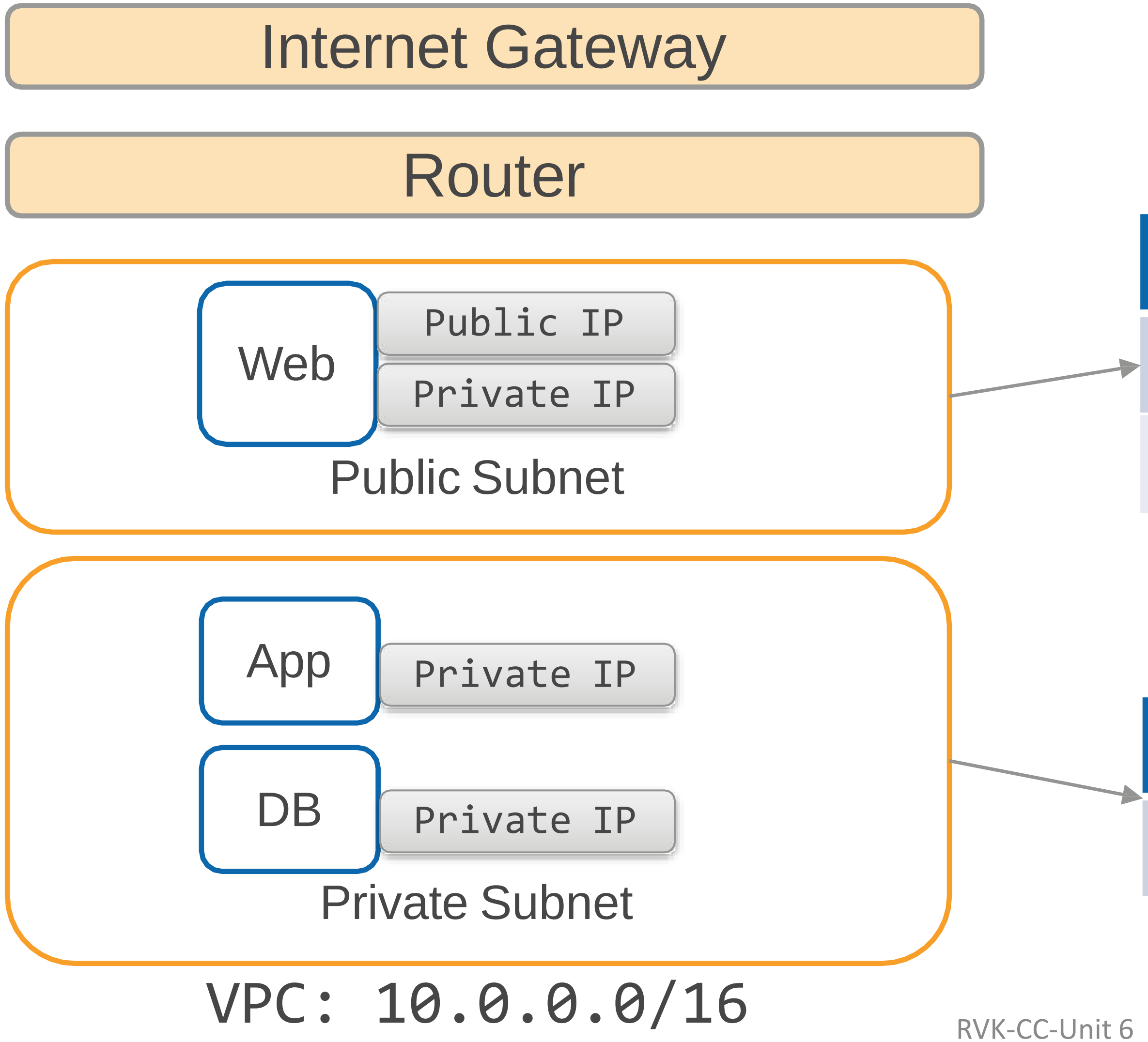
Private Subnet

VPC: 10.0.0.0/16

Main RouteTable

Destination	Target
10.0.0.0/16	local

VPC Internet Gateway Route



Internet Gateway

Router

Web

Public IP

Private IP

Public Subnet

App

Private IP

DB

Private IP

Private Subnet

VPC: 10.0.0.0/16

Public Route Table

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	IGW-id

Main Route Table

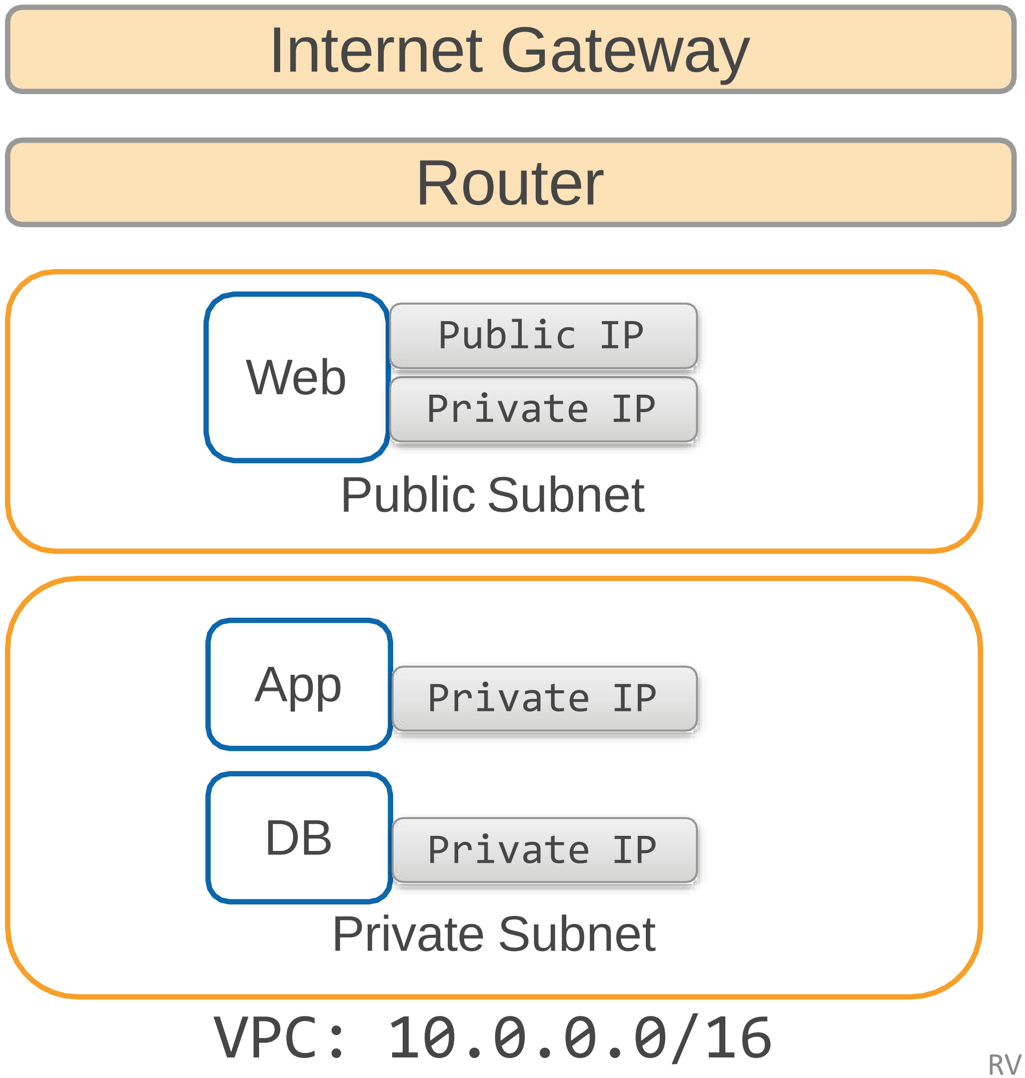
Destination	Target
10.0.0.0/16	local

Firewall

Security Group

Network Access Control List (NACL)

Firewall



Internet Gateway

Router

Web

Public IP

Private IP

Public Subnet

App

Private IP

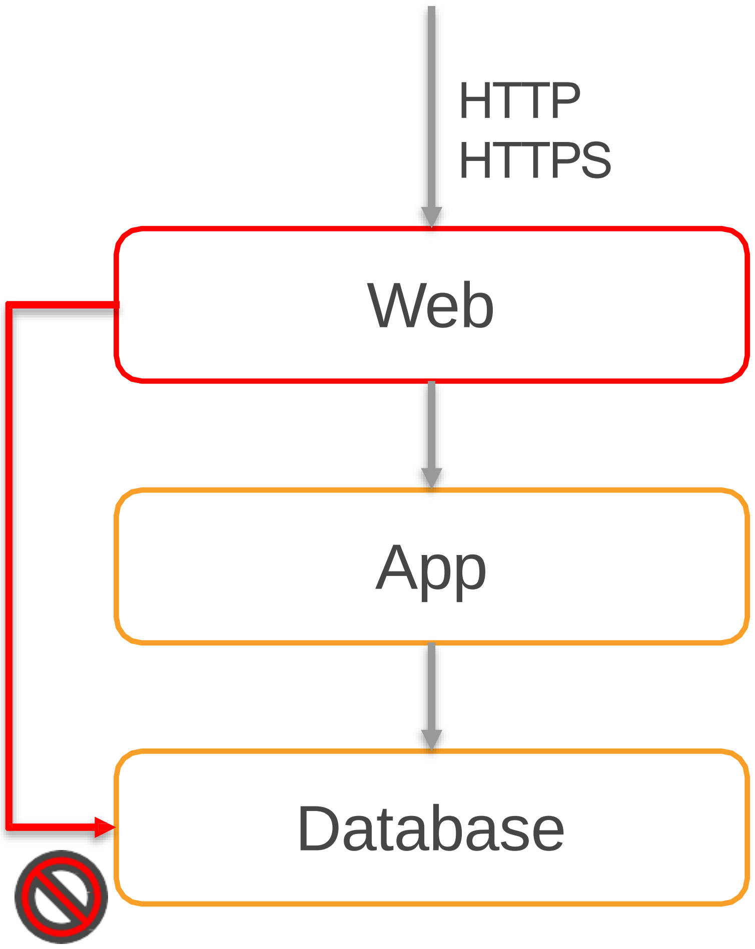
DB

Private IP

Private Subnet

VPC: 10.0.0.0/16

Security Group and Network ACL



HTTP

HTTPS

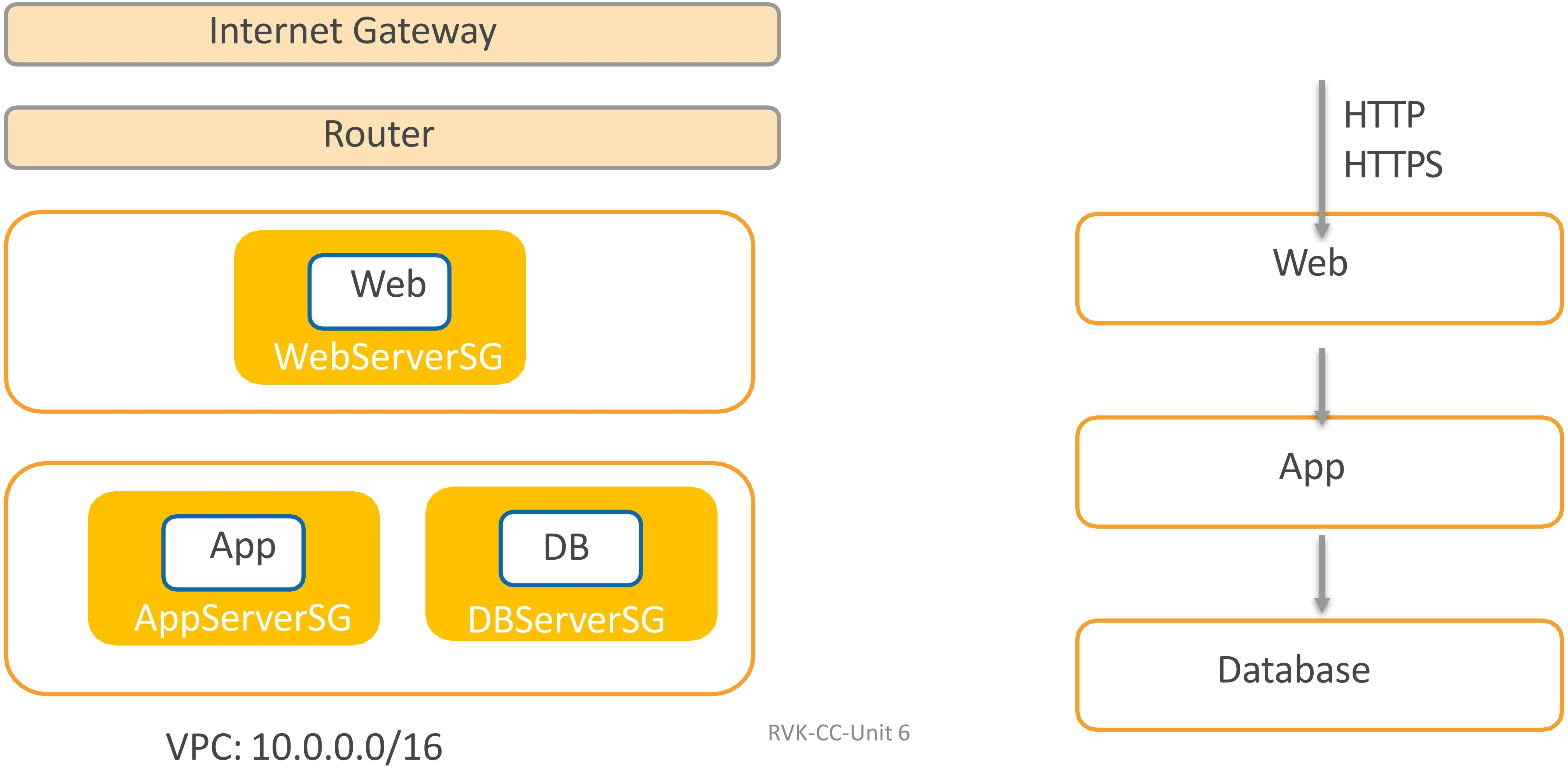
Web

App

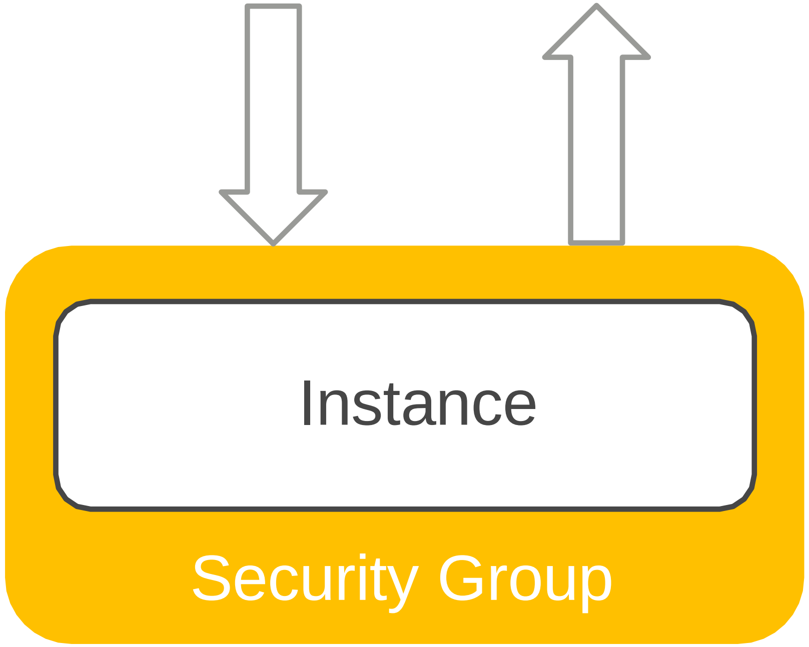
Database

Security Group

A security group acts as a firewall that controls the traffic allowed to and from the resources in your virtual private cloud (VPC). You can choose the ports and protocols to allow for inbound traffic and for outbound traffic.



Security Group – Instance Firewall



Specify what traffic is **ALLOWED**

Default Security Group

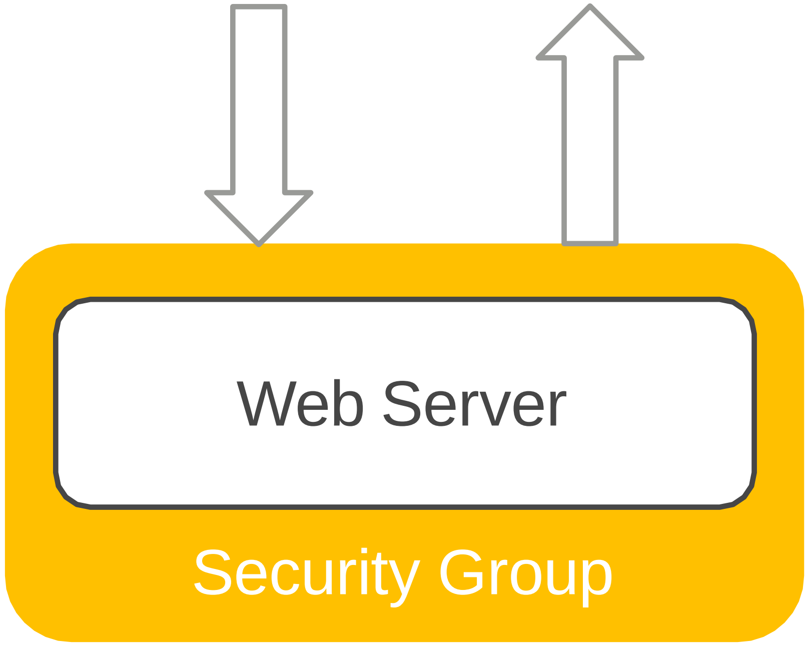
Inbound Rules

Source	Protocol	Port Range	Type
Default SG-ID	ALL	ALL	All Traffic

Outbound Rules

Destination	Protocol	Port Range	Type
0.0.0.0/0	ALL	ALL	All Traffic

Web Server Security Group



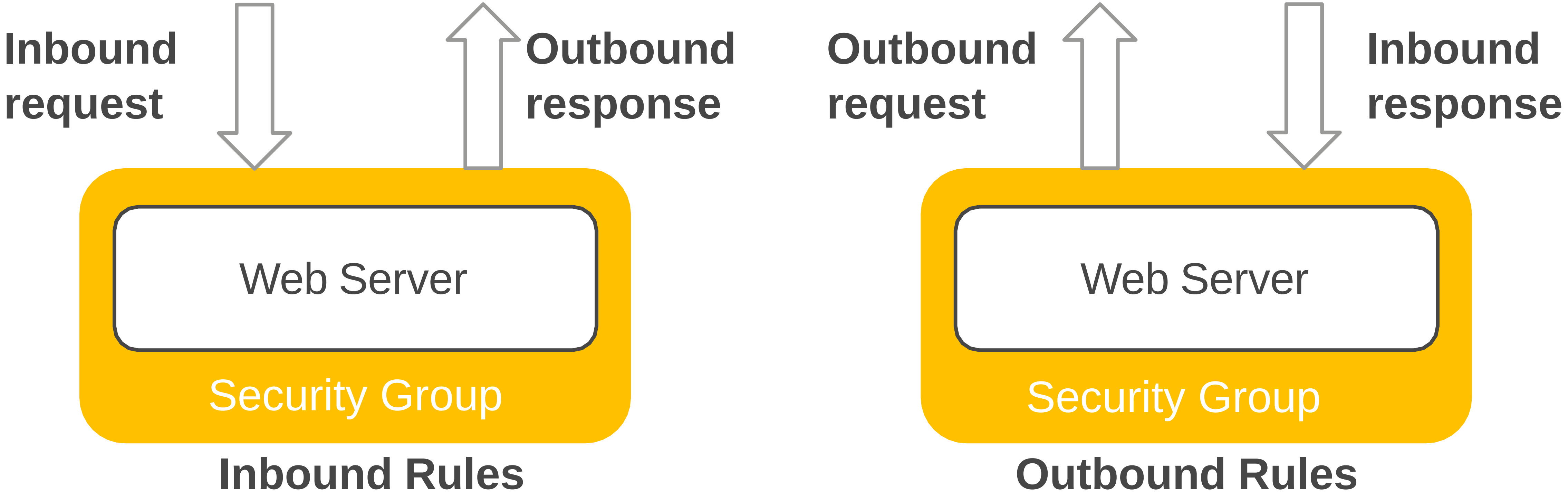
Inbound Rules

Source	Protocol	Port Range	Type
0.0.0.0/0	TCP	80	HTTP
0.0.0.0/0	TCP	443	HTTPS

Outbound Rules

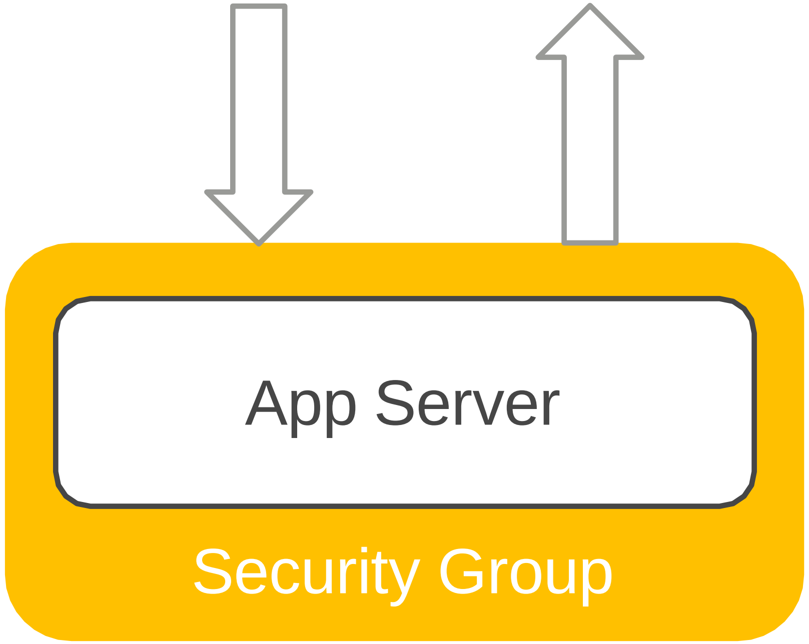
Destination	Protocol	Port Range	Type
0.0.0.0/0	ALL	ALL	All Traffic

Security Group is Stateful



If a request is allowed, the response for the request is automatically allowed

App Server Security Group



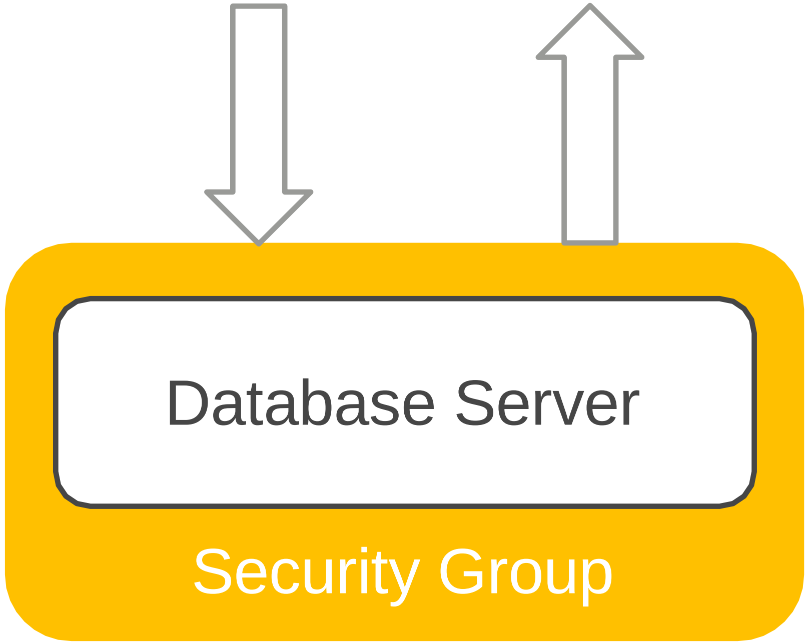
Inbound Rules

Source	Protocol	Port Range	Type
WebServerSG-ID	TCP	80	HTTP
WebServerSG-ID	TCP	443	HTTPS

Outbound Rules

Destination	Protocol	Port Range	Type
0.0.0.0/0	ALL	ALL	All Traffic

Database Server Security Group



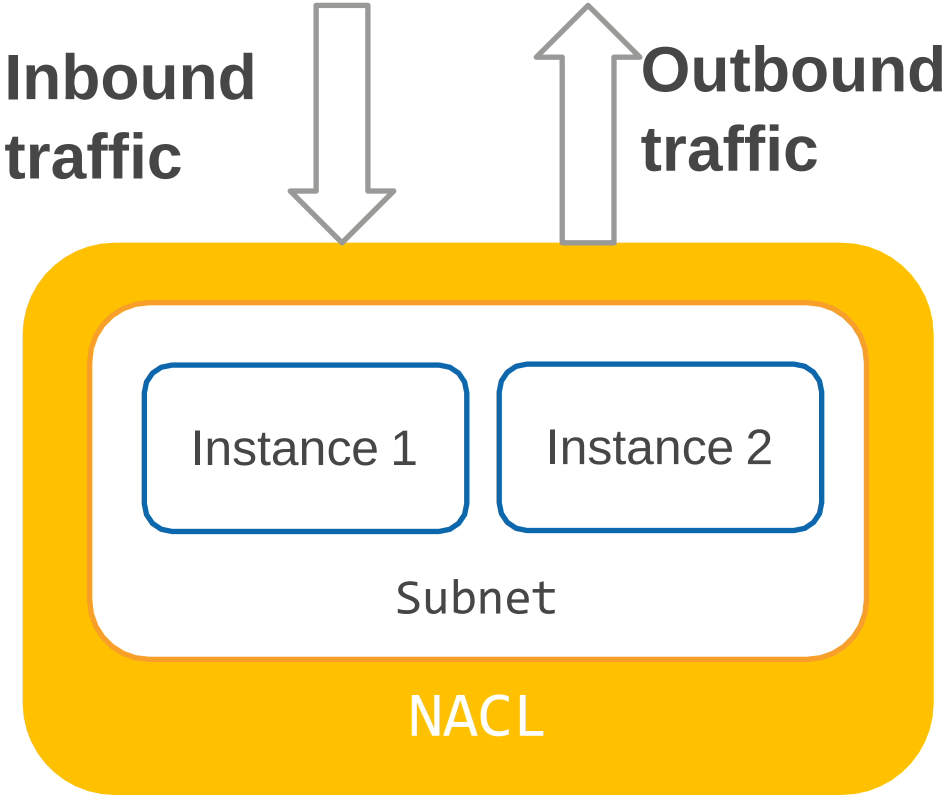
Inbound Rules

Source	Protocol	Port Range	Type
AppServerSG-ID	TCP	3306	MySQL Aurora

Outbound Rules

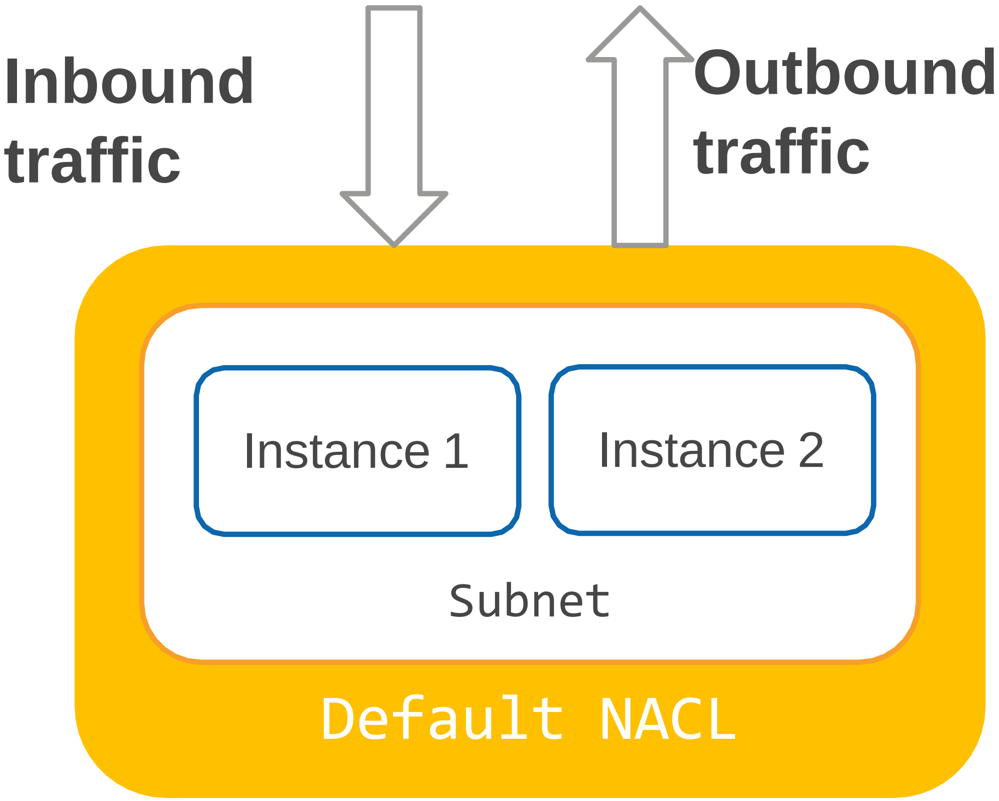
Destination	Protocol	Port Range	Type
0.0.0.0/0	ALL	ALL	All Traffic

Network Access Control List (NACL)– Subnet Firewall



- Specify what traffic is ALLOWED or DENIED in a subnet.
- All instances in the subnet are automatically protected.
- Stateless firewall – you need to allow both inbound and outbound traffic.
- Rules are evaluated in numeric order – lowest numbered rule that matches traffic decides the outcome.

Default Network ACL



Inbound traffic

Outbound traffic

Instance 1

Instance 2

Subnet

Default NACL

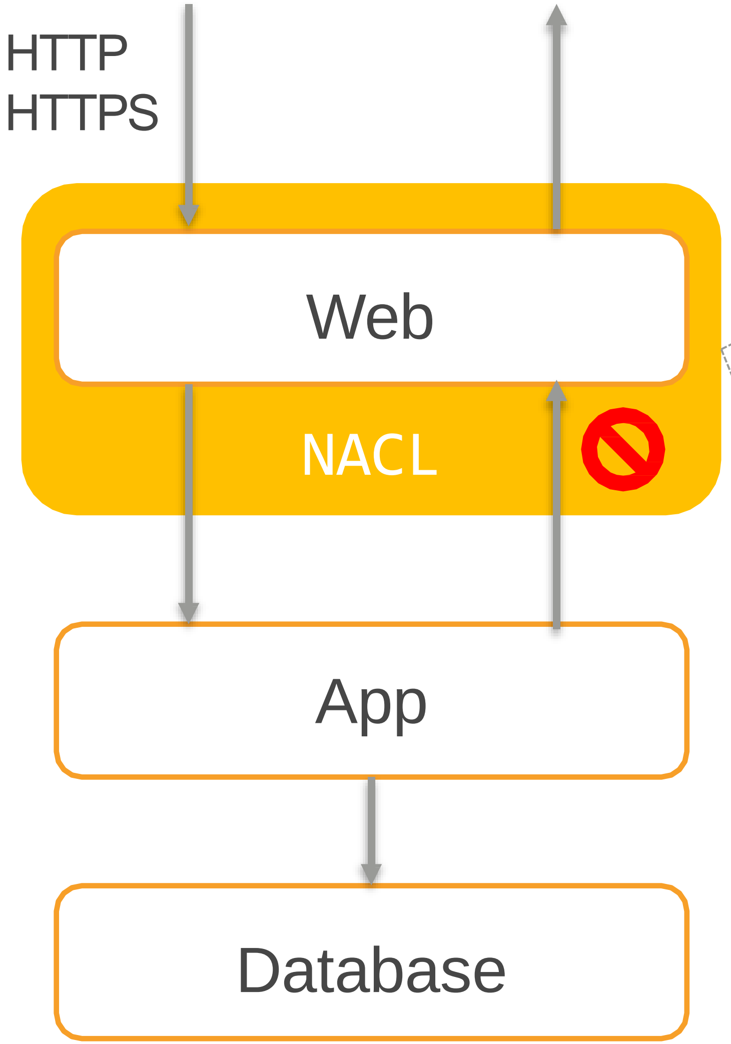
Inbound Rules

Rule #	Protocol	Port Range	Type	Source	Allow/ Deny
100	ALL	ALL	All Traffic	0.0.0.0/0	ALLOW
*	ALL	ALL	All Traffic	0.0.0.0/0	DENY

Outbound Rules

Rule #	Protocol	Port Range	Type	Destination	Allow/ Deny
100	ALL	ALL	All Traffic	0.0.0.0/0	ALLOW
*	ALL	ALL	All Traffic	0.0.0.0/0	DENY

Network ACL is tricky - Stateless



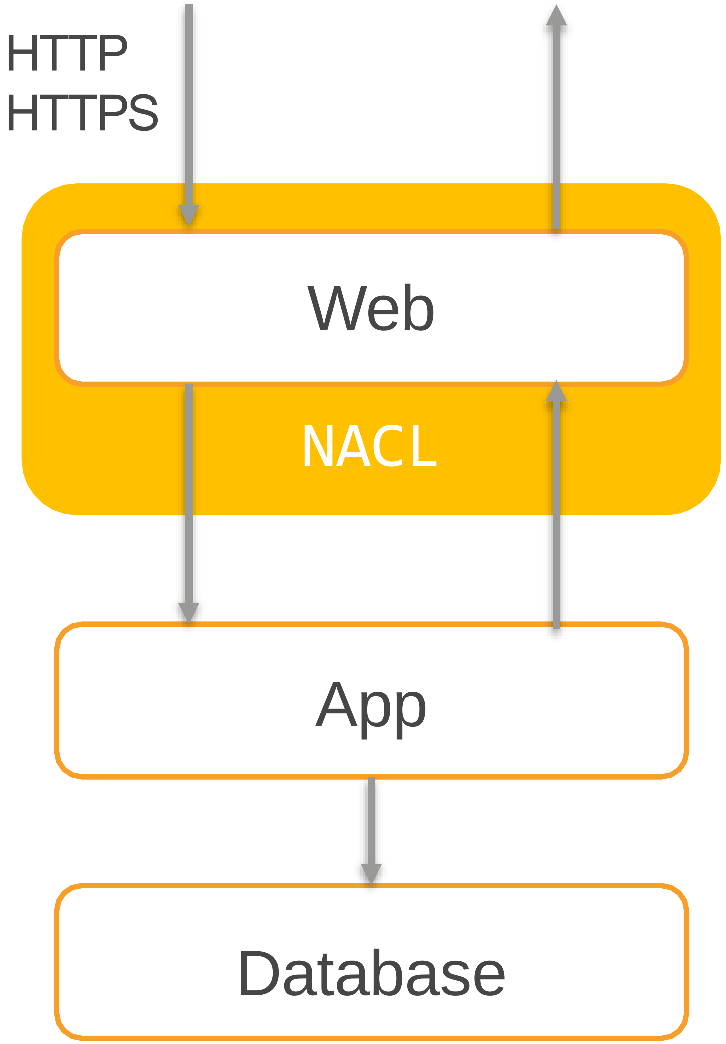
Public Subnet - Inbound Rules

Rule #	Protocol	Port Range	Type	Source	Allow/ Deny
100	TCP	80	HTTP	0.0.0.0/0	ALLOW
110	TCP	443	HTTPs	0.0.0.0/0	ALLOW
*	ALL	ALL	All Traffic	0.0.0.0/0	DENY

Public Subnet - Outbound Rules

Rule #	Protocol	Port Range	Type	Destination	Allow/ Deny
100	ALL	ALL	All Traffic	0.0.0.0/0	ALLOW
*	ALL	ALL	All Traffic	0.0.0.0/0	DENY

Network ACL – Fix Allow Local Traffic

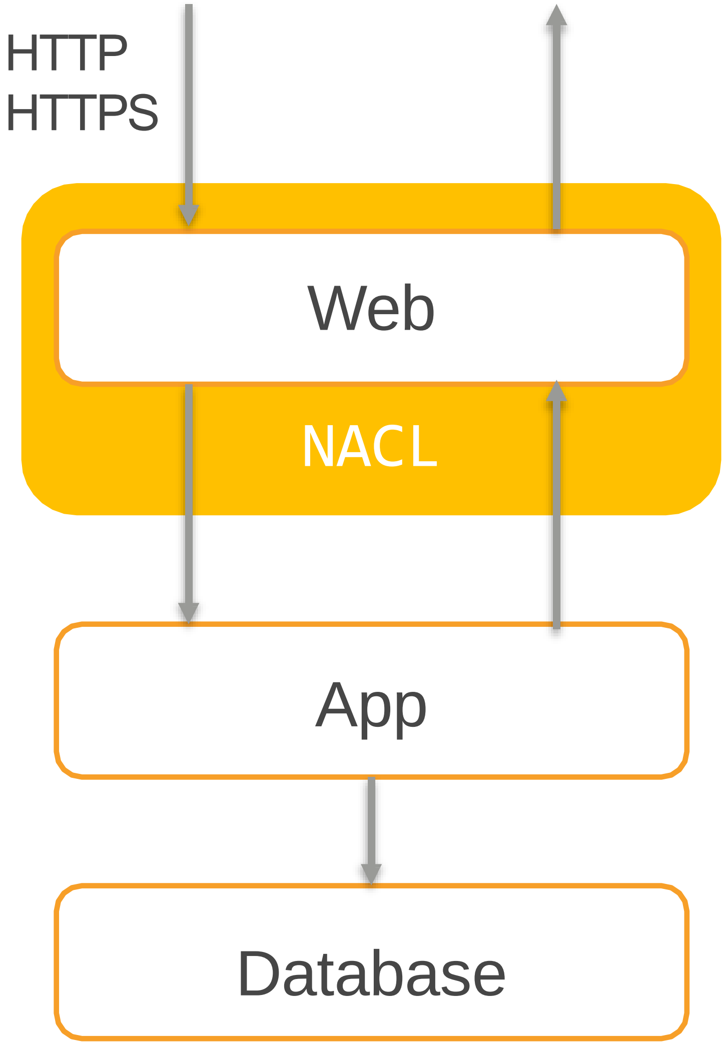


VPC: 10.0.0.0/16

Public Subnet - Inbound Rules

Rule #	Protocol	Port Range	Type	Source	Allow/ Deny
90	*ALL*	*ALL*	*All Traffic*	*10.0.0.0/ 16*	*ALLOW*
100	TCP	80	HTTP	0.0.0.0/0	ALLOW
110	TCP	443	HTTPS	0.0.0.0/0	ALLOW
*	ALL	ALL	All Traffic	0.0.0.0/0	DENY

Network ACL - Deny



DENY suspicious requests

Public Subnet - Inbound Rules

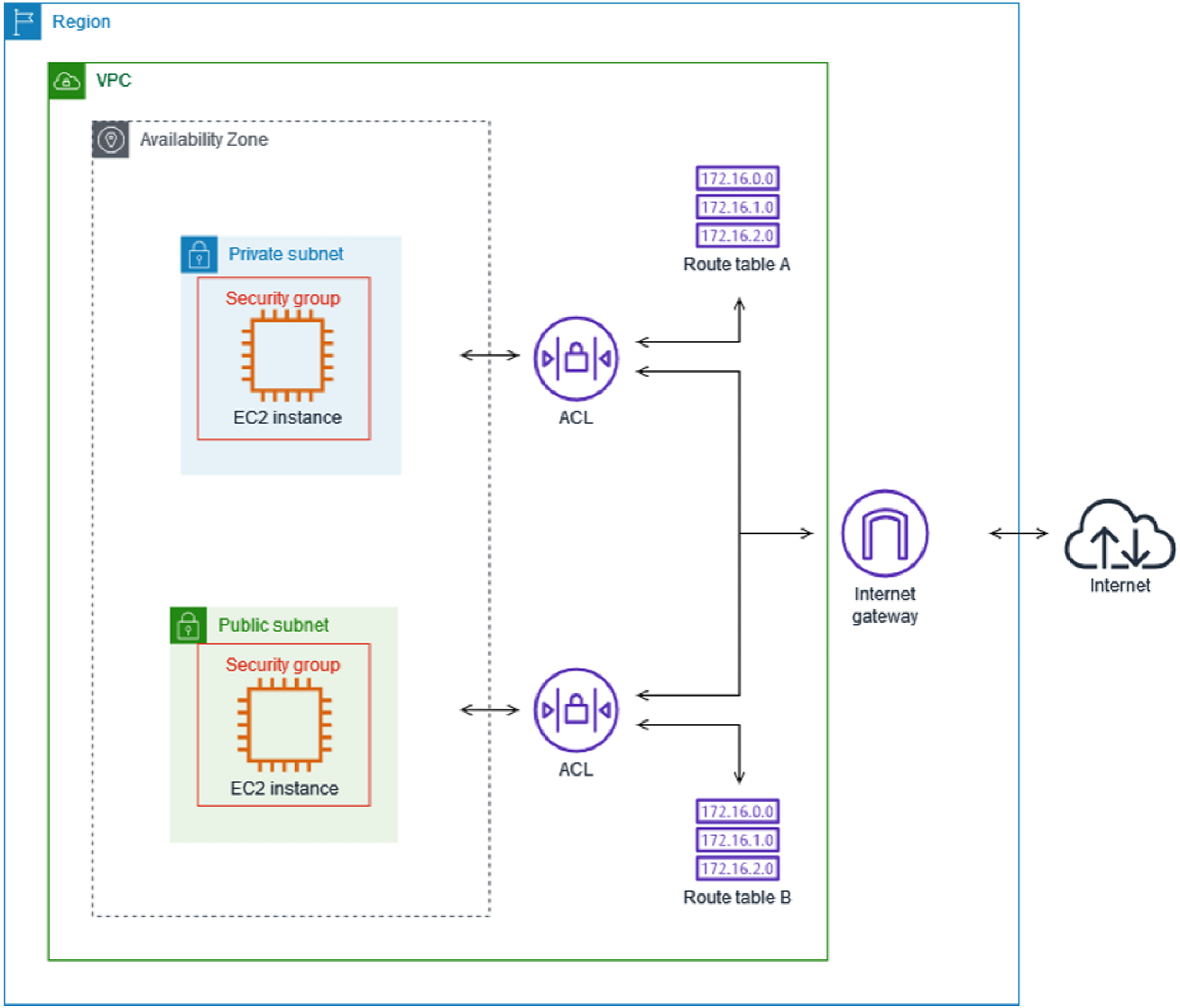
Rule #	Protocol	Port Range	Type	Source	Allow/ Deny
50	ALL	ALL	All Traffic	123.123.0.0/16	DENY
90	ALL	ALL	All Traffic	10.0.0.0/16	ALLOW
100	TCP	80	HTTP	0.0.0.0/0	ALLOW
110	TCP	443	HTTPS	0.0.0.0/0	ALLOW
*	ALL	ALL	All Traffic	0.0.0.0/0	DENY

Security Group and Network ACL

- Traffic from an internet gateway is routed to the appropriate subnet using the routes in the routing table.

- The rules of the network ACL that is associated with the subnet control which traffic is allowed to the subnet.

- The rules of the security group that is associated with an instance control which traffic is allowed to the instance.



Security Group Vs Network ACL

Security Group	Network ACL
Operates at the instance level.	Operates at the subnet level
Applies to an instance only if it is associated with the instance.	Applies to all instances deployed in the associated subnet (providing an additional layer of defense if security group rules are too permissive)
Supports allow rules only.	Supports allow rules and deny rules.
Evaluates all rules before deciding whether to allow traffic.	Evaluates rules in order, starting with the lowest numbered rule, when deciding whether to allow traffic
Stateful: Return traffic is allowed, regardless of the rules.	Stateless: Return traffic must be explicitly allowed by the rules.

Private, Public and Elastic IP

Private, Public, Elastic IP

1 **Private IP** – Each instance is assigned a Private IP. Stays for the life of the instance.

2 **Public IP**– Optional. Enabled when launching the instance. Required to send or receive traffic from the internet.

3 **Elastic IP**– Optional. Persistent / Static IP address assigned to your account / region. Required to send or receive traffic from the internet. You can reassign to any instance in the region.

VPC CIDR

VPC

```
10.0.0.0/16  (IPv4)
2600:1f16:e3f:7000::/56 (IPv6)
```

IPv4 and IPv6 Traffic are routed separately

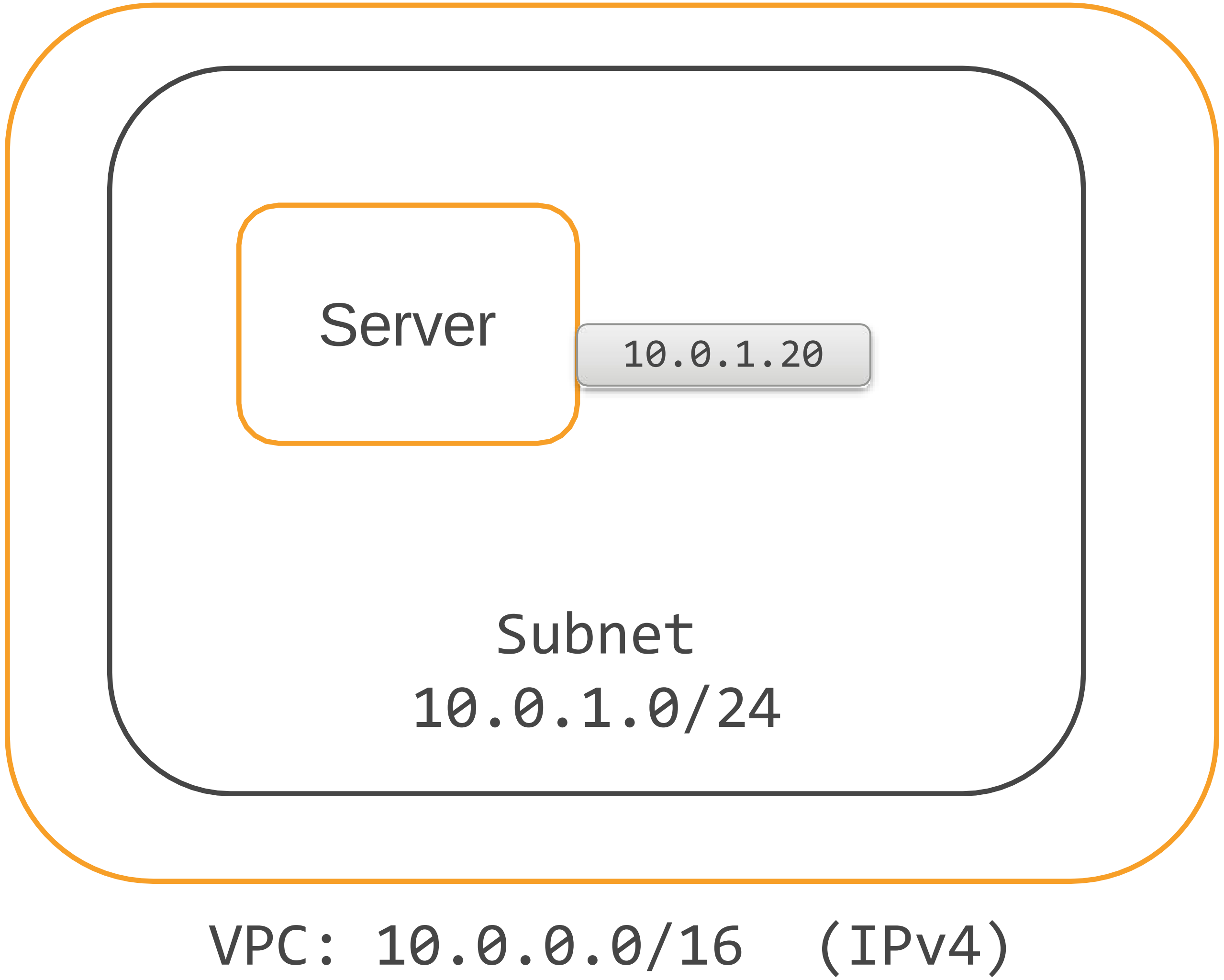
Configure

- Route table
- Security Group
- Network ACL

Private IPv4 CIDR

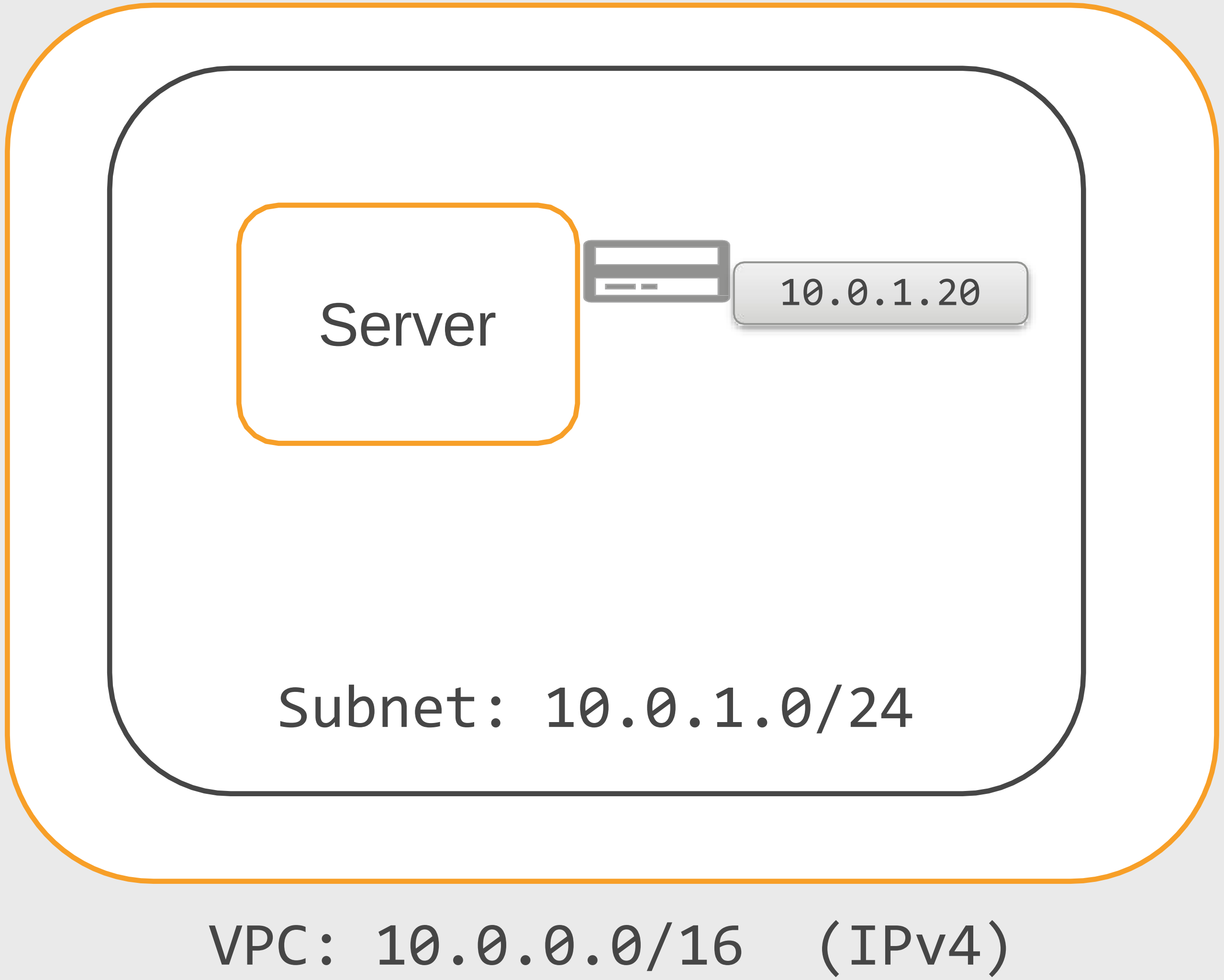
```
10.0.0.0 - 10.255.255.255   (10.0.0.0/8 prefix)
172.16.0.0 - 172.31.255.255   (172.16.0.0/12 prefix)
192.168.0.0 - 192.168.255.255 (192.168.0.0/16 prefix)
```

Private IP



Private IP automatically assigned from subnet CIDR block

Elastic Network Interface (ENI)

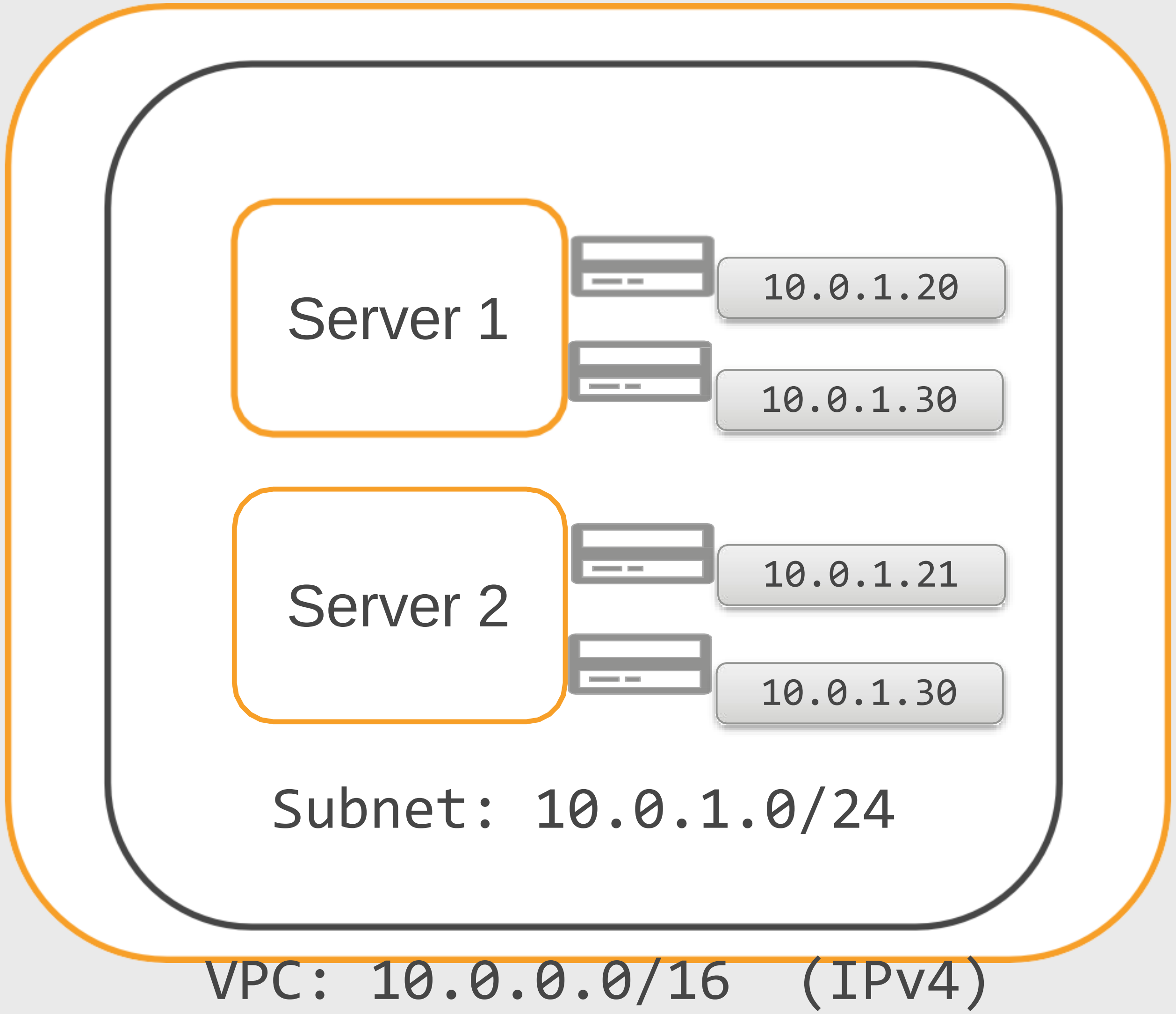


IP address is assigned to the primary network interface eth0

Private DNS Hostname

Primary network interface and private IP address stays with the instance until instance is terminated

Multiple Elastic Network Interfaces (ENI)

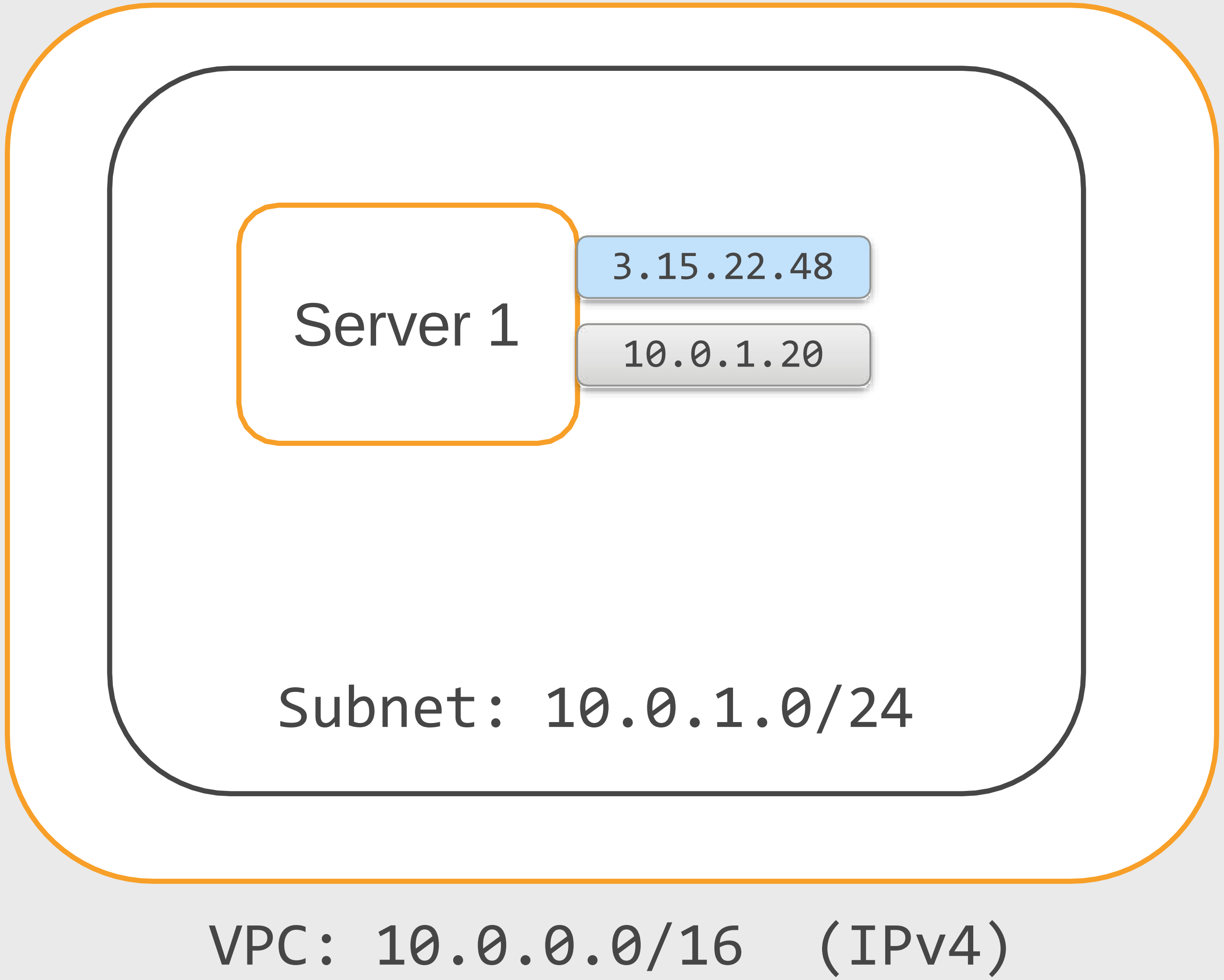


Multiple network interfaces can be attached to an instance

Secondary ENI can be detached and attached to another instance

Network traffic to that IP address is redirected to the new instance

Public IP



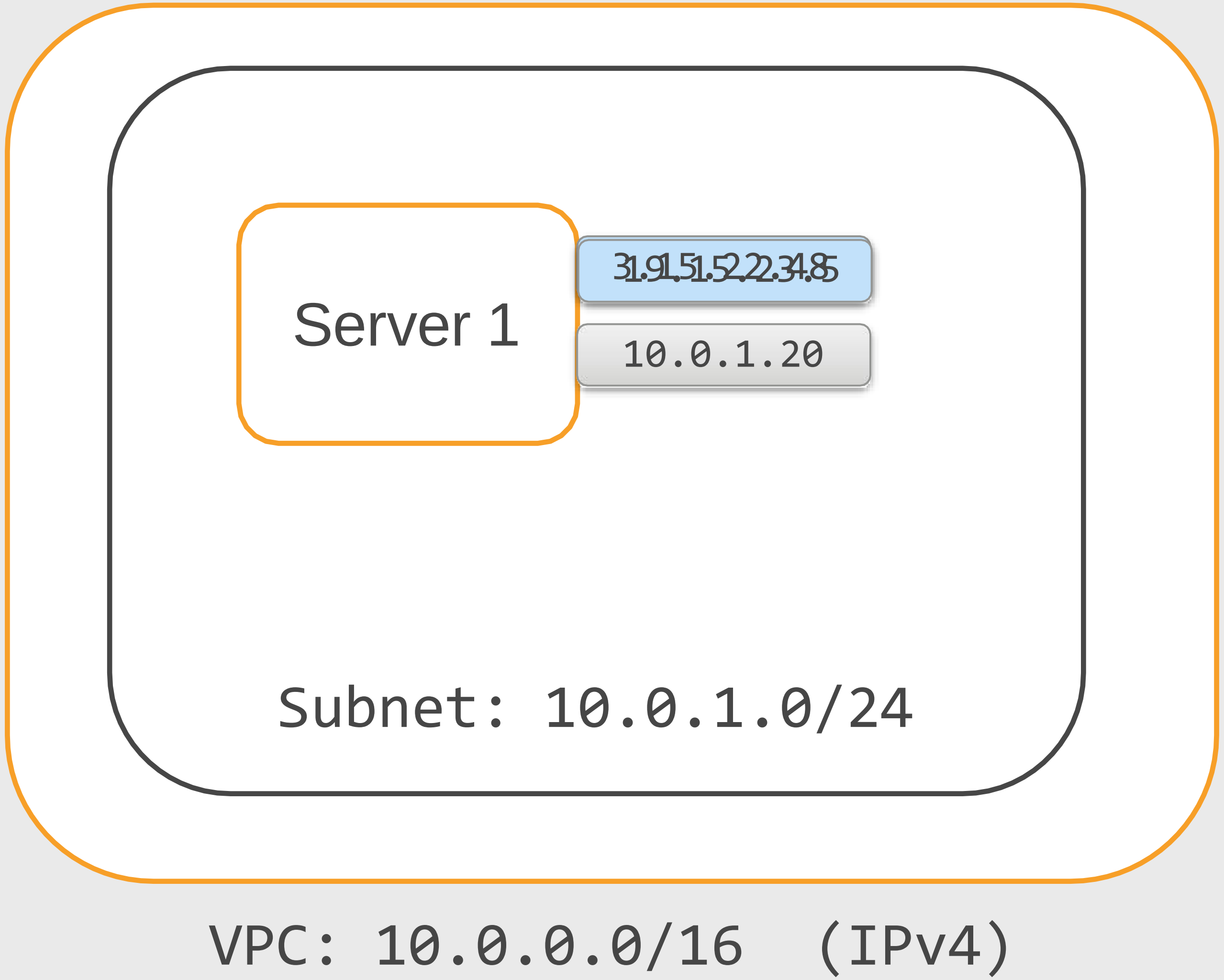
Public IP required to send or receive request from the internet

Public IP Assignment:

- Specify at the time of launching the instance
- Subnet setting to auto-assign public IP

Assigned from Amazon's Public IP pool

Public IP – Instance Start/Stop/Terminate



Stop or Terminate instance

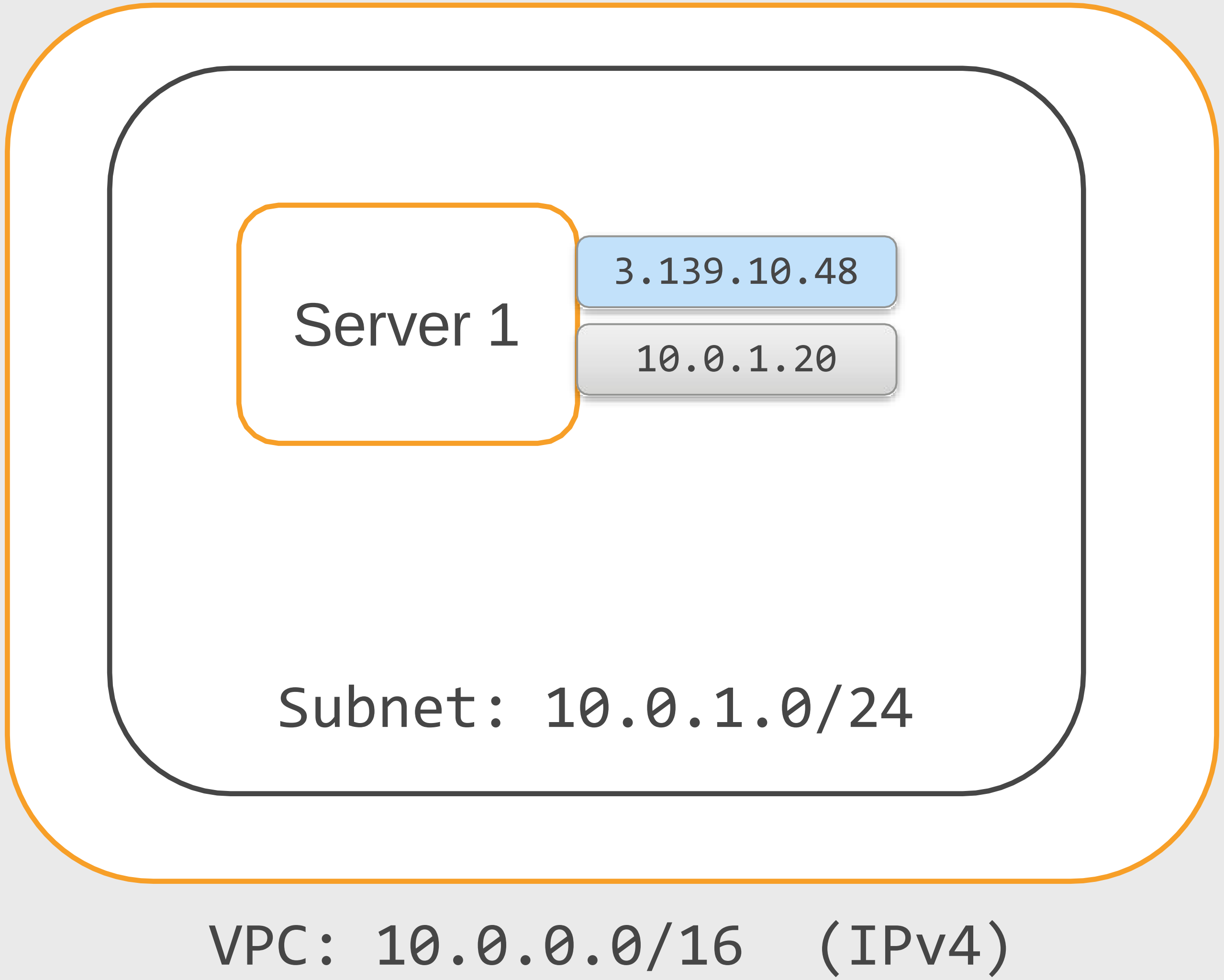
- Public IP is released back to pool

Restart a stopped instance

- New Public IP is assigned

Public IP will change if you stop and restart an instance

Elastic IP



Elastic IP is static-public IP address

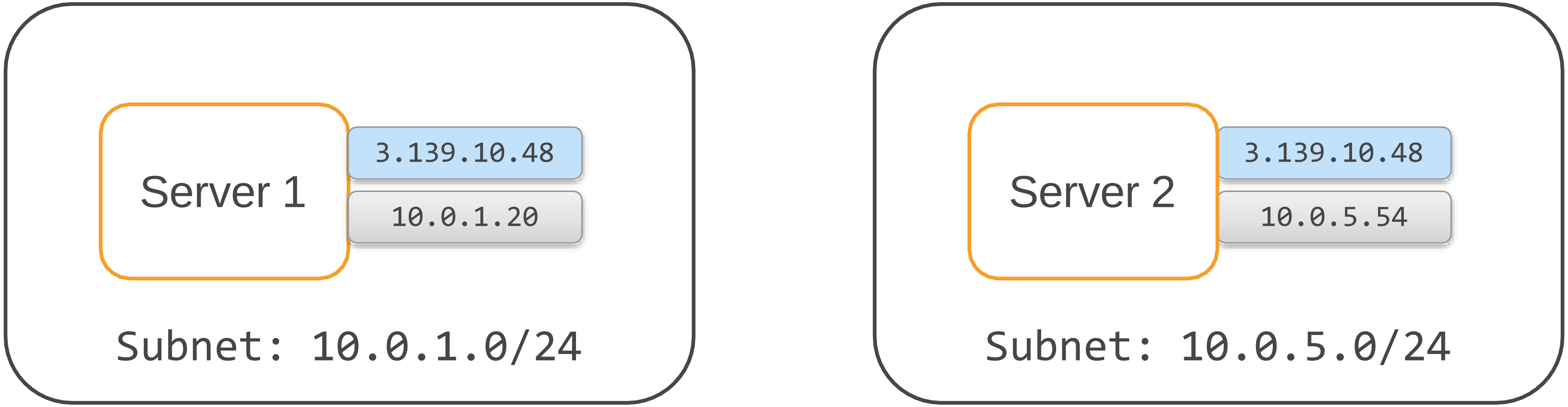
An Elastic IP address comes from Amazon's pool of IPv4 addresses, or from a custom IPv4 address pool that you have brought to your AWS account.

Assign to any instance

Stays attached to stopped instance

Limit of 5 Elastic IP per account per region

Elastic IP – Move to a Different Instance



- Detach and attach to a different instance in the same region in your account
- Redirect traffic to the new instance.
  - By using an Elastic IP address, you can mask the failure of an instance or software by rapidly remapping the address to another instance in your account.
- Elastic IP remains allocated to your account until you release it

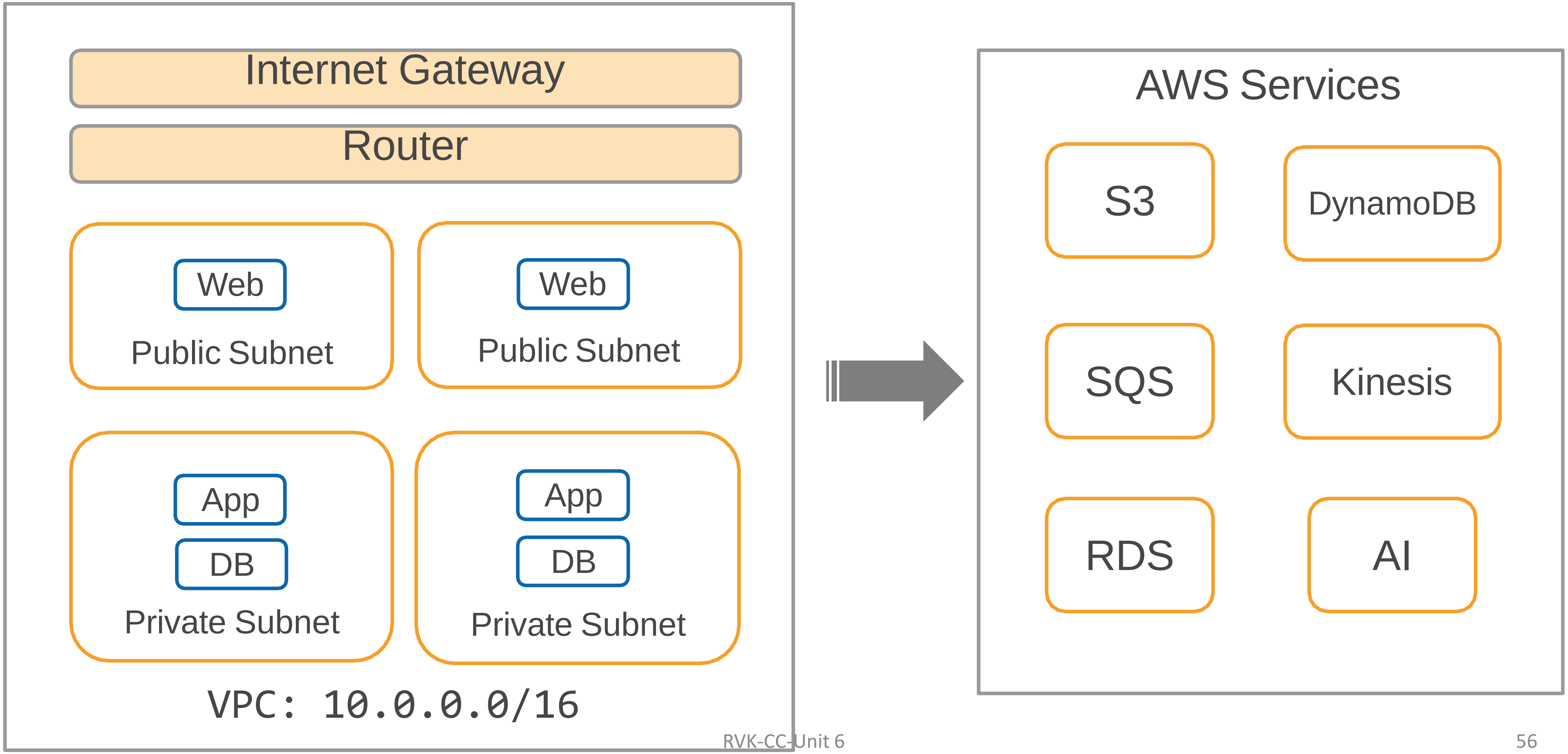
Integrating with other AWS Services

Internet

Gateway Endpoint

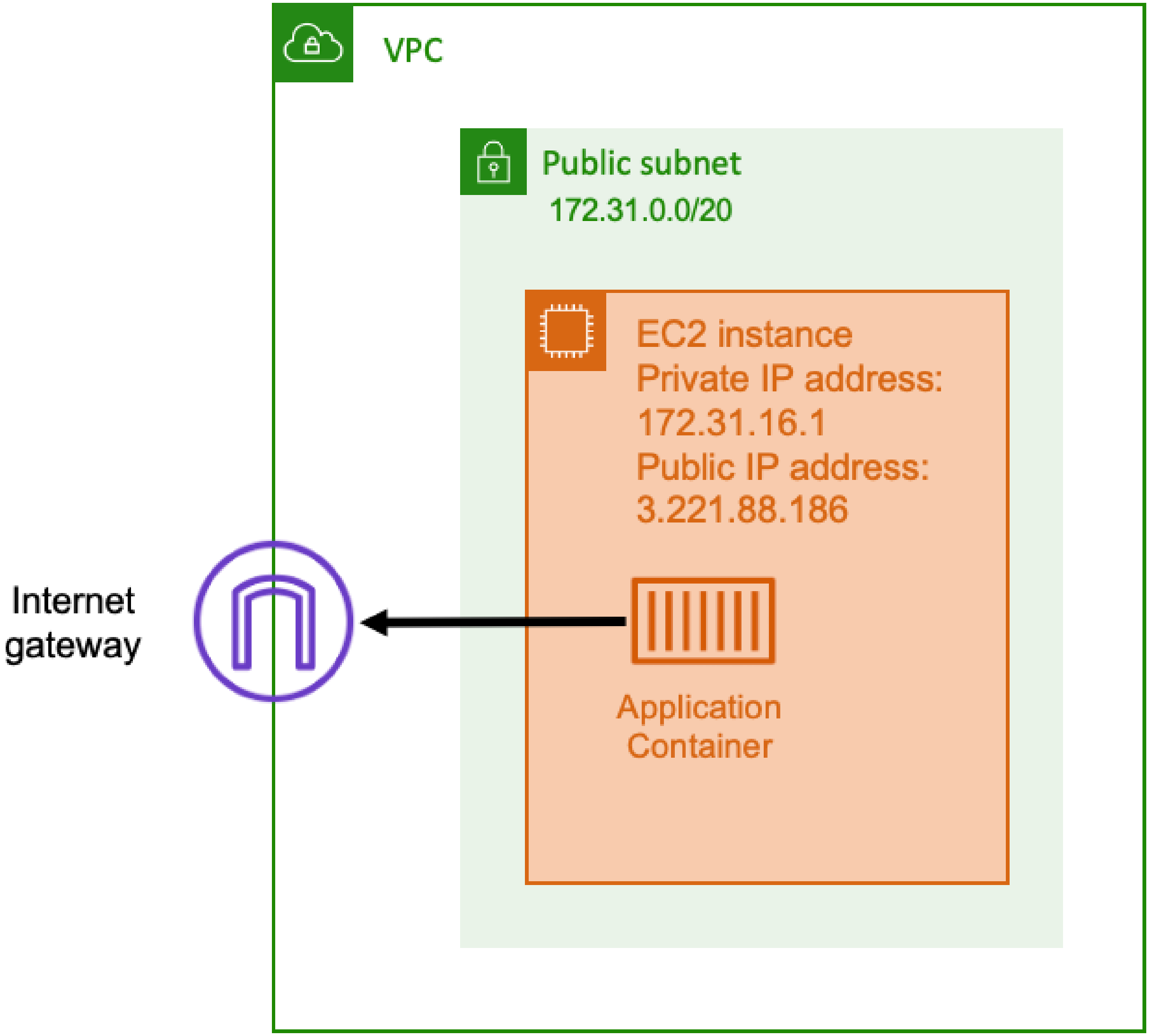
Interface Endpoint

How to integrate VPC with other AWS services?

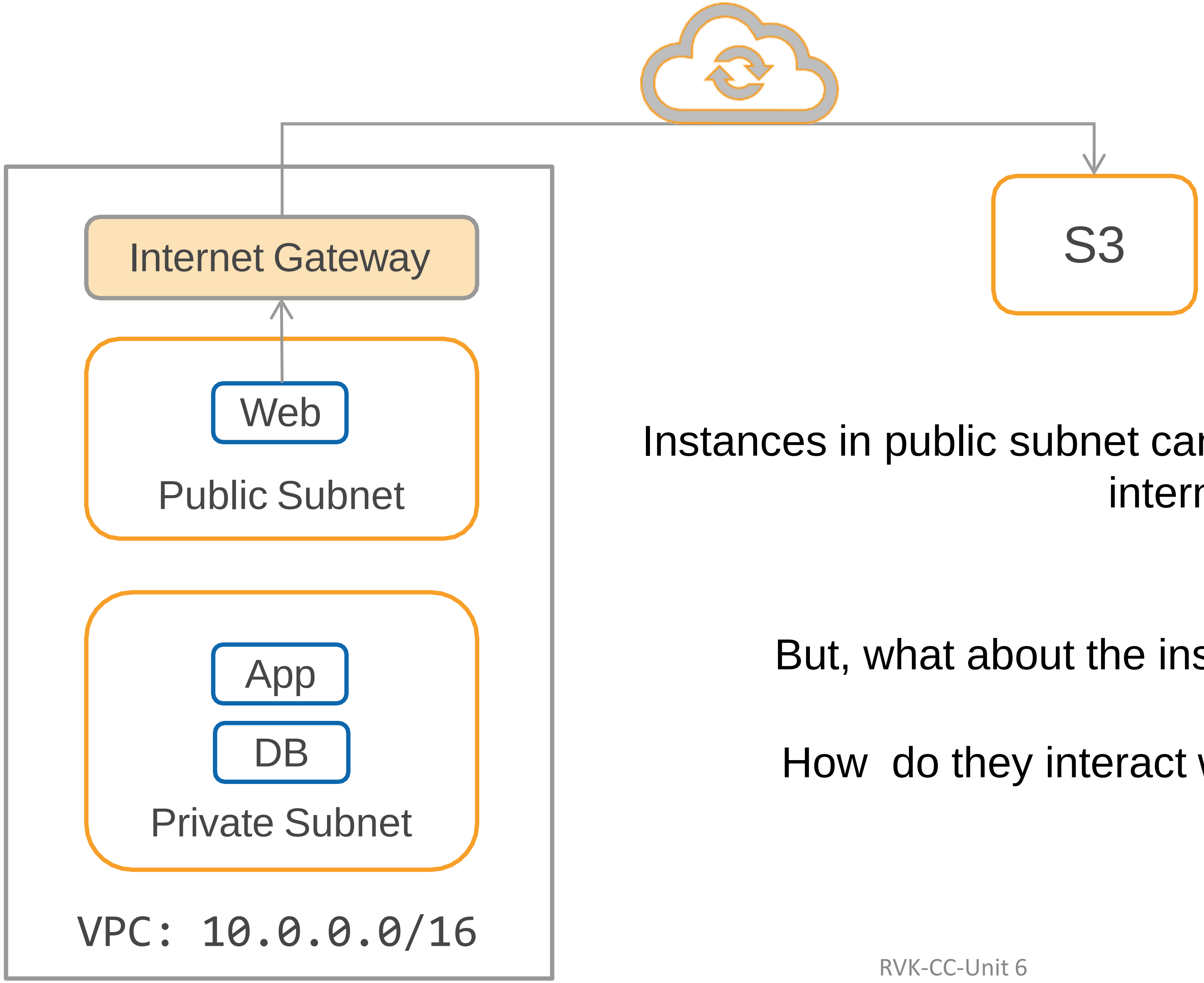


Using a Public Subnet and Internet Gateway

- This network architecture facilitates direct communication between the host that runs your application and other hosts on the internet.
- The communication is bi-directional. This means that not only can you establish an outbound connection to any other host on the internet, but other hosts on the internet might also attempt to connect to your host.
- Therefore, you should pay close attention to your security group and firewall rules.



Using a Public Subnet and Internet Gateway (cont..)



Instances in public subnet can make outbound calls to the internet.

But, what about the instances in private subnet?

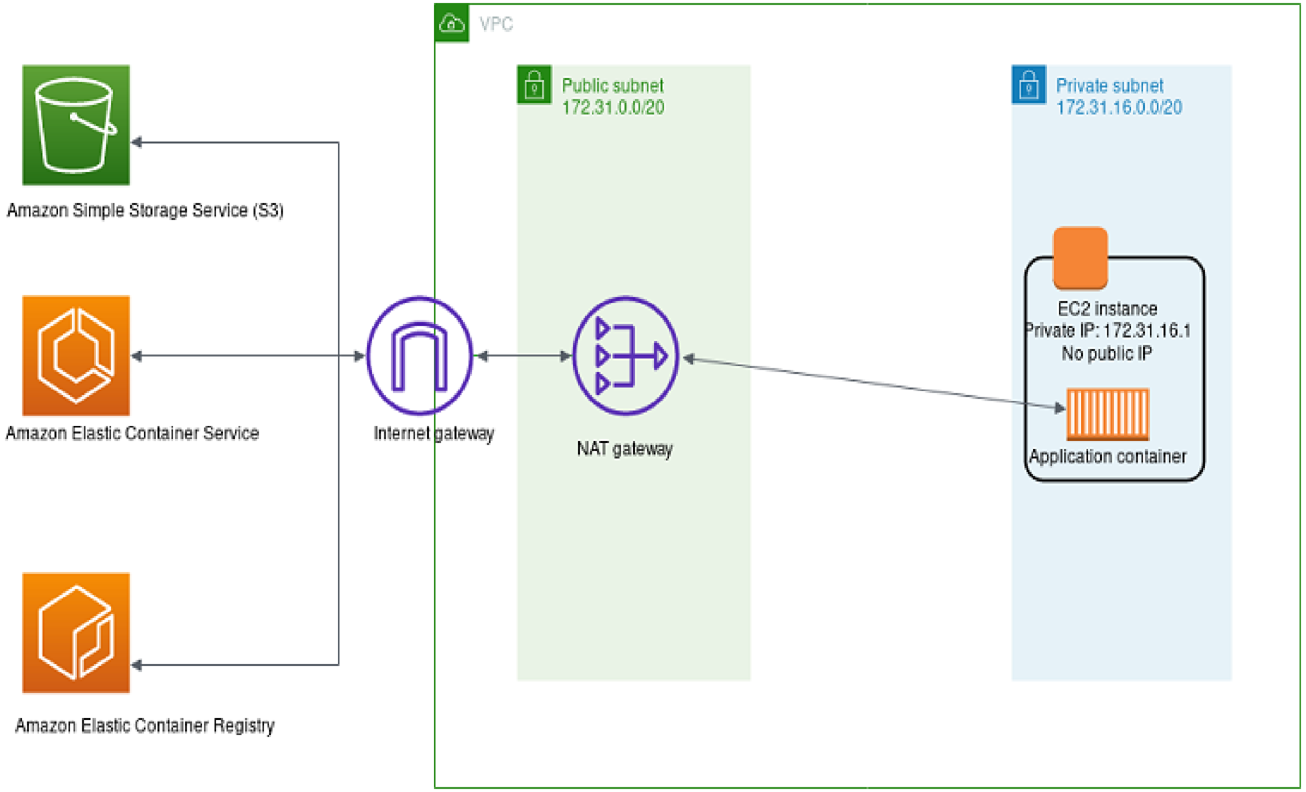
How do they interact with other AWS services?

Using a Private Subnet and NAT Gateway

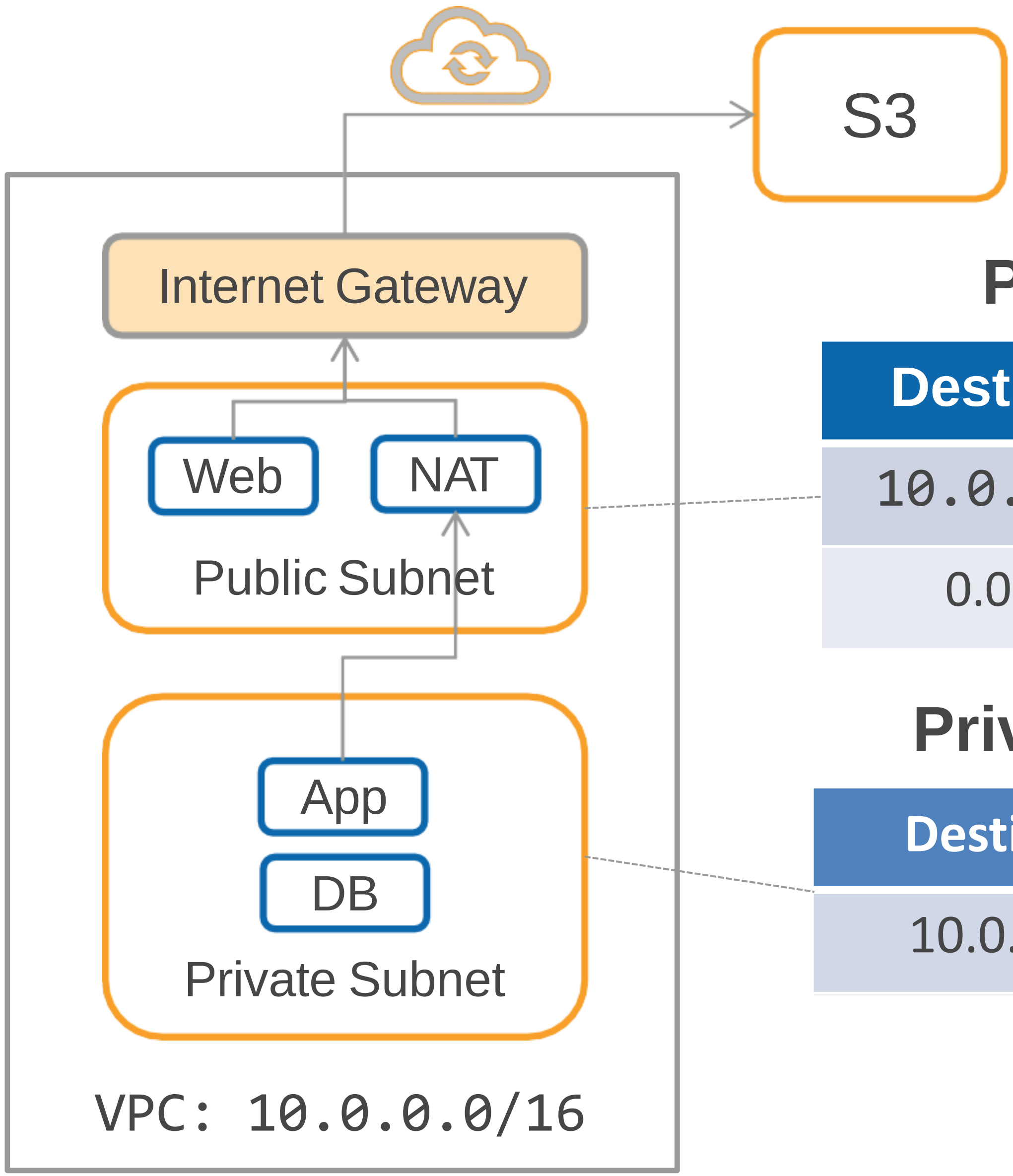
- Using a Network Address Translation (NAT) gateway is the easiest way to ensure that your Amazon Elastic Container Services (ECS) tasks can access other AWS services.

- With a private subnet, you can use a NAT gateway to enable a host inside a private subnet to connect to the internet.

- Drawbacks:
  - You can't limit what destinations the NAT gateway can communicate with.
  - NAT gateways charge for every GB of data that passes through.



Using a Private Subnet and NAT Gateway (cont..)



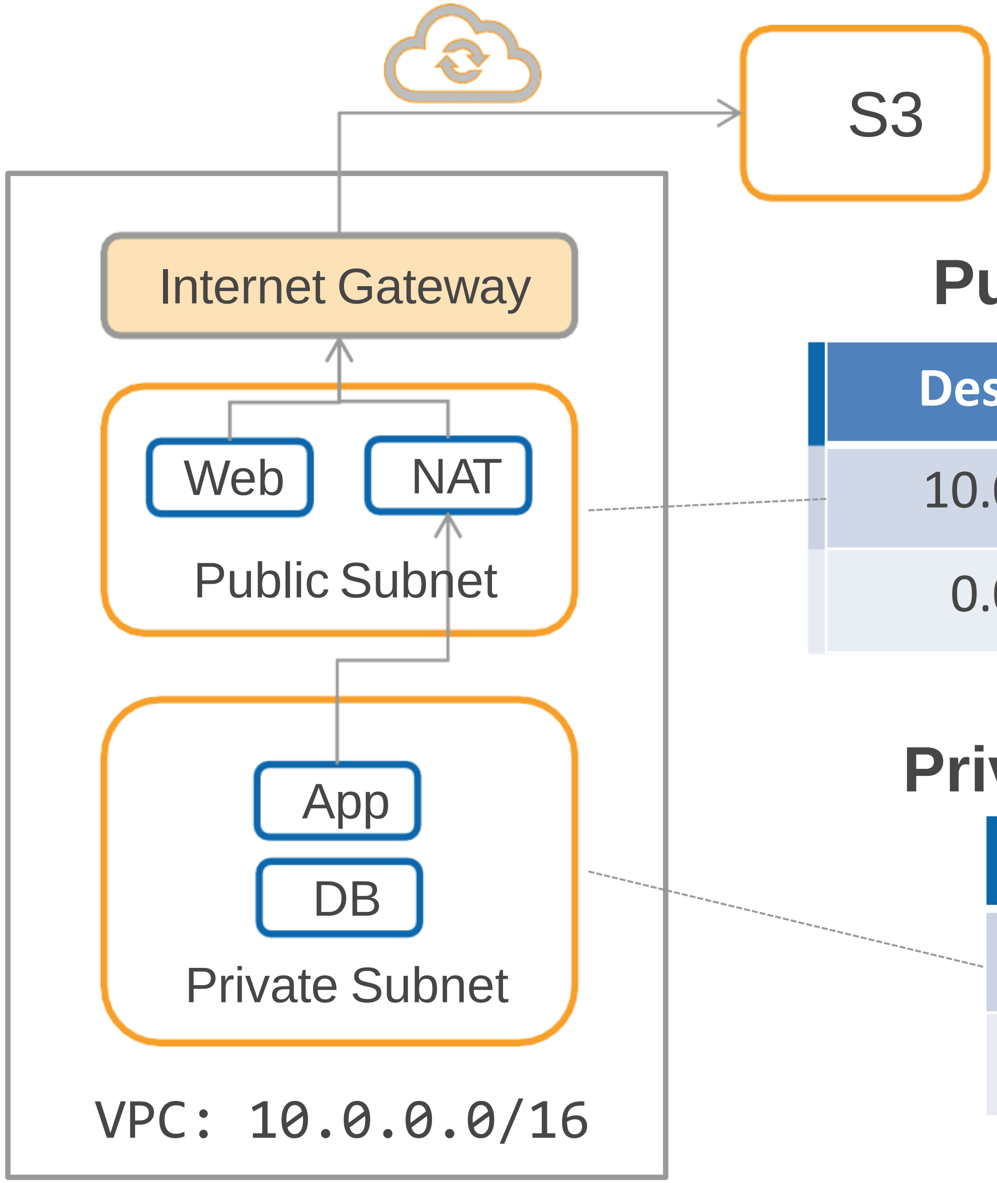
Public Route Table

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	IGW-id

Private Route Table

Destination	Target
10.0.0.0/16	local

Using a Private Subnet and NAT Gateway (cont..)



Public Route Table

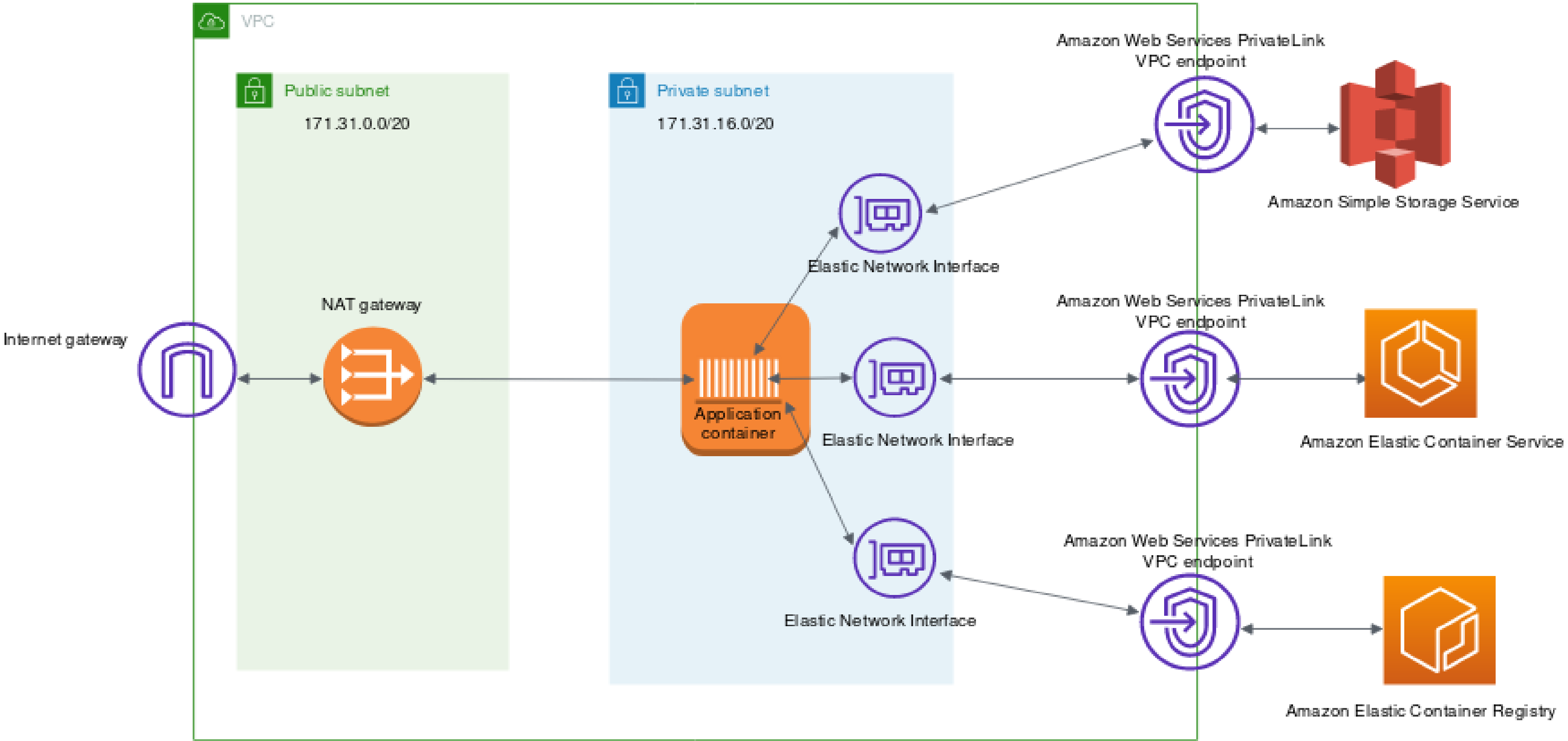
Destination	Target
10.0.0.0/16	local
0.0.0.0/0	IGW-id

Private Route Table (include path to NAT)

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	NAT-GW-id

Why Not Talk Directly to AWS Services?

- AWS PrivateLink provides private connectivity between VPCs, AWS services, and your on-premises networks without exposing your traffic to the public internet.
- A VPC endpoint enables private connections between your VPC and supported AWS services and VPC endpoint services. Traffic between your VPC and the other service doesn't leave the Amazon network.
- A VPC endpoint doesn't require public IP addresses, an internet gateway, virtual private gateway, NAT device, VPN connection, or AWS Direct Connect connection.


VPC
Public subnet
171.31.0.0/20
Private subnet
171.31.16.0/20
Internet gateway
NAT gateway
Application container
Elastic Network Interface
Elastic Network Interface
Elastic Network Interface
Amazon Web Services PrivateLink VPC endpoint
Amazon Web Services PrivateLink VPC endpoint
Amazon Web Services PrivateLink VPC endpoint
Amazon Simple Storage Service
Amazon Elastic Container Service
Amazon Elastic Container Registry

Amazon VPC Endpoint Types

- Interface – It creates an interface endpoint to send traffic to endpoint services that use a Network Load Balancer to distribute traffic. Traffic destined for the endpoint service is resolved using DNS. All newer services use interface endpoint.

- GatewayLoadBalancer - It creates a Gateway Load Balancer endpoint to send traffic to a fleet of virtual appliances using private IP addresses. You route traffic from your VPC to the Gateway Load Balancer endpoint using route tables. The Gateway Load Balancer distributes traffic to the virtual appliances and can scale with demand.

- Gateway - It creates a gateway endpoint to send traffic to Amazon S3 or DynamoDB using private IP addresses. You route traffic from your VPC to the gateway endpoint using route tables. Gateway endpoints do not enable AWS PrivateLink.

Amazon VPC Endpoint Types

Gateway Endpoint - S3, DynamoDB

Interface Endpoint – All newer services use interface endpoint

Gateway Endpoint

1 With endpoint, you can access S3 and DynamoDB using Private IP address

2 Endpoint is regional - Used for S3 and DynamoDB in the same region

3 For other regions, use internet gateway + NAT

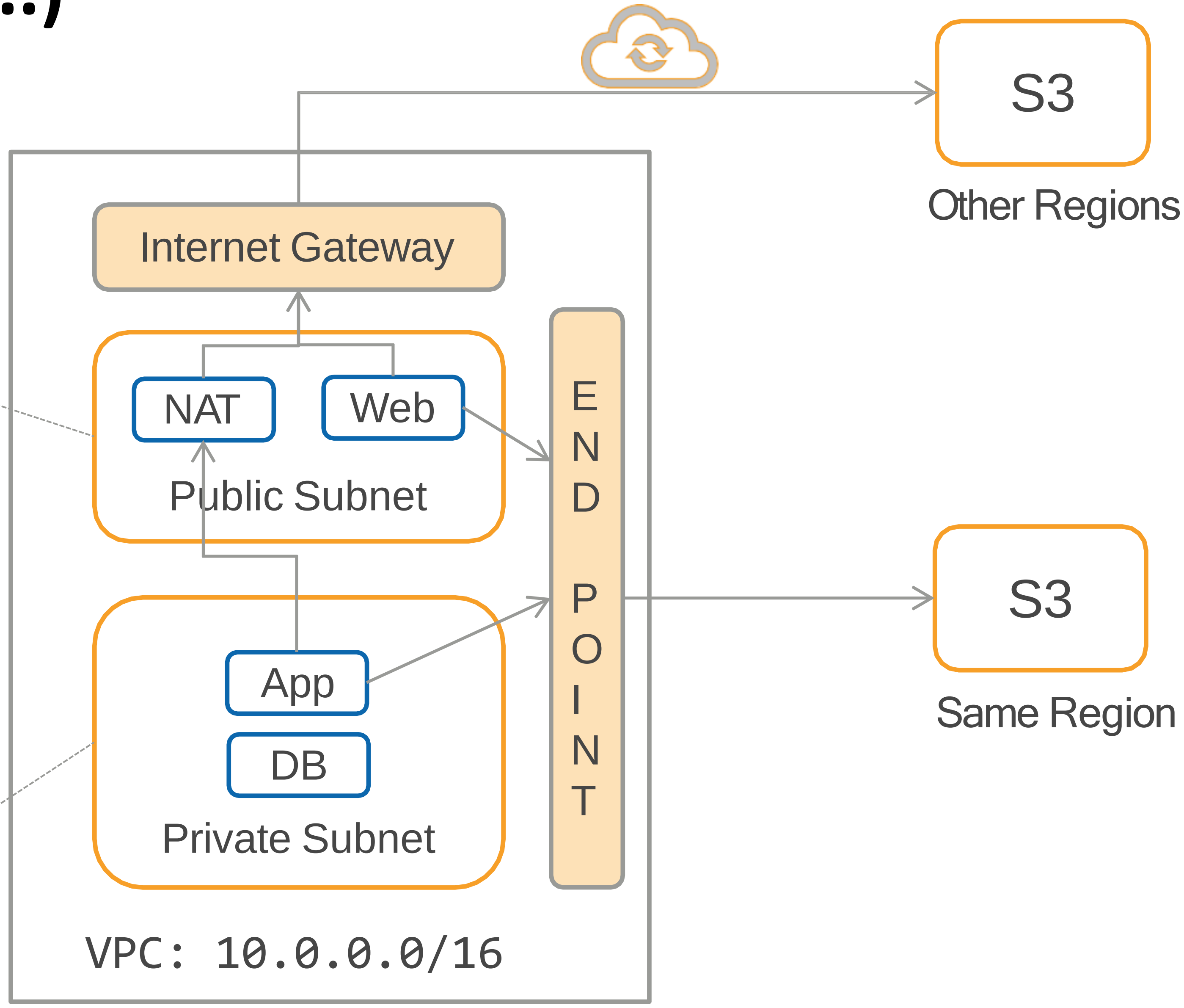
Gateway Endpoint (cont..)

Public Route Table

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	IGW-id
Pl-id	VPCE-id

Private Route Table

Destination	Target
10.0.0.0/16	Local
0.0.0.0/0	NAT-GW-id
Pl-id	VPCE-id



Interface Endpoint

1. Interface endpoints are also known as PrivateLink

2. Privately interact with many AWS services (same-region)

3. Interface endpoint creates a network interface with private IP (easy to remember)

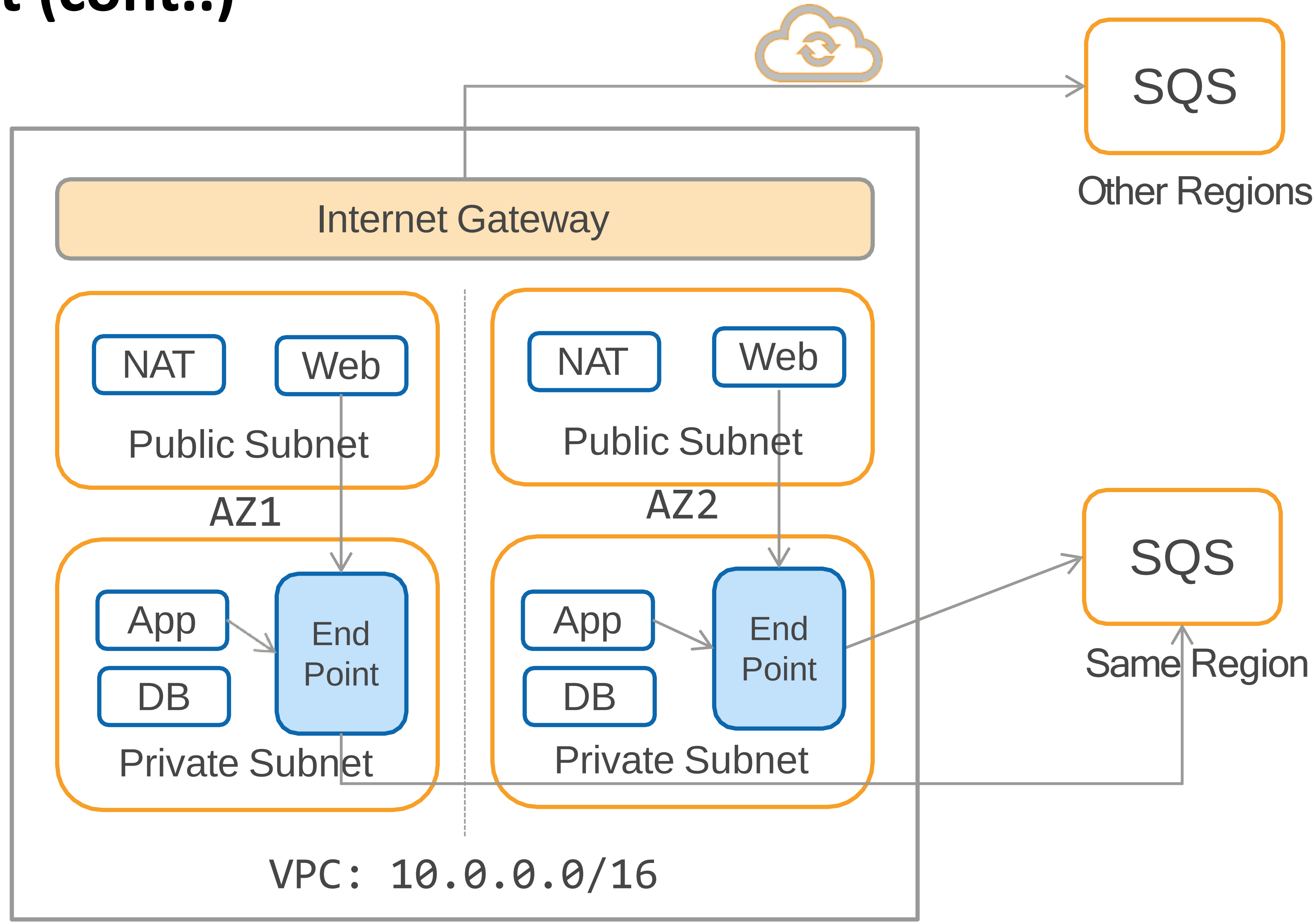
4. Flexibility to expose your service to other customers

Interface Endpoint (cont..)

No need to update route table to use endpoint – Private IP

```
Queue Name:
sqs.us-east-1.amazonaws.com
```

With Private DNS HostName option, Service DNS name is automatically mapped to Endpoint IP address



Summary – Integrating with AWS services

- **Internet**

Useful for both cross-region, same-region access Public instances – Internet Gateway.

Private instances – NAT + Internet Gateway.

- **Gateway Endpoint**

Private connectivity to S3, DynamoDB in the same region.

For other regions, use the internet.

- **Interface Endpoint**

Private connectivity to many AWS services in the same region.

For high availability, create an interface endpoint in each AZ.

For other regions, use the internet.

Google Cloud Network Architecture

For more details refer to: [Google Networking Architecture]

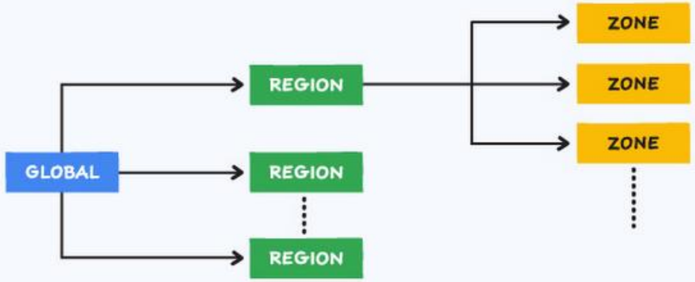
Introduction to Cloud Networking

#GCPSketchnote @PVERGADIA THECLOUDGIRL.DEV

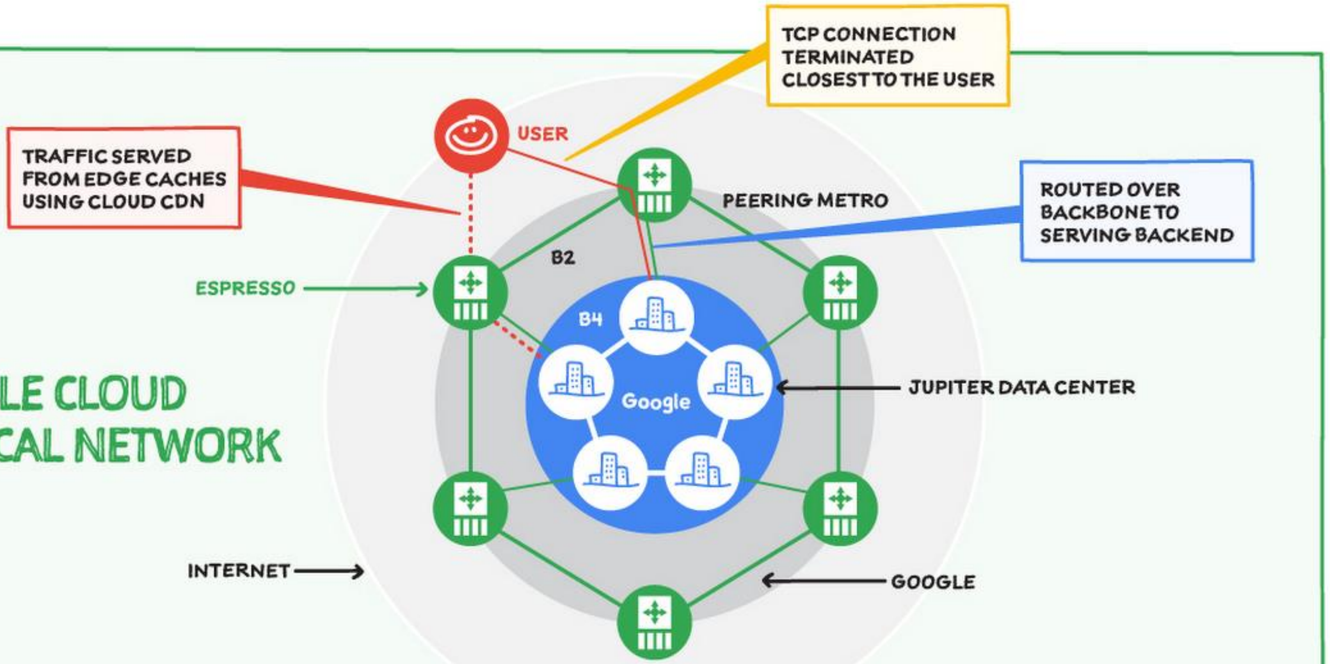
08.17.2021

GLOBAL INFRASTRUCTURE

27	REGIONS
82	ZONES
146	NETWORK EDGE LOCATIONS
AVAILABLE IN 200+	COUNTRIES & TERRITORIES
14	SUB-SEA CABLES
113	INTERCONNECT LOCATIONS



GOOGLE CLOUD PHYSICAL NETWORK



TCP CONNECTION TERMINATED CLOSEST TO THE USER

TRAFFIC SERVED FROM EDGE CACHES USING CLOUD CDN

ROUTED OVER BACKBONE TO SERVING BACKEND

Google Cloud NETWORKING SERVICES

CONNECT

- VPC
- CLOUD DNS
- CLOUD VPN
- CLOUD ROUTER
- DEDICATED INTERCONNECT
- PARTNER INTERCONNECT

SECURE

- CLOUD ARMOR
- FIREWALL RULES
- PACKET MIRRORING
- CLOUD IAP
- CLOUD NAT

SCALE

- CLOUD LOAD BALANCER
- CLOUD CDN

OPTIMIZE

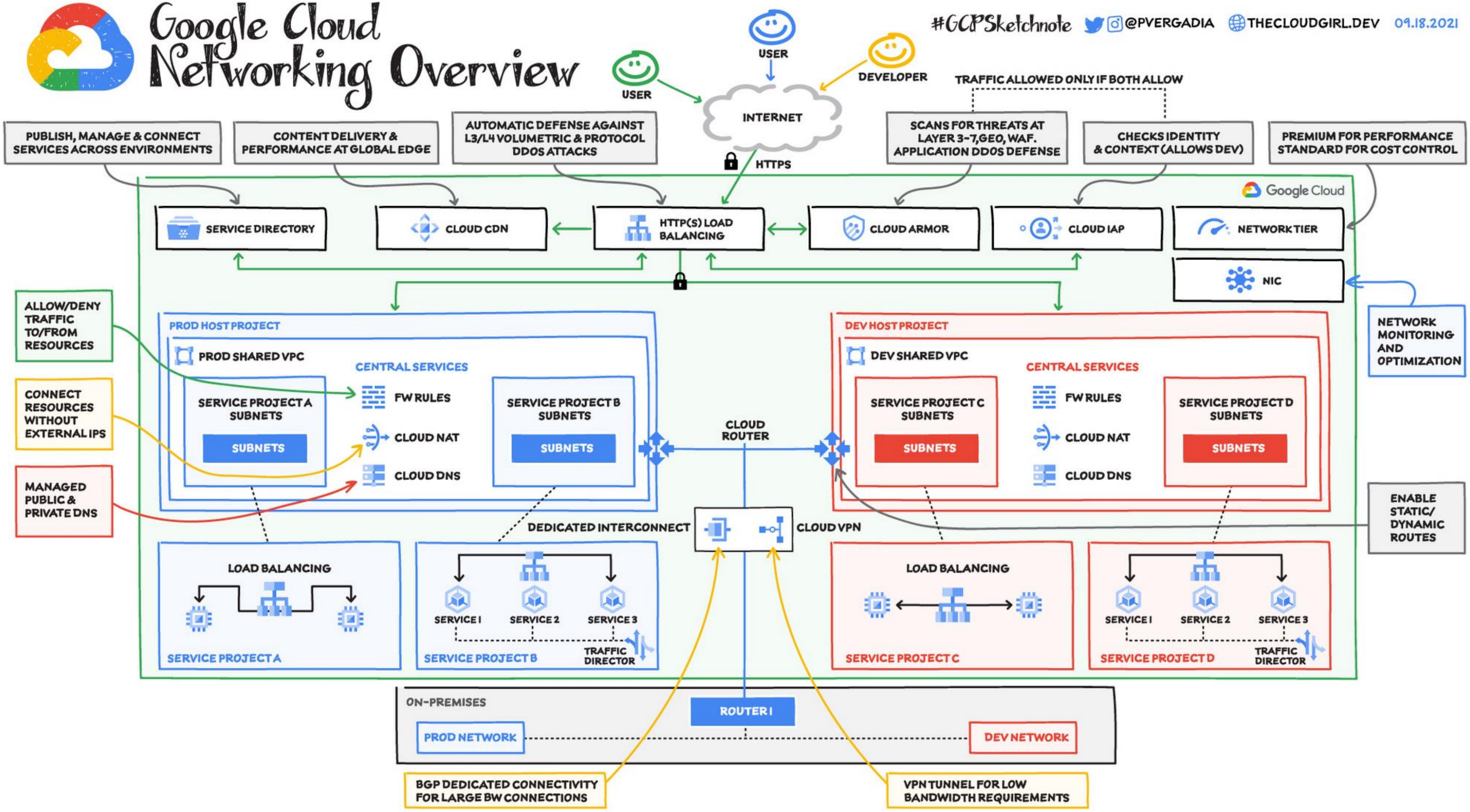
- PREMIUM TIER
- STANDARD TIER
- NETWORK INTELLIGENCE CENTER

MODERNIZE

- GKE NETWORKING (+ ON-PREM) (IN ANTHOS)
- TRAFFIC DIRECTOR
- SERVICE DIRECTORY

Google Cloud Networking Overview

#GCPSketchnote @PVERGADIA THECLOUDGIRL.DEV 09.18.2021



USER

USER

DEVELOPER

INTERNET

HTTPS

TRAFFIC ALLOWED ONLY IF BOTH ALLOW

PUBLISH, MANAGE & CONNECT SERVICES ACROSS ENVIRONMENTS

CONTENT DELIVERY & PERFORMANCE AT GLOBAL EDGE

AUTOMATIC DEFENSE AGAINST L3/L4 VOLUMETRIC & PROTOCOL DDOS ATTACKS

SCANS FOR THREATS AT LAYER 3-7, GEO, WAF. APPLICATION DDOS DEFENSE

CHECKS IDENTITY & CONTEXT (ALLOWS DEV)

PREMIUM FOR PERFORMANCE STANDARD FOR COST CONTROL

Google Cloud

SERVICE DIRECTORY

CLOUD CDN

HTTP(S) LOAD BALANCING

CLOUD ARMOR

CLOUD IAP

NETWORK TIER

NIC

NETWORK MONITORING AND OPTIMIZATION

ALLOW/DENY TRAFFIC TO/FROM RESOURCES

CONNECT RESOURCES WITHOUT EXTERNAL IPS

MANAGED PUBLIC & PRIVATE DNS

PROD HOST PROJECT

PROD SHARED VPC

CENTRAL SERVICES

SERVICE PROJECT A SUBNETS

SUBNETS

FW RULES

CLOUD NAT

CLOUD DNS

SERVICE PROJECT B SUBNETS

SUBNETS

CLOUD ROUTER

DEV HOST PROJECT

DEV SHARED VPC

CENTRAL SERVICES

SERVICE PROJECT C SUBNETS

SUBNETS

FW RULES

CLOUD NAT

CLOUD DNS

SERVICE PROJECT D SUBNETS

SUBNETS

ENABLE STATIC/ DYNAMIC ROUTES

DEDICATED INTERCONNECT

CLOUD VPN

LOAD BALANCING

SERVICE PROJECT A

SERVICE 1

SERVICE 2

SERVICE 3

TRAFFIC DIRECTOR

SERVICE PROJECT B

LOAD BALANCING

SERVICE PROJECT C

SERVICE 1

SERVICE 2

SERVICE 3

TRAFFIC DIRECTOR

SERVICE PROJECT D

ON-PREMISES

ROUTER 1

PROD NETWORK

DEV NETWORK

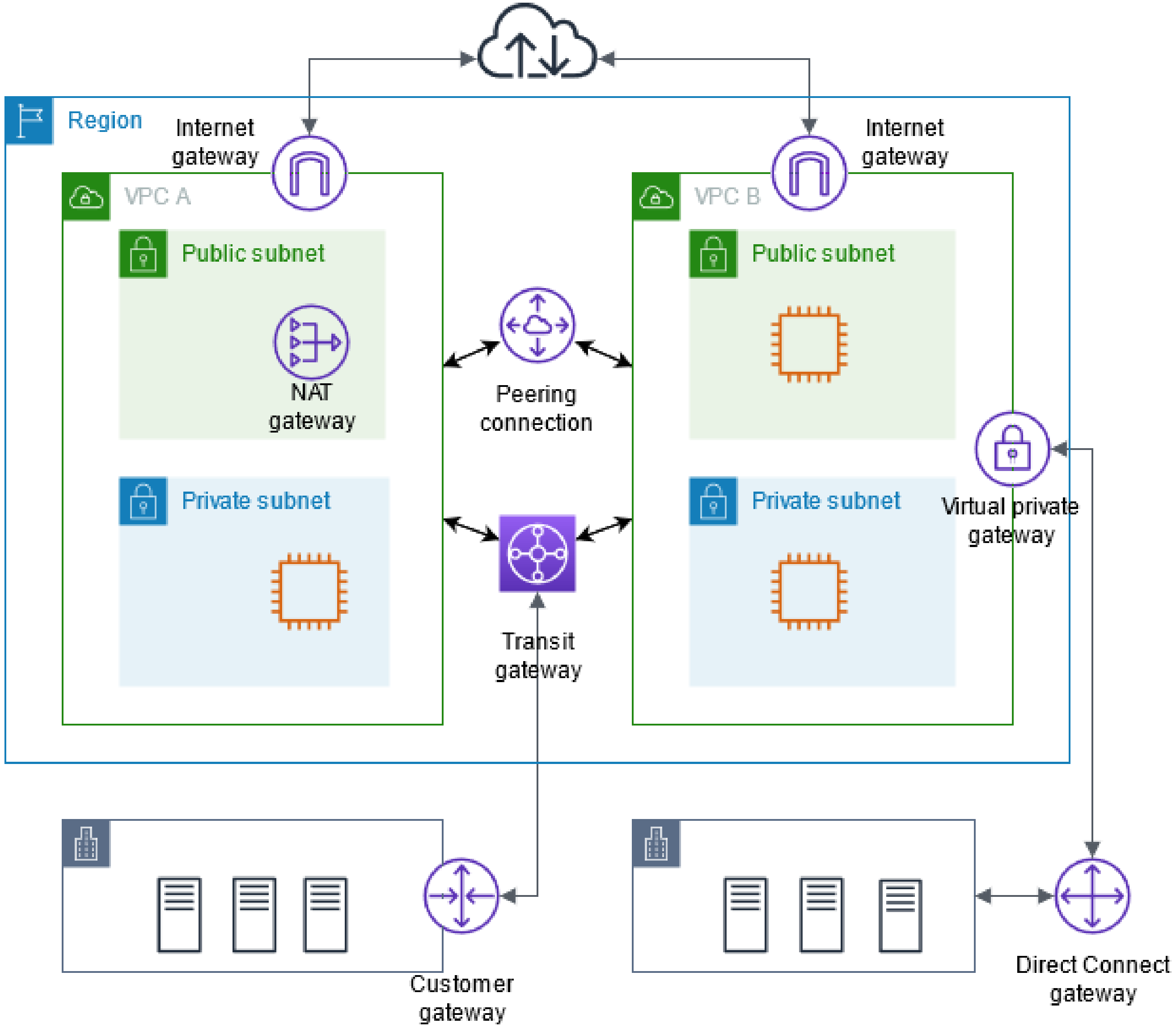
BGP DEDICATED CONNECTIVITY FOR LARGE BW CONNECTIONS

VPN TUNNEL FOR LOW BANDWIDTH REQUIREMENTS

Hybrid Cloud

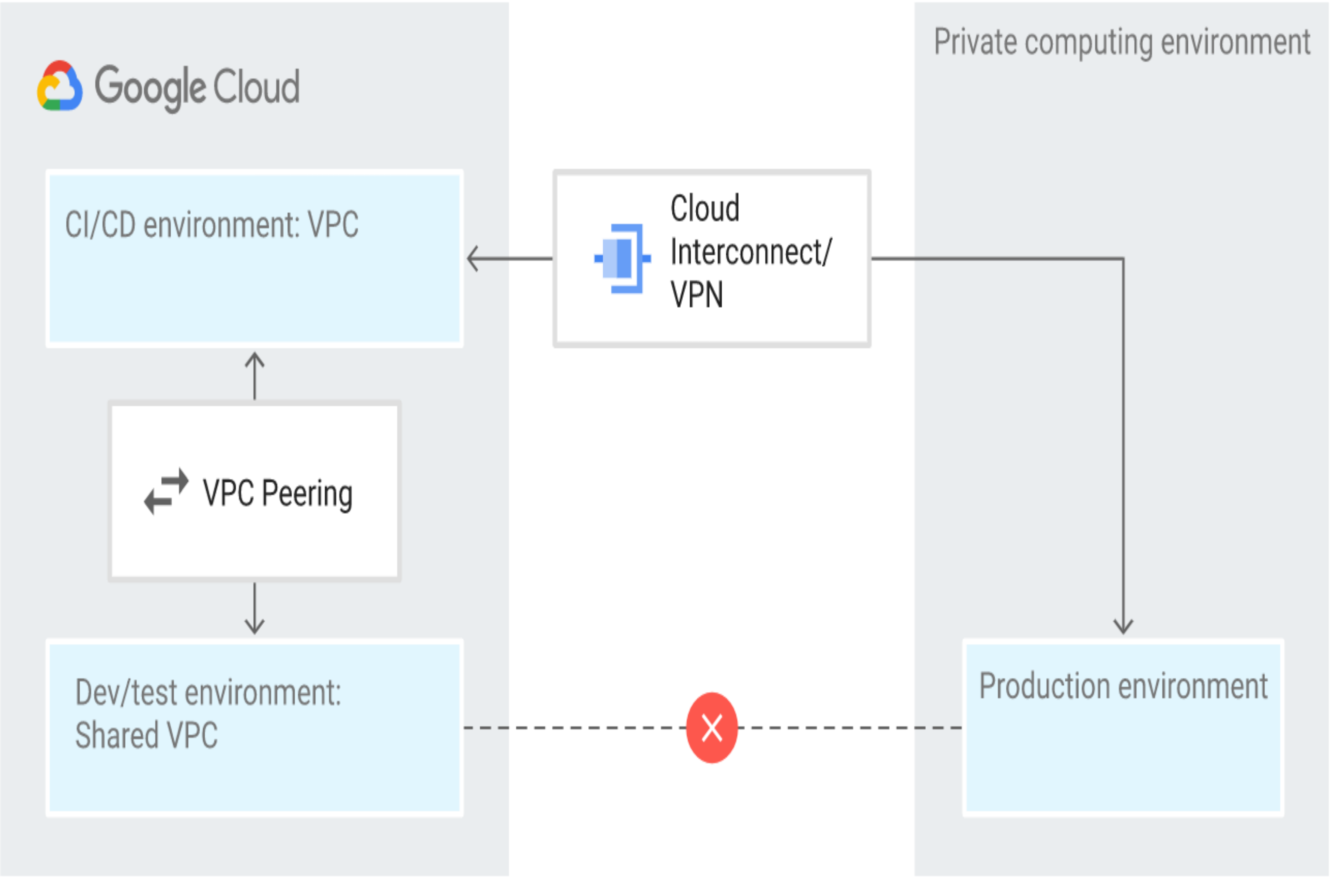
Example: Amazon Hybrid Cloud

- VPC A is connected to the internet through an internet gateway.
- The EC2 instance in the private subnet of VPC A can connect to the internet using the NAT gateway in the public subnet of VPC A.
- VPC B is connected to the internet through an internet gateway.
- The EC2 instance in the public subnet of VPC B can connect to the internet using the internet gateway.
- VPC A and VPC B are connected to each other through a VPC peering connection and a transit gateway.
- The transit gateway has a VPN attachment to a data center.
- VPC B has a AWS Direct Connect connection to a data center.



Example: Google Hybrid Cloud

- Two separate virtual private clouds (VPCs)—1)Shared VPC for development and testing workloads, and 2)an additional VPC for all CI/CD and administrative tooling.
- The two VPCs are peered, allowing cross-VPC communication that uses internal IP addresses.
- The peering allows CI/CD and administrative systems to deploy and manage development and testing workloads.
- Additionally, you connect the CI/CD VPC to the network running the production workloads in the private computing environment.
- You establish this connection by using either Cloud Interconnect or Cloud VPN. This connection allows you to deploy and manage production workloads.



Load Balancing

Load Balancing

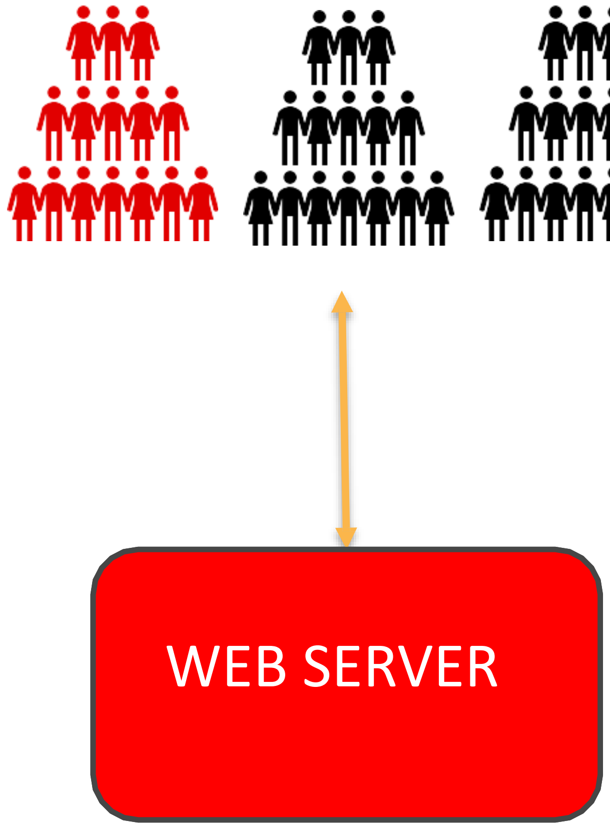
- Load Balancing automatically distributes incoming application traffic across multiple targets and virtual appliances in one or more Availability Zones (AZs). It makes decisions on where to send incoming requests based on algorithms for optimizing network traffic.

- The more popular load balancer algorithms include:
  - **Round robin:** Under this setup, the load balancer makes decisions based on the sequence of servers in your network. Thus, Server 1 is the first to handle a request, then Server 2 and onward. As it does not account for server load, this algorithm can lead to some servers getting overloaded with requests. A variation of this is the weighted round-robin, which assigns weights to servers based on their capabilities.
  - **Least connection:** With this setup, incoming requests are sent to servers with the least number of active connections. Thus, it helps avoid the potential problem of server overload.
  - **Least bandwidth consumption:** This setup measures the amount of traffic transmitted to and from servers, with the server having the least bandwidth consumption eventually getting the request. This is like the least packets method, which bases the load balancer's forwarding decision on the number of packets the server transmits over the network.
  - **Least response time:** Under this setup, the load balancer sends monitoring requests to servers to determine how fast they can serve a request before forwarding the actual request to the server that can handle the request faster.
  - **Hashing:** With this setup, the load balancer relies on hash data from incoming network packets, including the Internet Protocol (IP) addresses of the source and destination. This can be complicated to set up, making it more difficult than the other algorithms discussed in this section.
  - **Custom load:** This setup involves querying server loads based on data provided by SNMP (Simple Network Management Protocol), including memory usage, CPU usage, and response time, with the load balancer relying on the data to make its routing decision.

Issues with Load Balancers

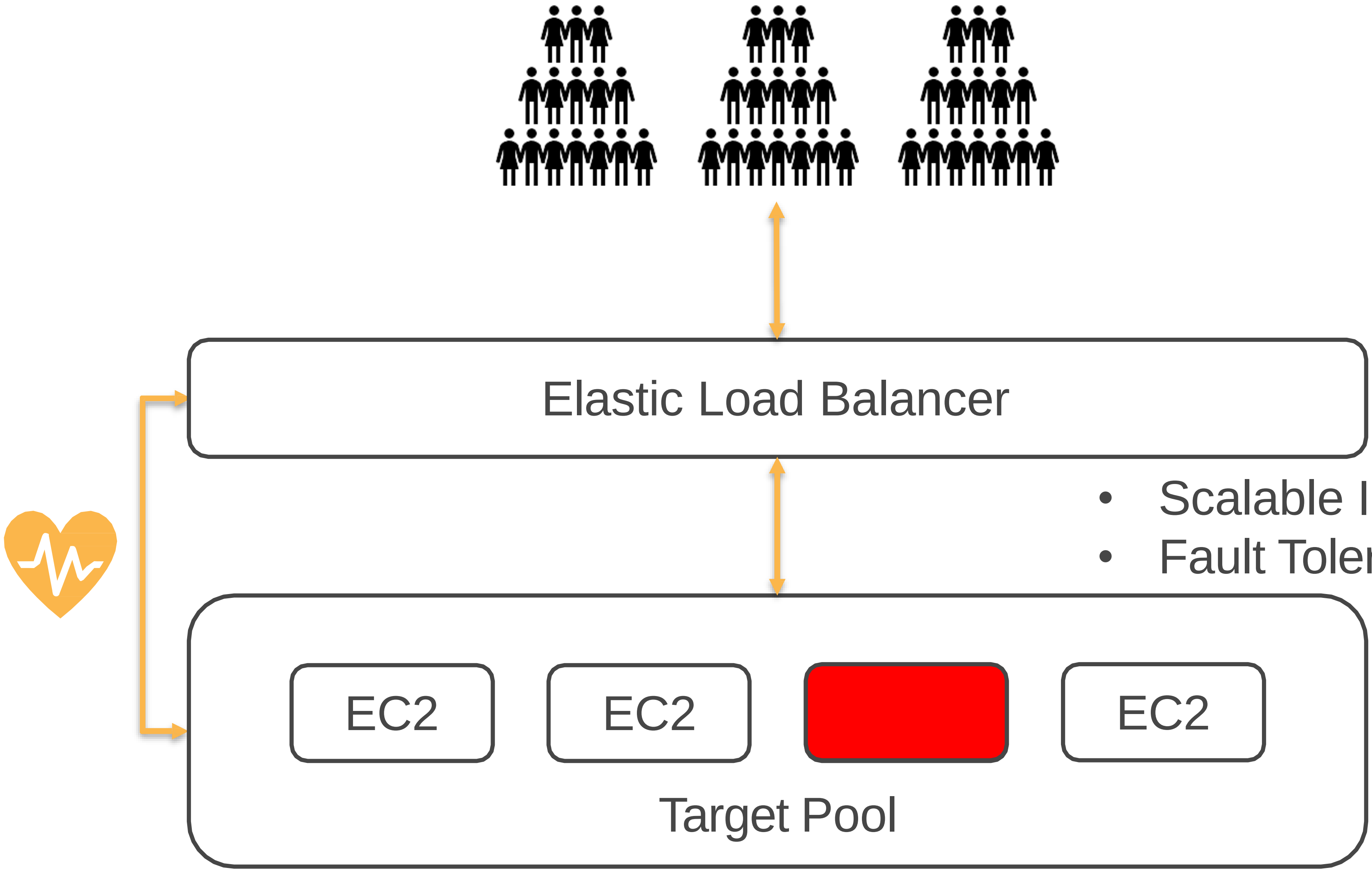
- It can make your network more complex than it already is.
- If your load balancer goes down, it can take your whole network down as well.
  - A failover mechanism for your load balancing infrastructure must be implemented to prevent this from happening.
  - One way to achieve this is through redundant routers that can switch traffic from one load balancer to another in case of failure.
  - With such a mechanism in place, when the primary load balancer fails, a backup load balancer takes over its functions until the primary load balancer goes back online.
- The failover requirement above means that load balancers can lead to higher operational costs for your network.
- Load balancer misconfiguration can also bring about network problems.
  - For example, it can happen that load balancers may detect failure incorrectly in healthy servers and reroute traffic to other servers needlessly. This can be due to several factors, including too frequent health checks and too short time-out periods for when a response is expected from your servers during health checks.
  - Adjustments to the frequency of health checks and making timeout periods longer can help resolve this issue.

Elastic Load Balancing Motivation



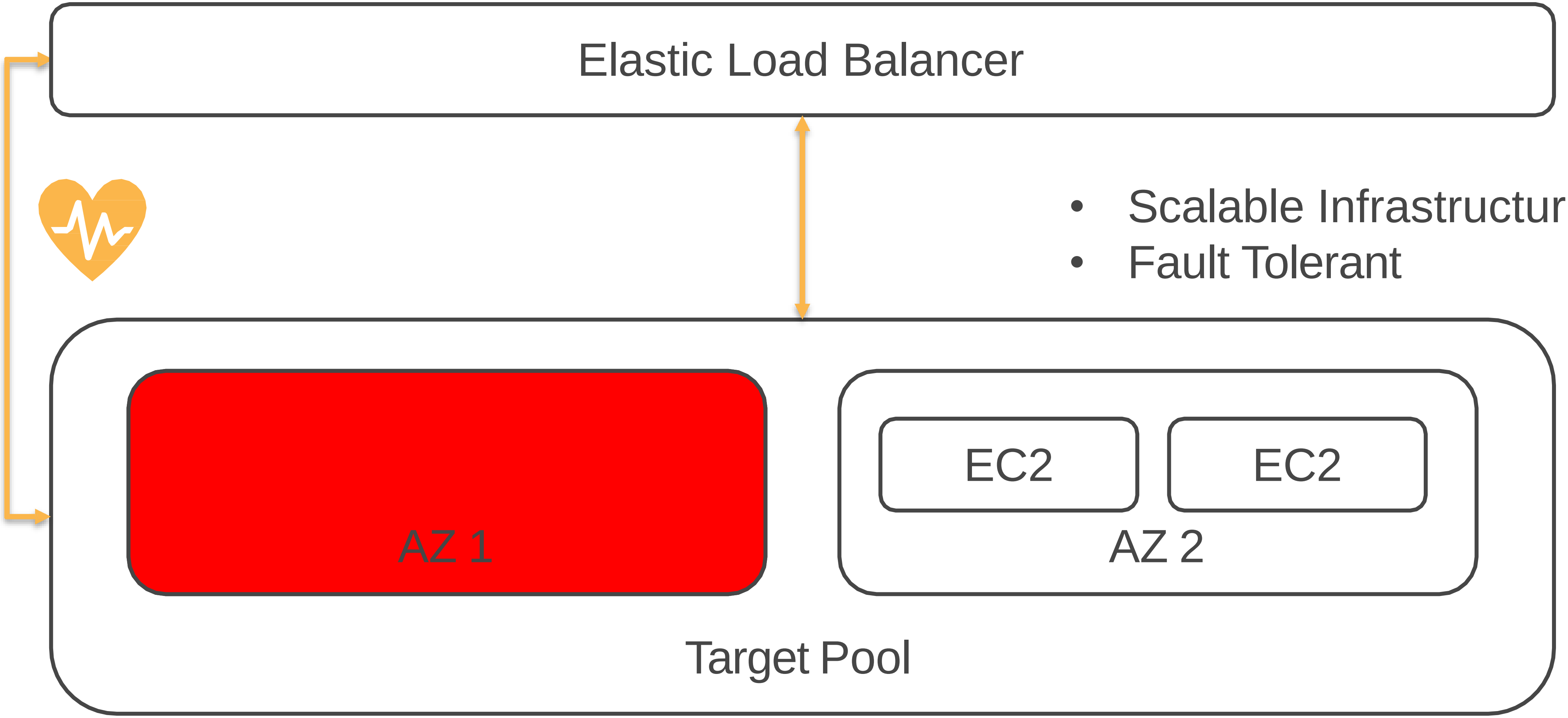
- Scalability Challenges
- Single point of failure

Elastic Load Balancing



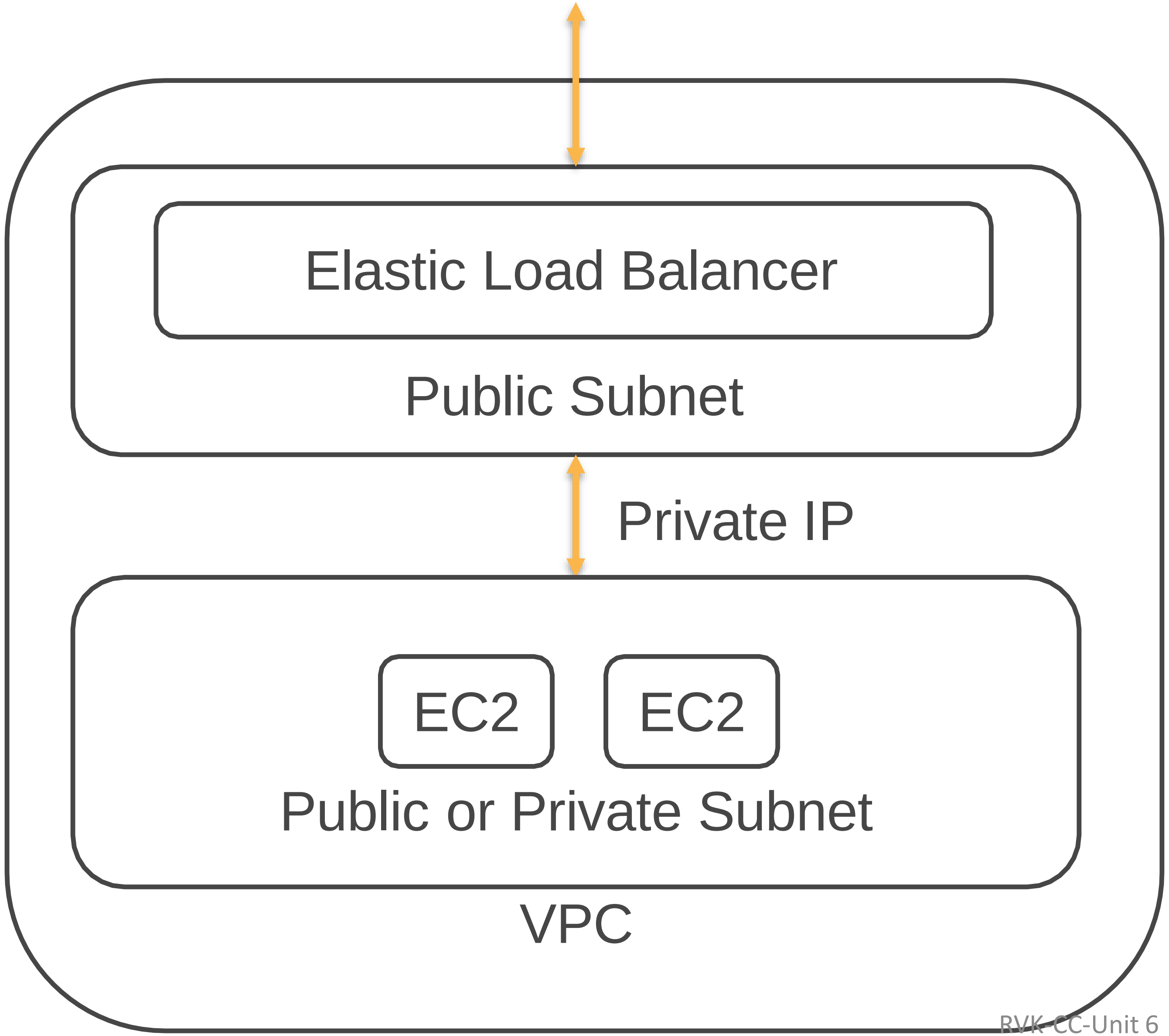
- Scalable Infrastructure
- Fault Tolerant

Elastic Load Balancing



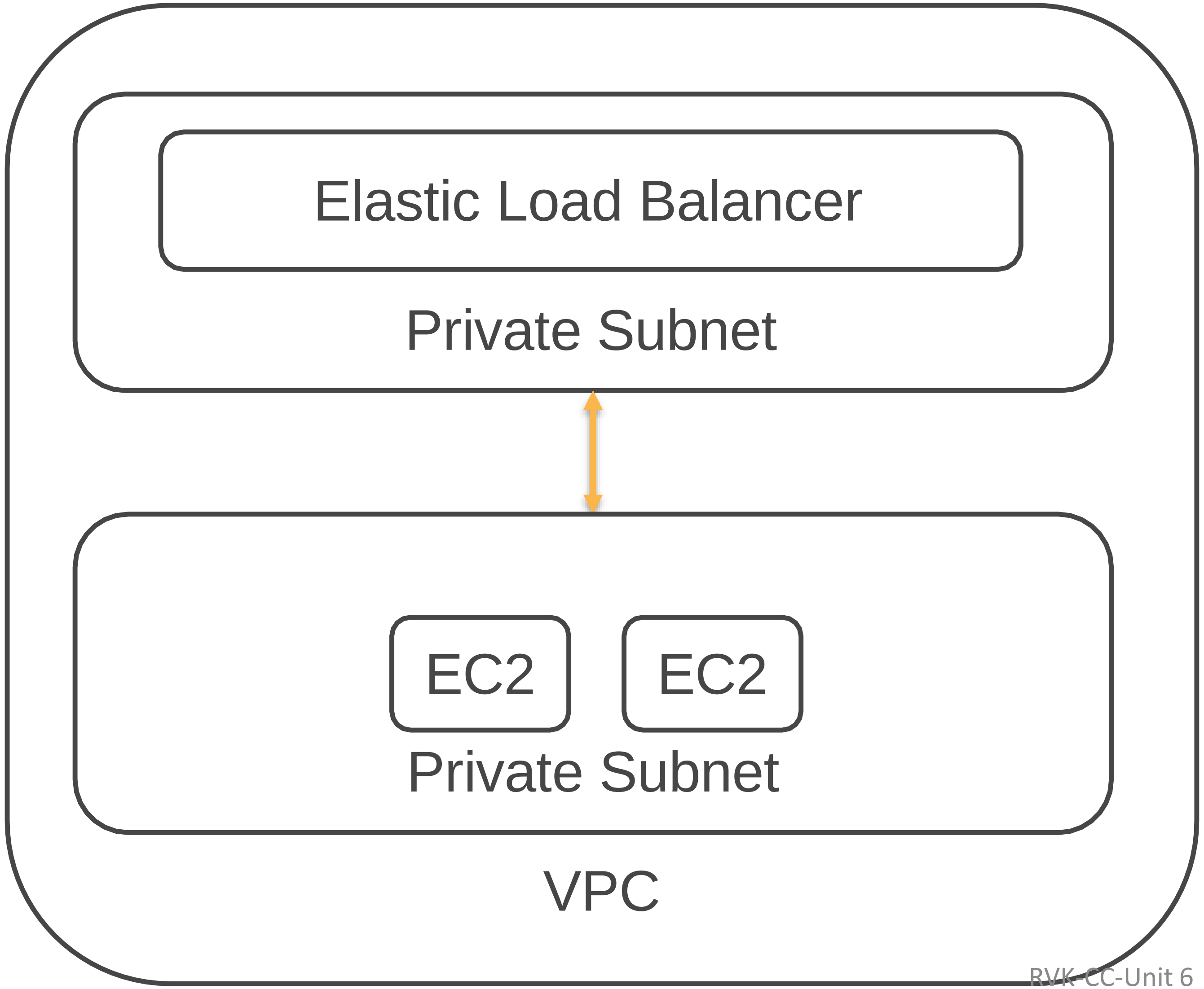
- Scalable Infrastructure
- Fault Tolerant

Elastic Load Balancing – Internet Facing



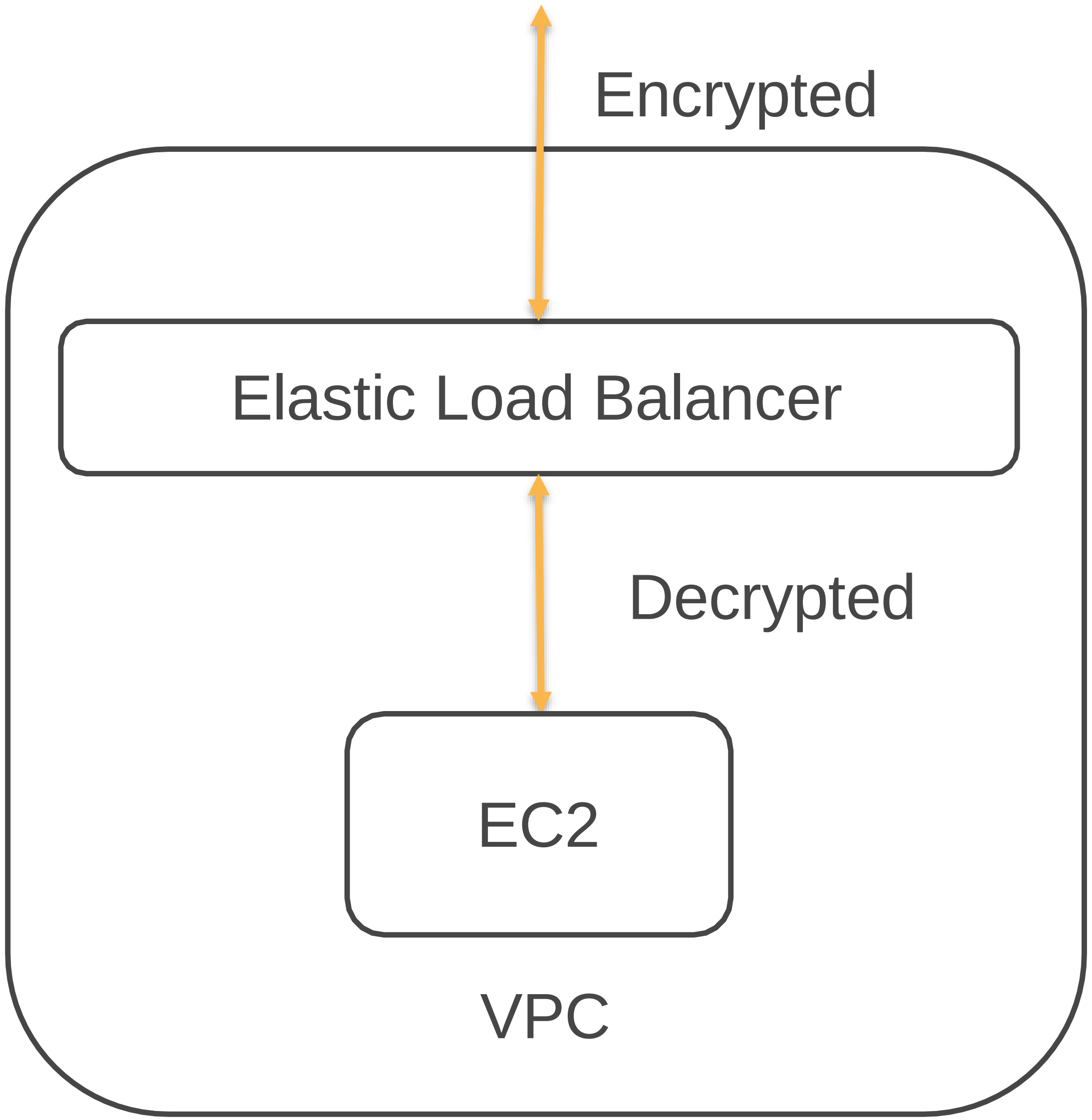
- Load Balancer is accessible from the internet.
  - Load Balancer talks to EC2 instance using Private IP
  - EC2 instances can be in public or private subnet
  - Reduces attack surface – EC2 instance configured only for private traffic
  - DDoS Protection

Elastic Load Balancing – Internal Facing



- Load Balancer is accessible only inside VPC

Elastic Load Balancing – Security



- Offload SSL/TLS
- Integrated Certificate Management
- User Authentication – Cognito (Application Load Balancer)
    - Internet Identity Providers
    - SAML
    - OpenID Connect

ELB Concepts

CloudWatch Monitoring

- Real time monitoring of key metrics

Connection Draining

- When deregistering instance, allow in-flight requests to complete
- Default wait time is 5 minutes (300 seconds)
- After wait time elapses, instance is deregistered

Sticky Sessions

- Route requests from a client to same target
- Used for stateful application - servers cache user data
- Disabled by default the Sticky session

HTTP/2

- Multiple requests sent on the same connection
- Efficient use of network resources

ELB Concepts

WebSockets

- Long running TCP Connection
- Bi-directional
- Server to Client Push notification support
- (eg. Gmail : Automatically receives mail alerts)

Cross Zone Load Balancing

- Enabled – distribute traffic evenly across all EC2 instances
- Disabled – distribute traffic evenly across availability zones

Load Balancer Access Logs

- Elastic Load Balancing provides access logs that capture detailed information about requests sent to your load balancer.

- Each log contains information such as the **time the request was received, the client's IP address, latencies, request paths, and server responses**.

- You can use these access logs to analyze traffic patterns and troubleshoot issues

- **Access logging is an optional feature of Elastic Load Balancing that is disabled by default**.

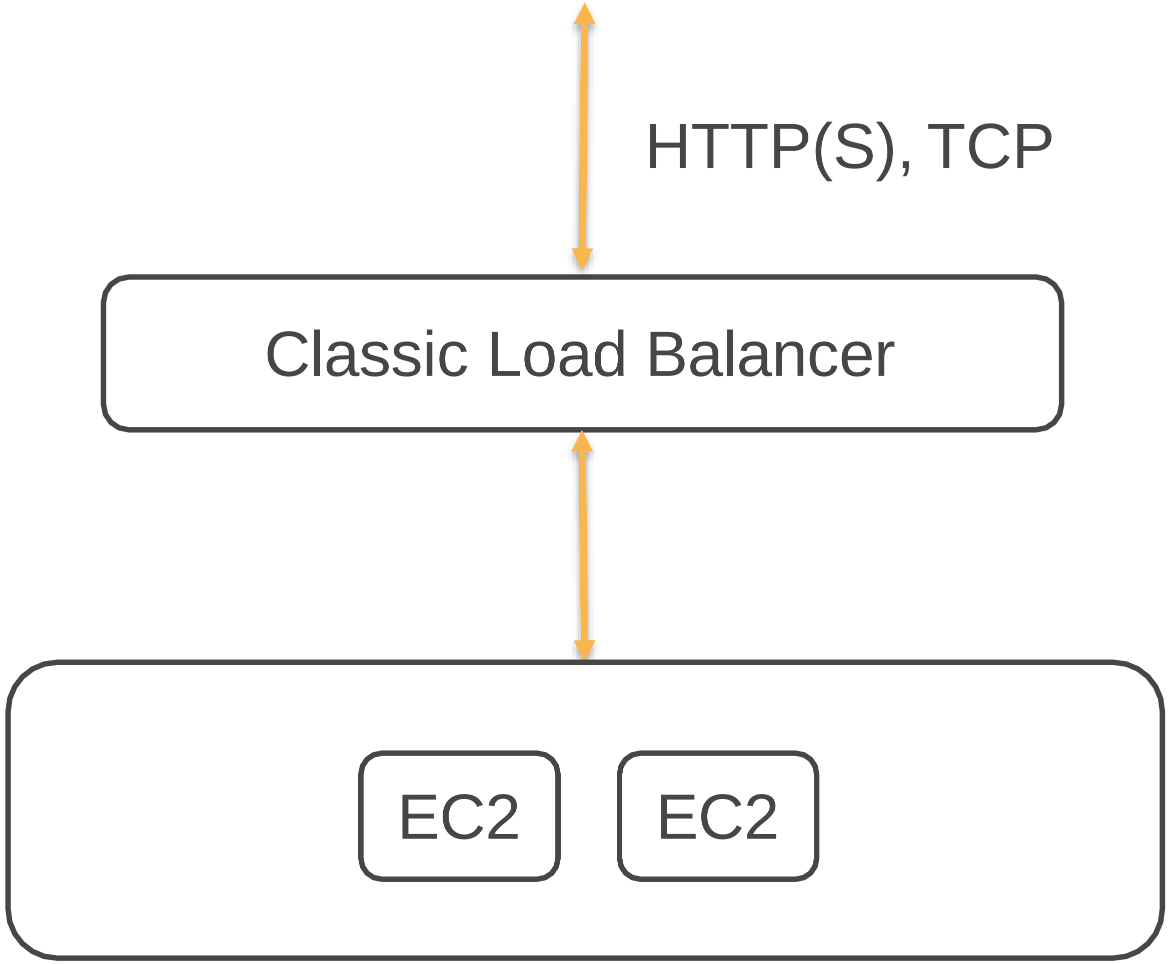
- After you enable access logging for your load balancer, Elastic Load Balancing captures the logs and stores them in the Amazon S3 bucket that you specify as compressed files.

Load Balancer Types

AWS Load Balancer - Types

Load Balancer	Use
Classic	• Basic load balancing across multiple EC2 instances • HTTP(S) (Layer 7) and TCP Support (Layer 4) • Recommended for legacy applications on EC2-Classic network
Application	• Load Balance across EC2 instances, Containers, Lambda, and Hybrid infrastructure • HTTP(S) traffic support (Layer 7) • Route traffic to target based on the content of the request
Network	• Load Balance across EC2 instances, Containers, Lambda, and Hybrid infrastructure • TCP, UDP traffic support (Layer 4) • Extreme performance

Classic Load Balancer



- Basic Load Balancing across multiple EC2 instances
- Supports HTTP(S) (Layer 7) and TCP (Layer 4) traffic
- Works both on EC2-Classic and VPC
- Previous generation product – recommended only for EC2-Classic

Application Load Balancer (ALB)

- It operates at the request level (layer 7), routing traffic to targets (EC2 instances, containers, IP addresses, and Lambda functions) based on the content of the request.

- Ideal for advanced load balancing of HTTP and HTTPS traffic. HTTP/2 and WebSocket Support

- It provides advanced request routing targeted at delivery of modern application architectures, including microservices and container-based applications.

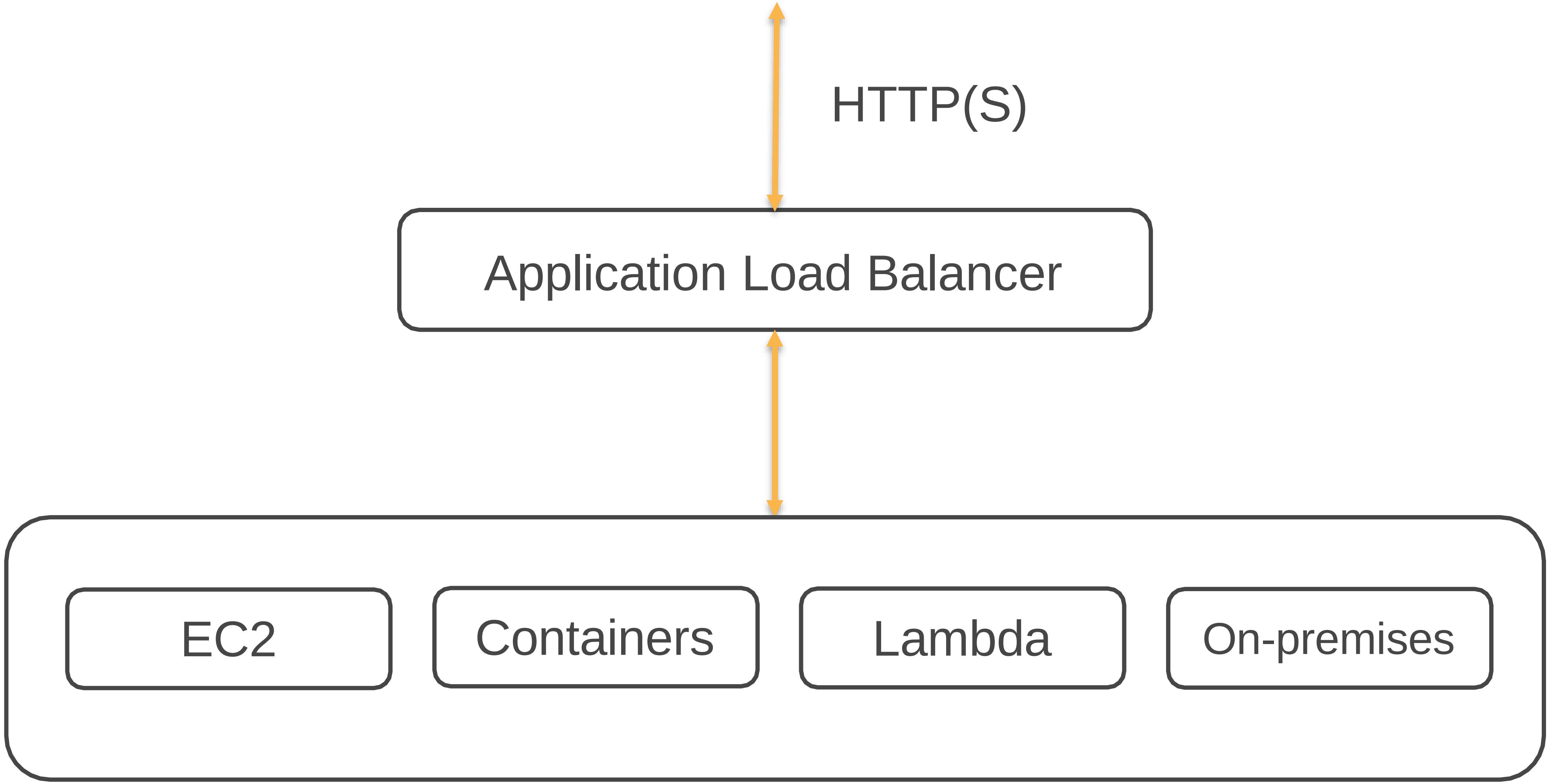
- It simplifies and improves the security of your application, by ensuring that the latest SSL/TLS ciphers and protocols are used at all times.

- Request Tracing – track individual request by unique ID across various services

- Support for hosting multiple websites (Server Name Indication)

- User Authentication - Cognito

Application Load Balancer (cont..)



HTTP(S)

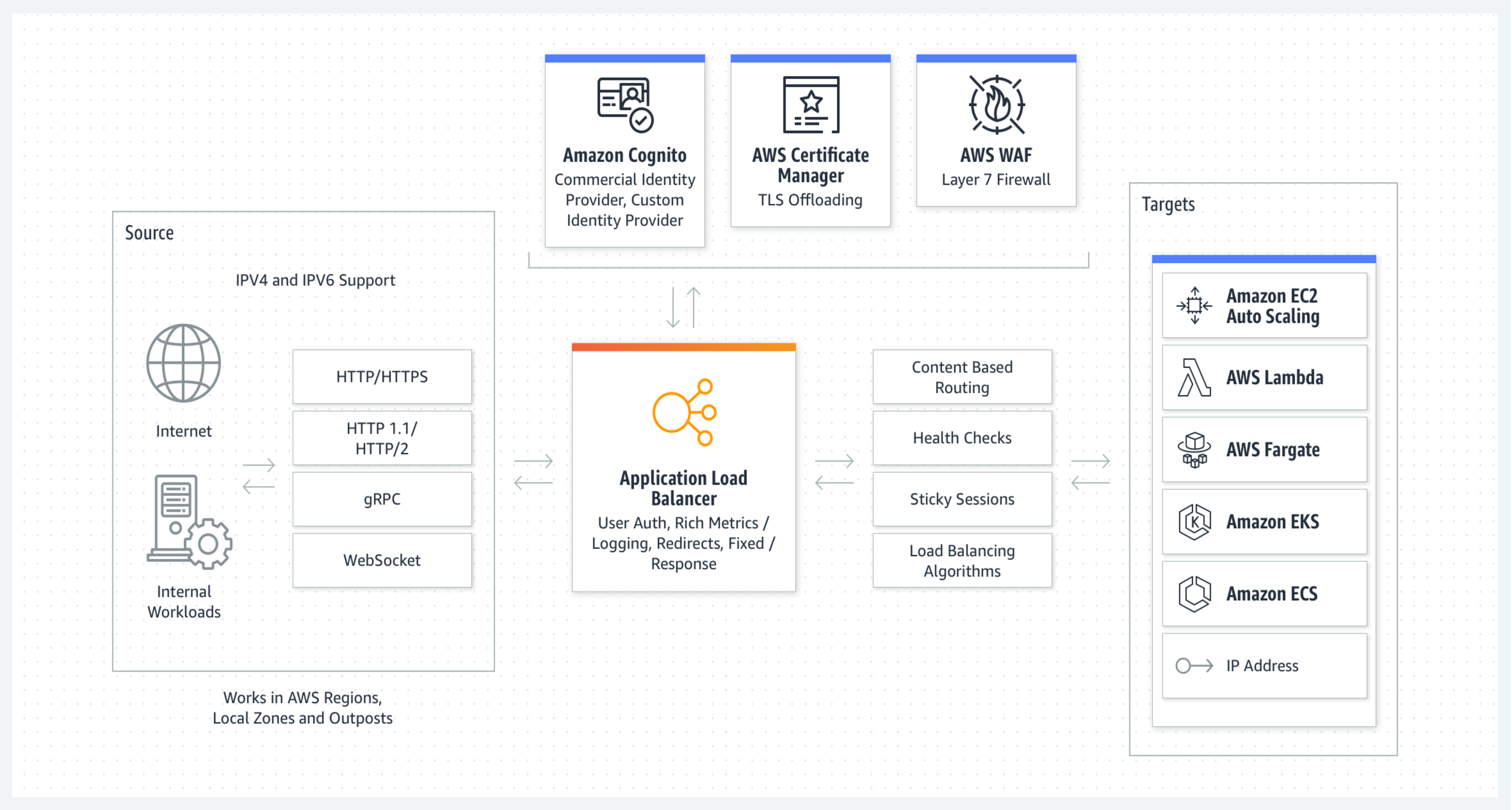
Application Load Balancer

EC2

Containers

Lambda

On-premises



Amazon Cognito
Commercial Identity Provider, Custom Identity Provider

AWS Certificate Manager
TLS Offloading

AWS WAF
Layer 7 Firewall

Source

IPV4 and IPV6 Support

Internet

Internal Workloads

HTTP/HTTPS

HTTP 1.1/ HTTP/2

gRPC

WebSocket

Works in AWS Regions, Local Zones and Outposts

Application Load Balancer
User Auth, Rich Metrics / Logging, Redirects, Fixed / Response

Content Based Routing

Health Checks

Sticky Sessions

Load Balancing Algorithms

Targets

Amazon EC2 Auto Scaling

AWS Lambda

AWS Fargate

Amazon EKS

Amazon ECS

IP Address

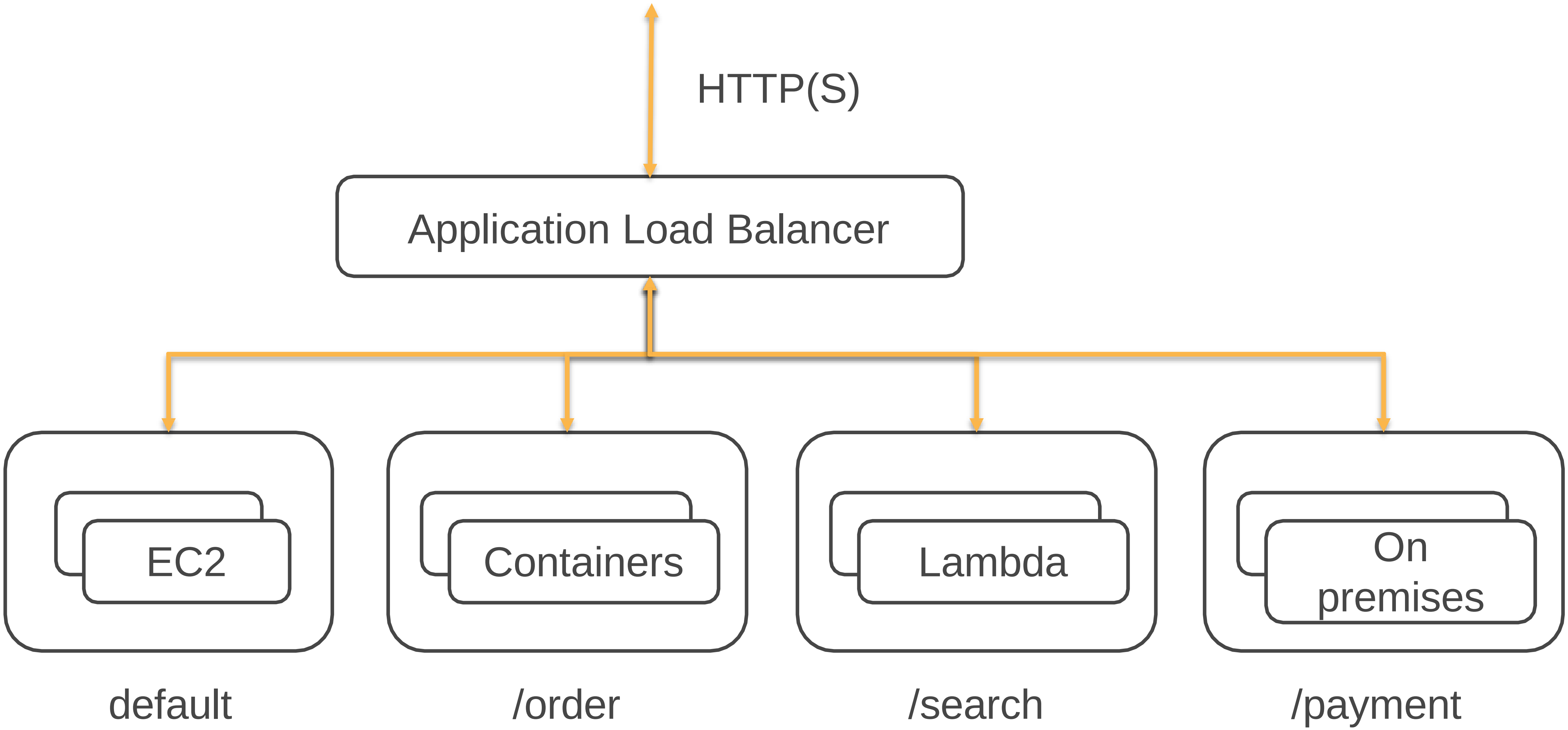
Application Load Balancer

Application Load Balancer - Routing

Traffic is routed through

- Path based
- Host HTTP header (support for multiple domains)
- Any standard or custom HTTP header
- Query string parameter based
- Source IP based (from where request is originating)

Application Load Balancer - Routing



HTTP(S)

Application Load Balancer

EC2

Containers

Lambda

On premises

default

/order

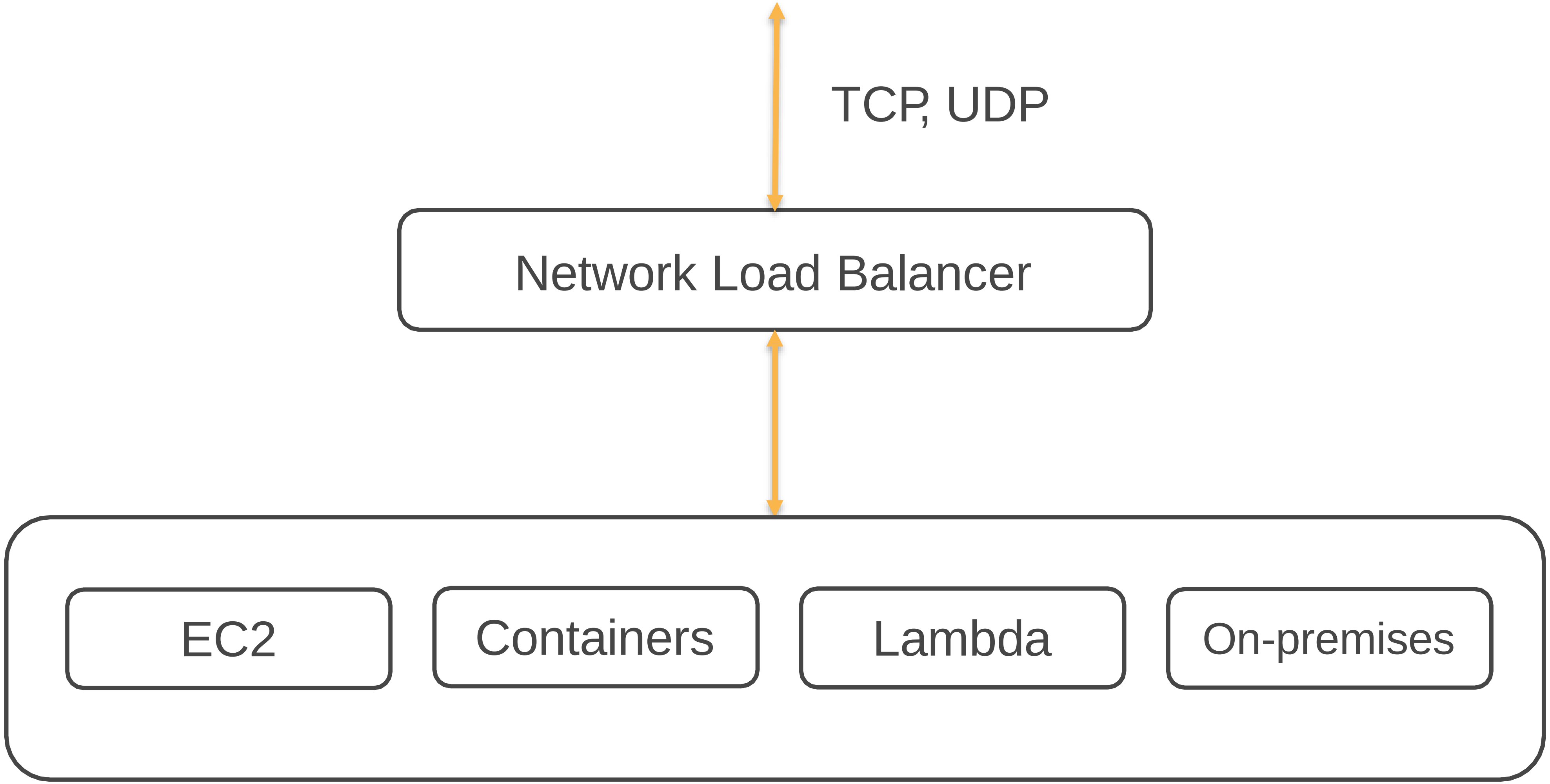
/search

/payment

Network Load Balancer (NLB)

- It operates at the connection level (Layer 4), routing connections to targets (Amazon EC2 instances, microservices, and containers) within Amazon VPC, based on IP protocol data.
- Ideal for load balancing of both TCP and UDP traffic.
- NLB is capable of handling millions of requests per second while maintaining ultra-low latencies.
- It is optimized to handle sudden and volatile traffic patterns while using a single static IP address per Availability Zone.
- It is integrated with other popular AWS services such as Auto Scaling, Amazon EC2 Container Service (ECS), Amazon CloudFormation, and AWS Certificate Manager (ACM).
- Preserves Client IP (Source IP) – your application can use this for further processing
- WebSocket Support
- Private Link Support – Private communication between VPCs

Network Load Balancer (cont..)



TCP, UDP

Network Load Balancer

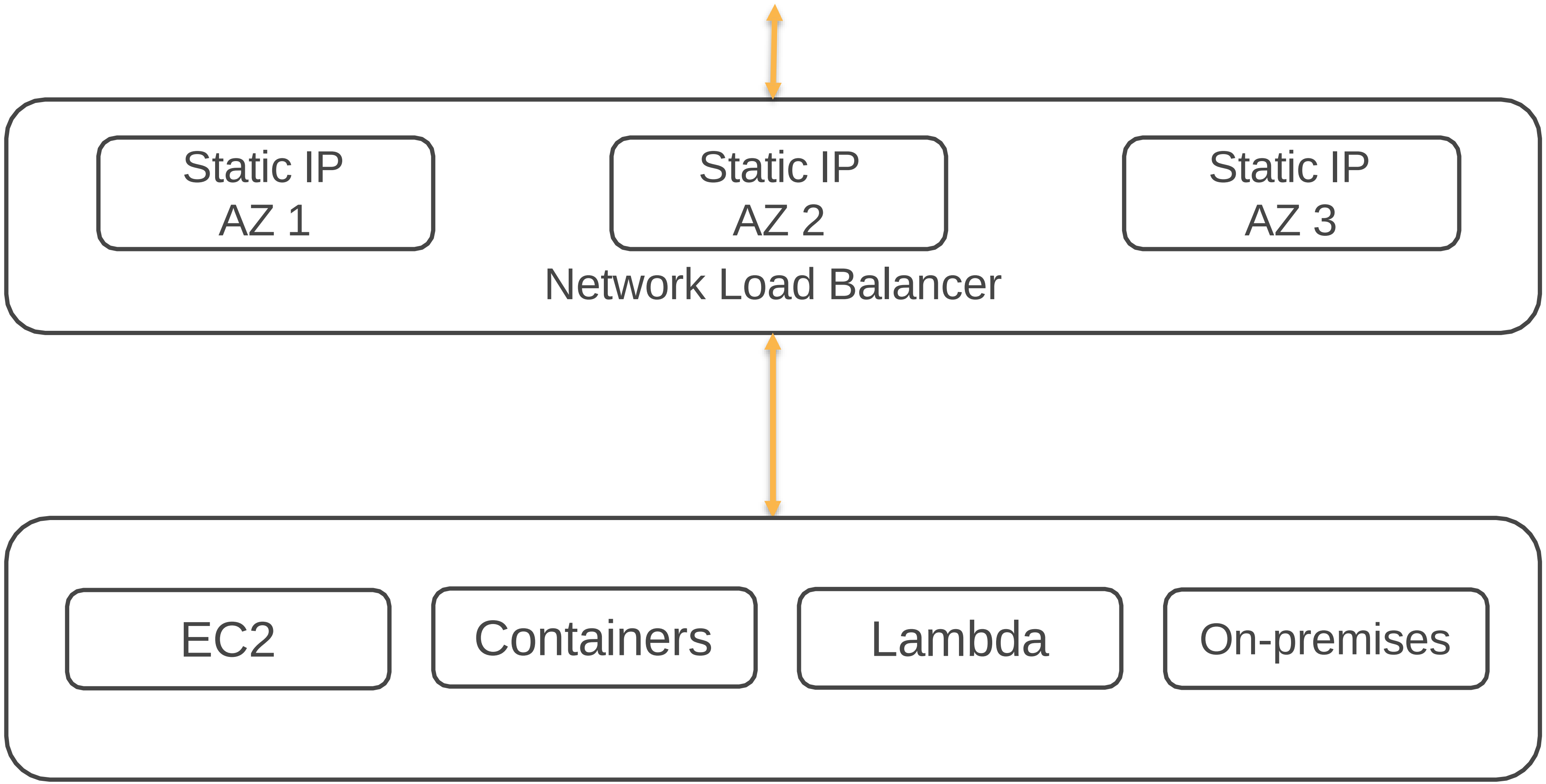
EC2

Containers

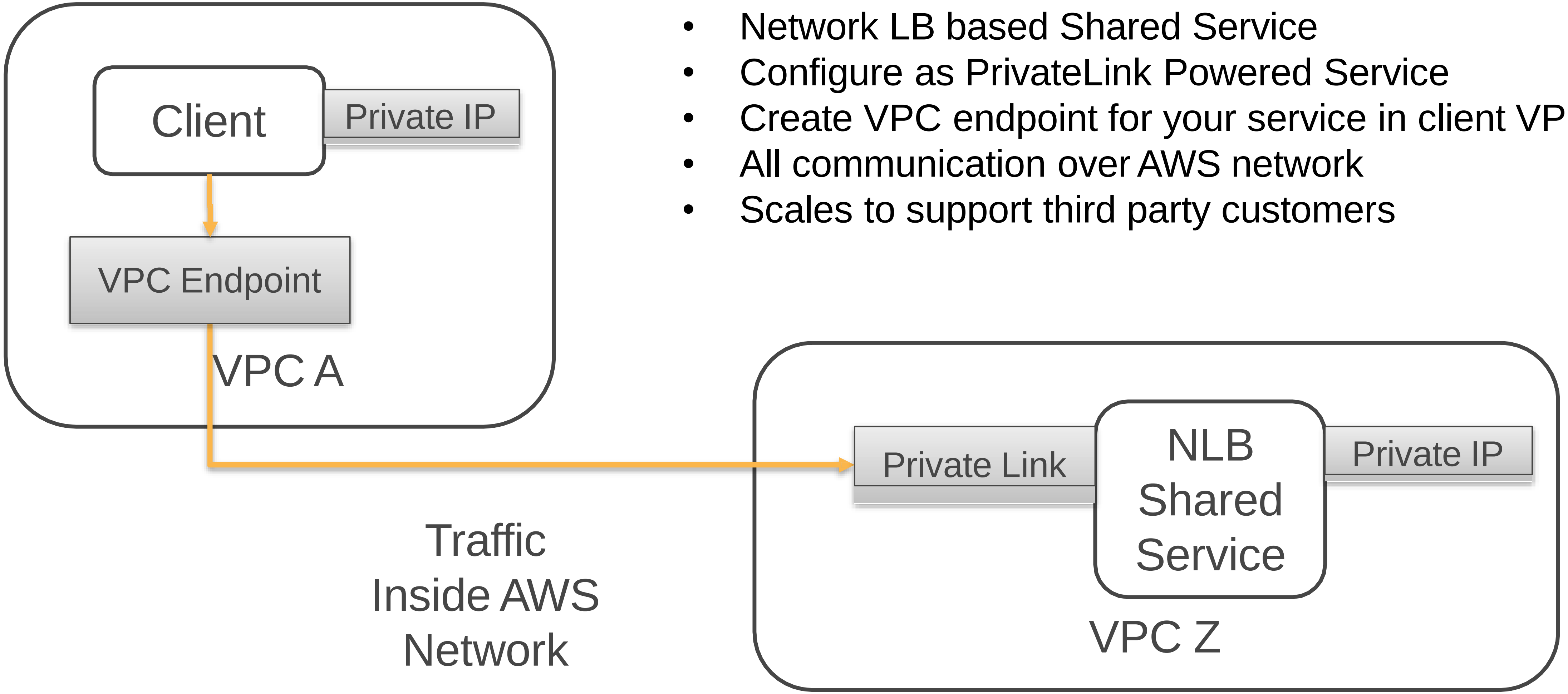
Lambda

On-premises

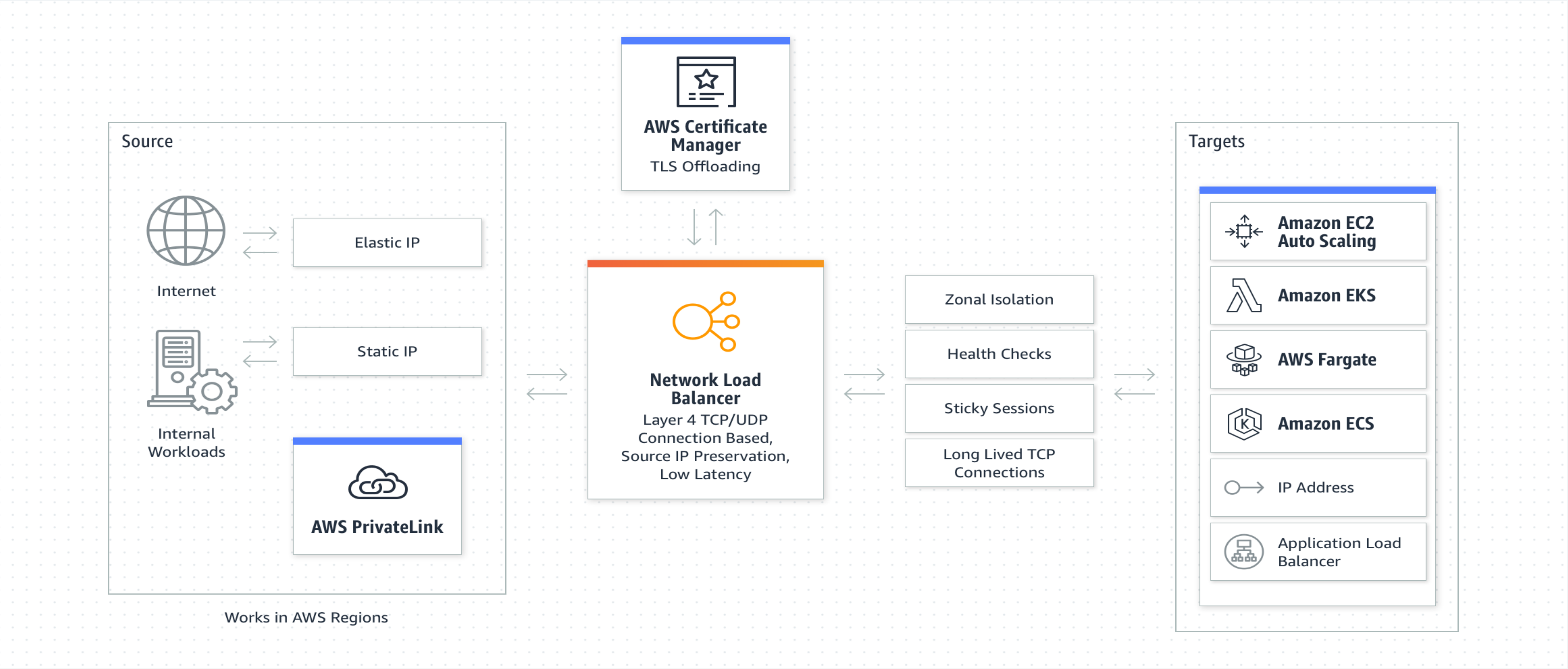
Network Load Balancer – Static IP



NLB-Private Link



- Network LB based Shared Service
- Configure as PrivateLink Powered Service
- Create VPC endpoint for your service in client VPC
- All communication over AWS network
- Scales to support third party customers



Network Load Balancer

Thank you!