Gokop Goteng

Designing Your Cloud Environment

What’s In This Module?

- **Part 1:** Choosing a Region
- **Part 2:** Selecting Availability Zones
- **Part 3:** Virtual Private Cloud (VPC)
- **Part 4:** Dividing VPCs and Subnets
- **Part 5:** Default VPCs and Default Subnets

Part 1: How to Choose a Region.

How to Choose a Region?

Data sovereignty and compliance

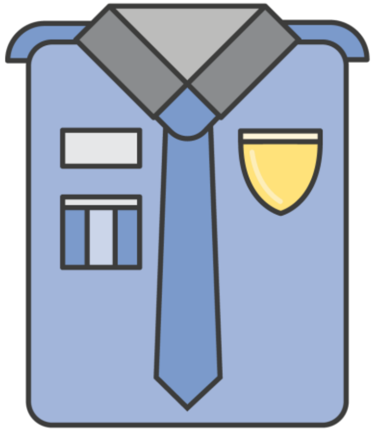
Proximity of users to data

Availability of services and features

Cost effectiveness

Data Sovereignty and Compliance

Where can you legally host your infrastructure?



What are the national and local data security laws?

Is customer data allowed outside of the country?

Can you meet governance requirements?

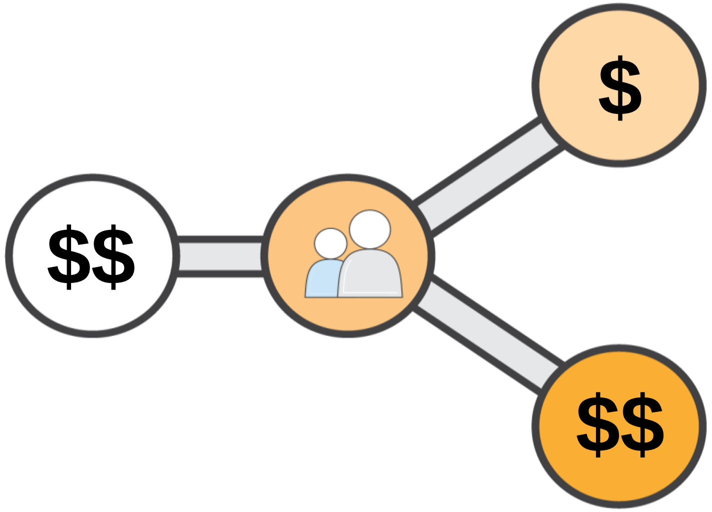
Did you know?
AWS opened its first carbon-neutral Region in 2011 and now offers five!

Learn more.

Proximity of Users to Data

What is the proximity to your user base?

Study: 100-ms delays can cost 1% in sales on Amazon.com



Equidistant regions?
Compare costs

Availability of Services and Features

What services and features are available?

Some services available in limited regions

Some services can cross-regions, but at increased latency

Services expanded to new regions regularly

Cost-effectiveness

Consider cost-effectiveness.

Service costs vary by region.

Some services (i.e. Amazon S3) have costs for transferring data out.

Consider replicating entire environment to another region.

Part 2: How Many Availability Zones Should You Use?

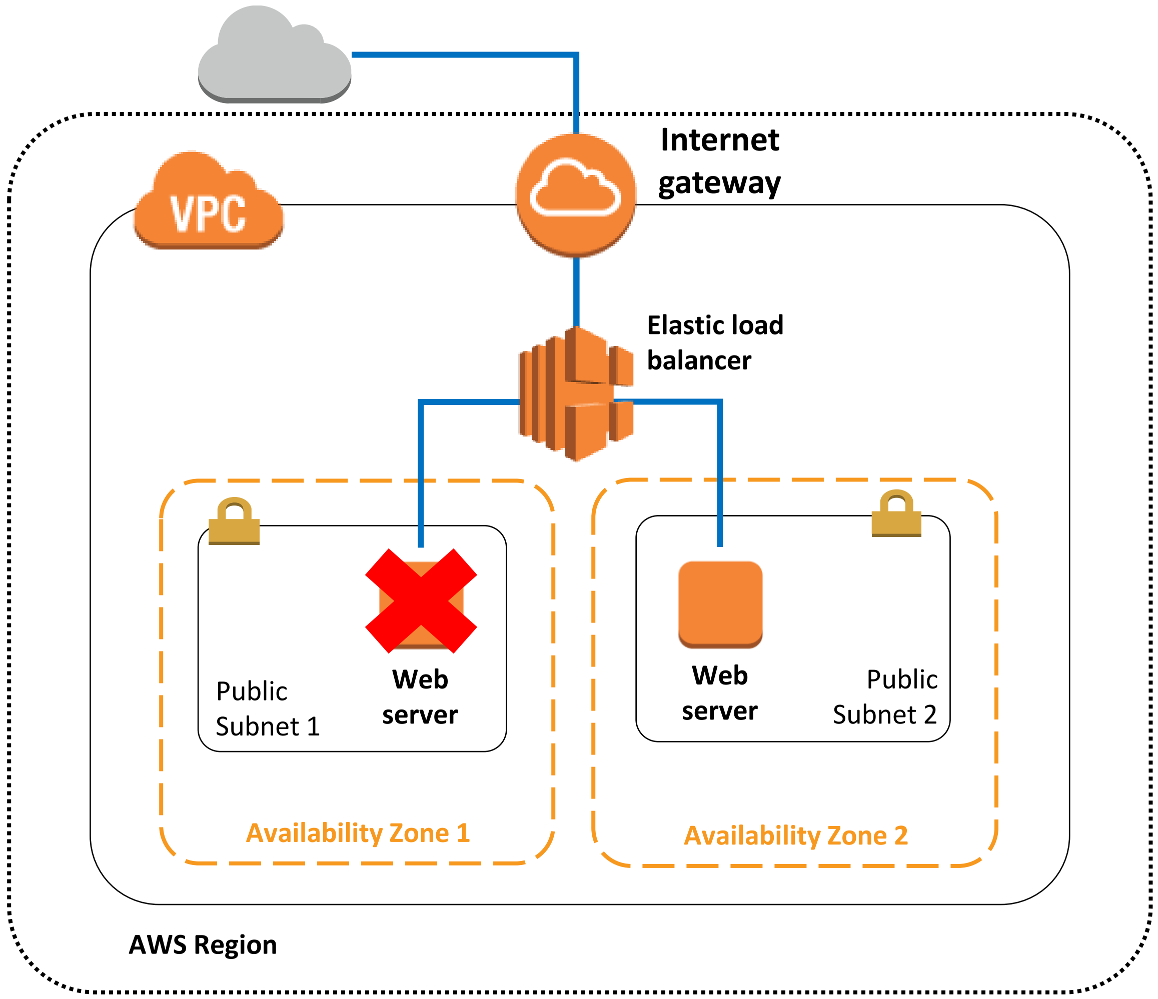
How Many Availability Zones Should You Use?

Recommendation: Start with two Availability Zones per Region.

- **Best practice:** If resources in one Availability Zone are unreachable, your application shouldn't fail.
- Most applications can support two Availability Zones.

Something to consider:
For heavy usage (Amazon DynamoDB) it may be beneficial to use more than two Availability Zones.

Using Two Availability Zones



Recommended Availability Zones

How many Availability Zones should be recommended for each scenario?

01

Applications that heavily use Amazon EC2 Spot Instances

2 Availability Zones
(or more based on price options)

02

Applications with MySQL, MS SQL Server, Oracle data sources

2 Availability Zones
to support active/passive

03

Applications with data sources Cassandra or MongoDB

2 or more Availability Zones
for extremely high availability

Part 3: Virtual Private Cloud (VPC)

Using One VPC

There are **limited** use cases where one VPC could be appropriate:

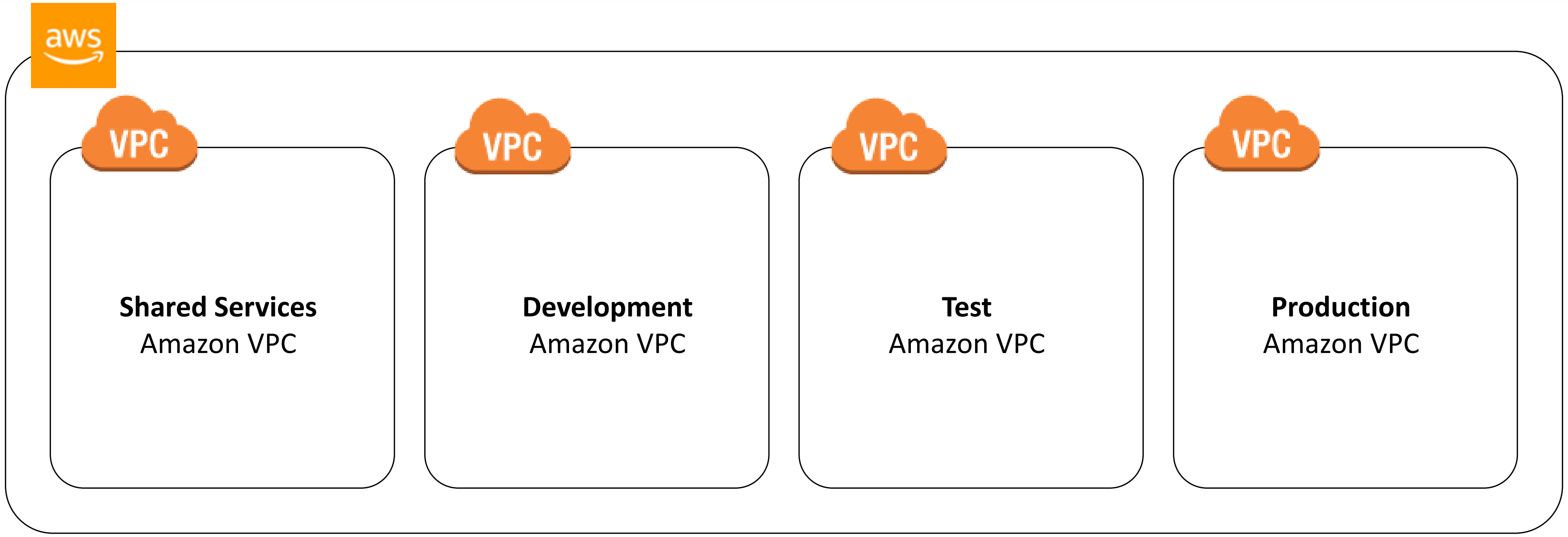
- High-performance computing environments
- Microsoft active directory for identity management
- Small, single applications managed by one person or very small team

For **most** use cases, there are two primary patterns for organizing your infrastructure:

Multi-VPC or Multi-Account

Multi-VPC and Multi-Account Patterns

Multi-VPC Pattern



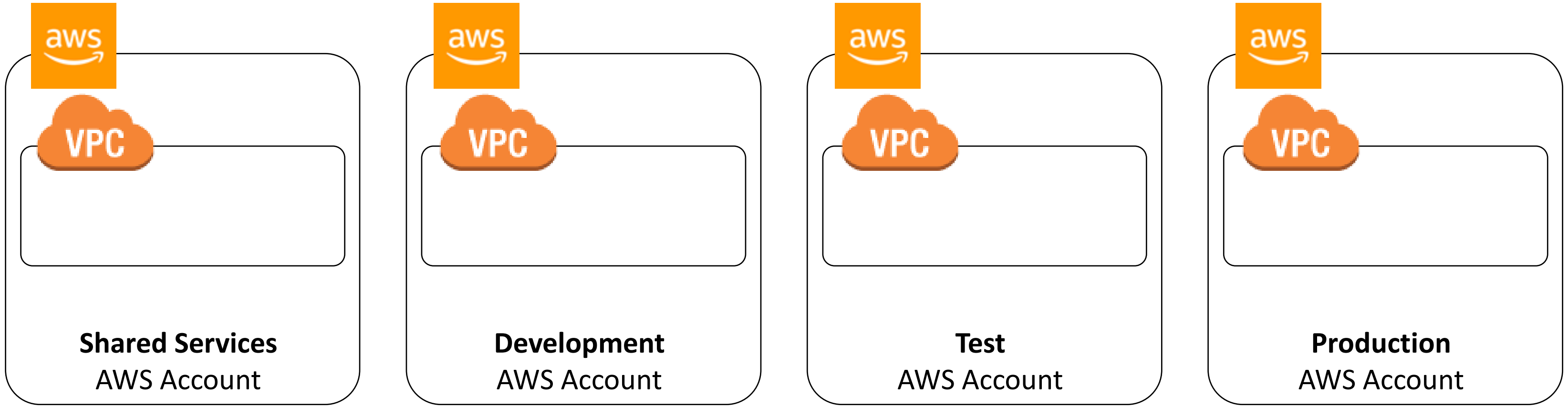
Shared Services
Amazon VPC

Development
Amazon VPC

Test
Amazon VPC

Production
Amazon VPC

Multi-Account Pattern



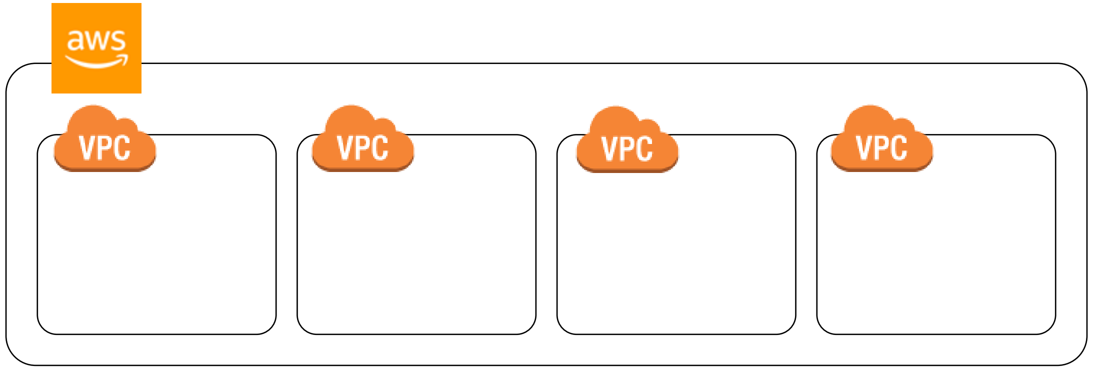
Shared Services
AWS Account

Development
AWS Account

Test
AWS Account

Production
AWS Account

Multi-VPC Pattern



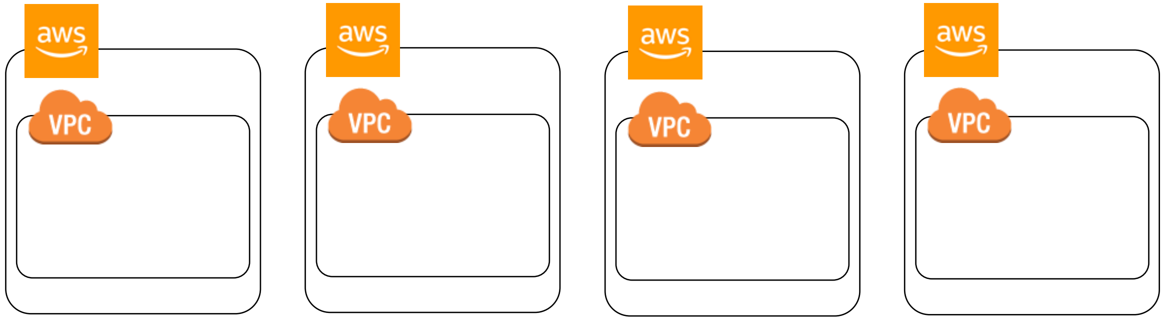
Best suited for:

- **Single team or organizations**, such as Managed Service Providers.
- Limited teams make **maintaining standards** and **managing access** far easier.

Exception:

- **Governance** and **compliance standards** may require workload isolation regardless of organizational complexity.

Multi-Account Pattern



Best suited for:

- **Larger organizations** and **organizations with multiple IT teams.**
- **Medium-sized organizations** that anticipate rapid growth.

Why?

- **Managing access** and **standards** can be challenging in more complex organizations.

Thoughts to consider:
Where will your team be in five years?

AWS Organizations

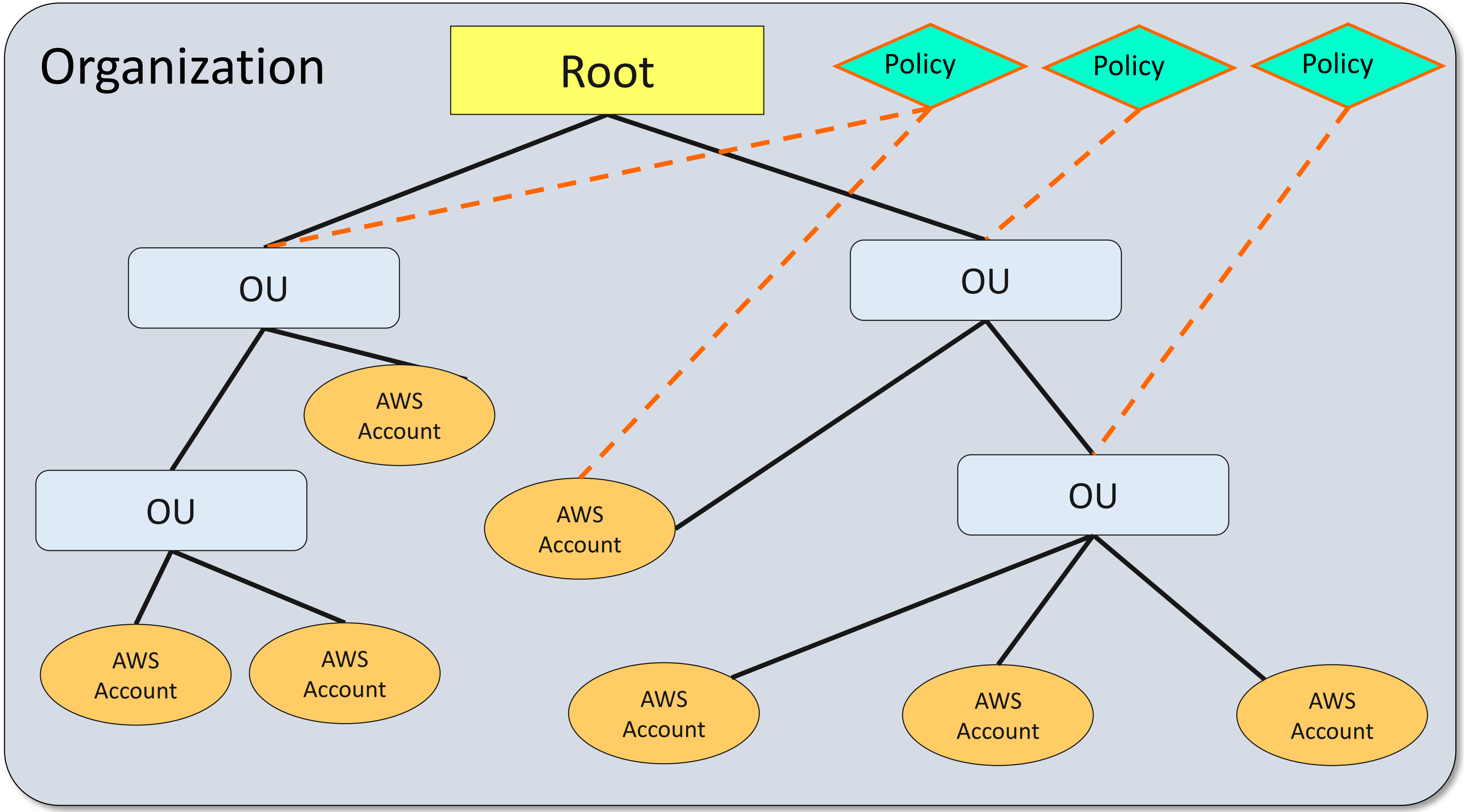
- Account management service.
- Consolidate multiple AWS accounts into an organization and arrange AWS accounts into organizational units.
- Consolidated billing and account management capabilities.

AWS Organizations

AWS Organizations: Features

- Hierarchical grouping of your accounts.
- Integration and support for AWS Identity and Access Management (IAM).
- Integration with other Amazon web services.
- Data replication with eventual consistency.
- Caching may improve performance.

AWS Organizations Key Concepts



Definitions

- Organization
- Root
- Organization Unit (OU)
- Account
- Invitation
- Handshake
- Service Control Policy

Other Important Considerations

The majority of AWS services **do not actually sit within a VPC.**

- Network traffic between AWS Regions traverse the AWS global network backbone by default.
- Sometimes traffic between regions uses the public internet.
- Amazon S3 and DynamoDB offer **VPC endpoints** (powered by PrivateLink) to connect without traversing the public internet.

Learn more.

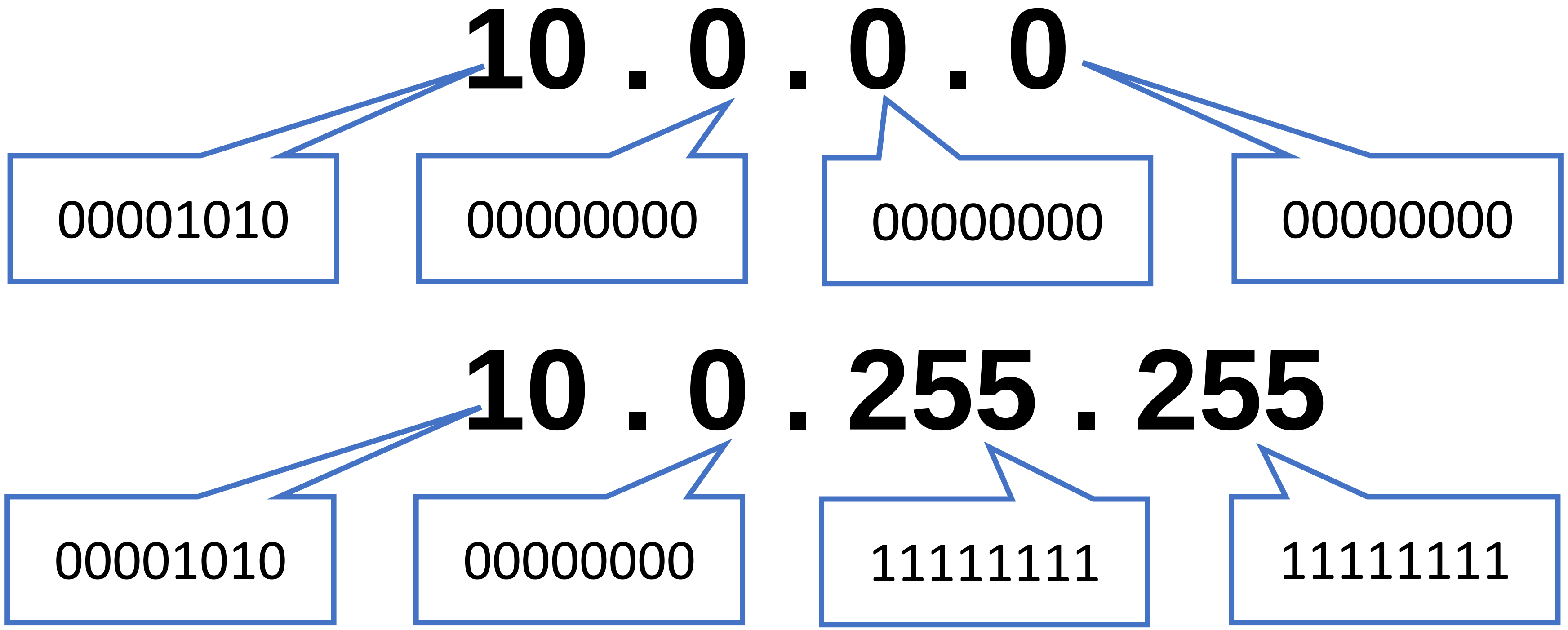
Part 4: Divide Your VPC into Subnets

VPCs and IP Addresses

- When you create your VPC, you specify its set of IP addresses with CIDR notation.
- **Classless Inter-Domain Routing (CIDR)** notation is a simplified way to show a specific range of IP addresses.
  - Example: 10.0.0.0/**16** = all IP addresses from 10.0.0.0 to 10.0.255.255
- **How does that work?** What does the **16** define?

IP Addresses and CIDR

Every set of 3 digits in an IP address represents a set of 8 binary values (8 bits).



10 . 0 . 0 . 0

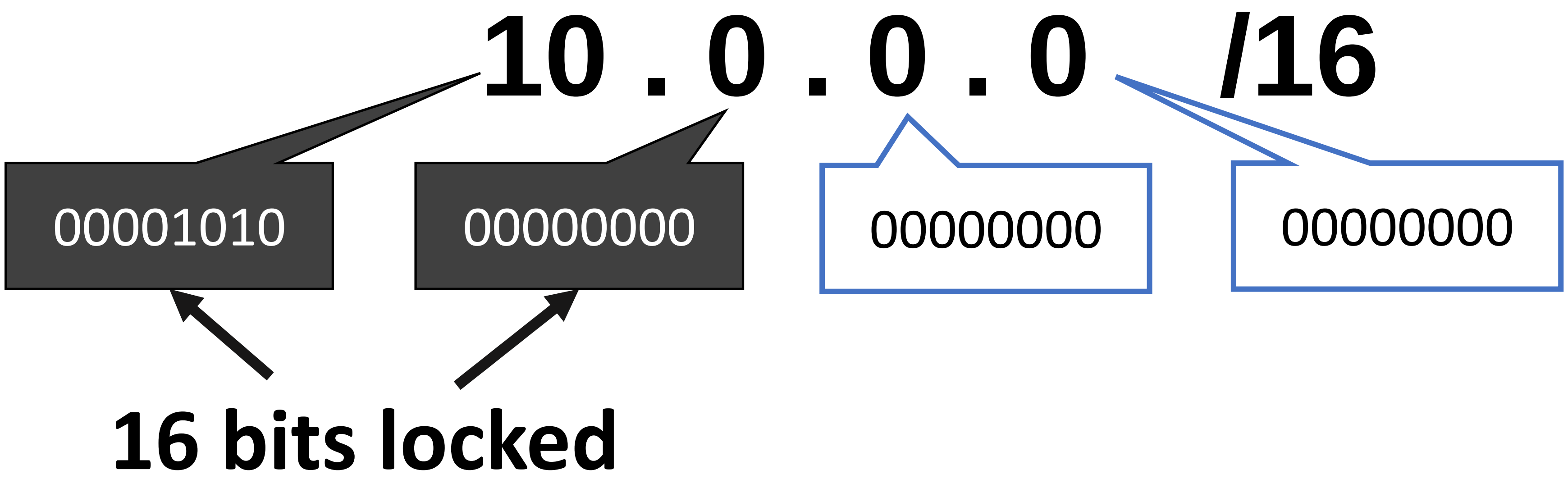
00001010 | 00000000 | 00000000 | 00000000

10 . 0 . 255 . 255

00001010 | 00000000 | 11111111 | 11111111

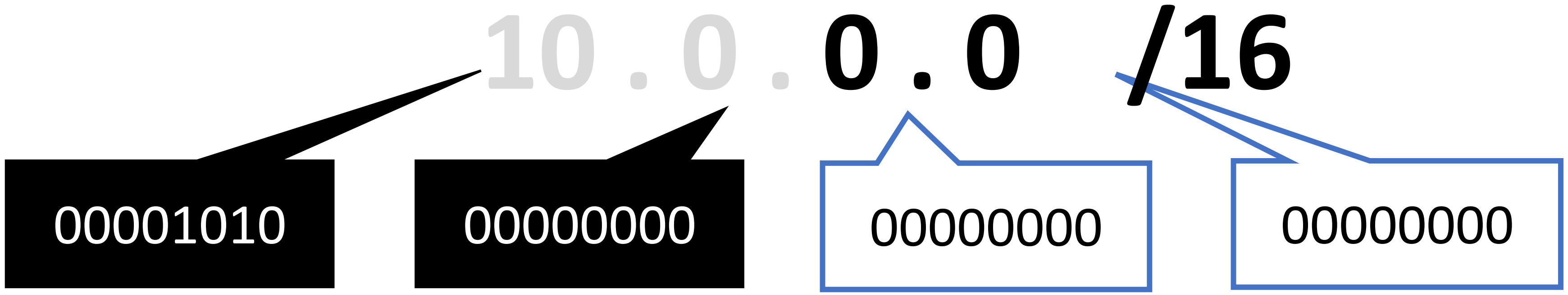
IP Addresses and CIDR: Part I

The 16 in the CIDR notation example represents how many of those bits are "locked down" and cannot change.



IP Addresses and CIDR: Part II

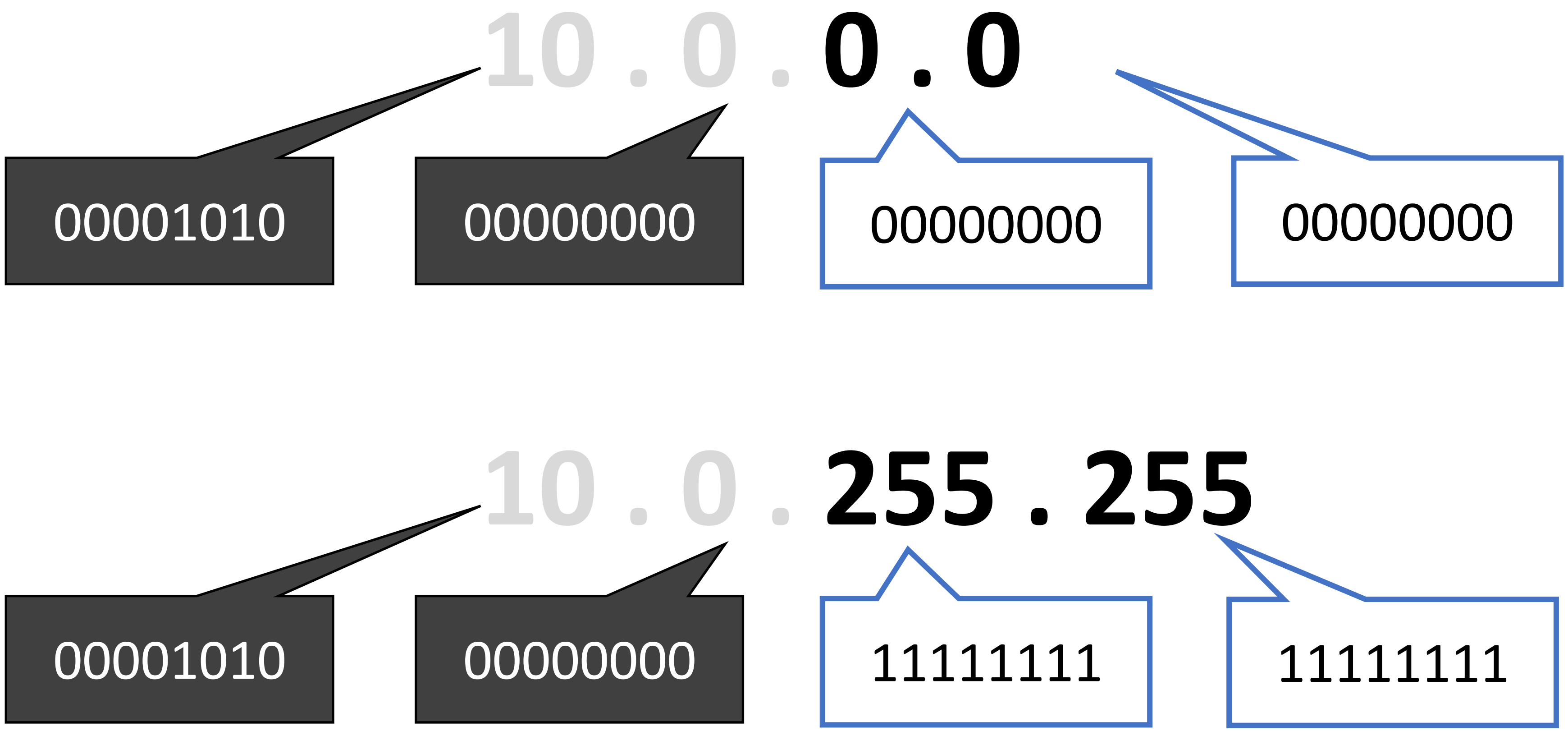
The unlocked bits can change between 1 and 0, allowing the full range of possible values.



CIDR Example: 10.0.0.0/16

Lowest possible IP

Highest possible IP



10 . 0 . 0 . 0

00001010 | 00000000 | 00000000 | 00000000

10 . 0 . 255 . 255

00001010 | 00000000 | 11111111 | 11111111

VPCs and IP Addresses

Amazon VPCs can use CIDR ranges between **/16** and **/28**.

For every **one step** a CIDR range increases, the total number of IP addresses is **cut in half.**

Dedicated Network Bits (CIDR)	Bits available to IPs (Total IPs)
/16	**65,536**
/17	**32,768**
/18	**16,384**
/19	**8,192**
/20	**4,096**
…	…
/28	**16**

What Are Subnets?

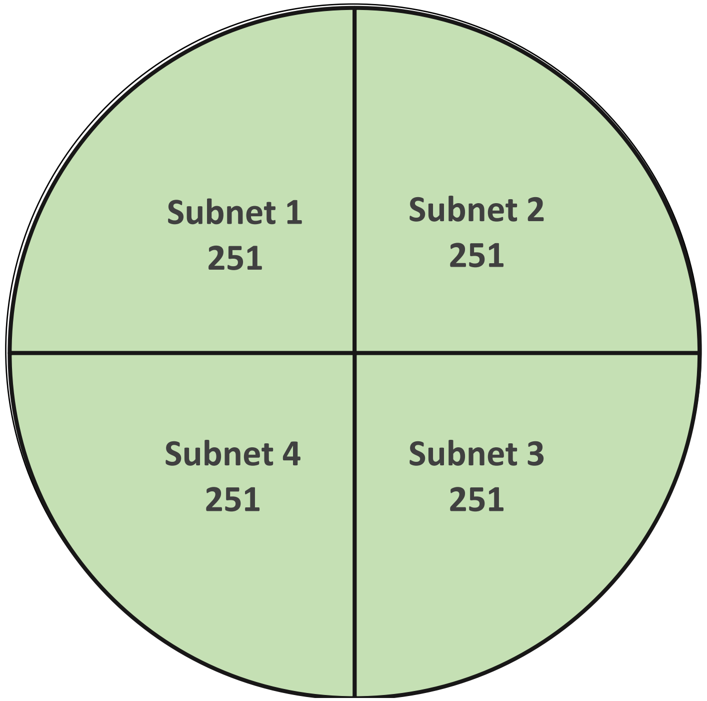
Subnets are **segments** or **partitions** of a network, divided by **CIDR range**.

Example:

A VPC with **CIDR /22** includes 1,024 total Ips (32-22=10, 2^{10}=1,024)

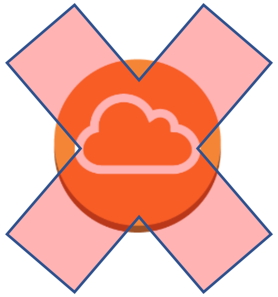
Note: In every subnet, the first four and last IP addresses are reserved for AWS use.

- 10.0.0.0: Network address.
- 10.0.0.1: Reserved by AWS for the VPC router.
- 10.0.0.2: Reserved by AWS for mapping to Amazon provided DNS.
- 10.0.0.3: Reserved by AWS for future use.
- 10.0.0.255: Network broadcast address.



Public and Private Subnets

Internet gateway



Internet gateway

Public subnets

- If a subnet’s traffic is routed to an internet gateway, the subnet is a *public subnet.*

Private subnets

- If a subnet’s traffic does not have a route to an internet gateway, the subnet is a *private subnet.*

Learn more.

Public and Private Subnets Part II

Recommendation: Use subnets to define internet accessibility.

Public subnets

- Include a routing table entry to an **internet gateway** to support inbound/outbound access to the public internet.

Private subnets

- Do not have a routing table entry to an internet gateway and are **not directly accessible** from the public internet.
- Use a "jump box" (NAT/proxy/bastion host) to support restricted, **outbound-only** public internet access.

How to Use Subnets

Recommendation: Use subnets to define Internet accessibility.

Public subnets

- Include a routing table entry to an **Internet gateway** to support inbound/outbound access to the public Internet.

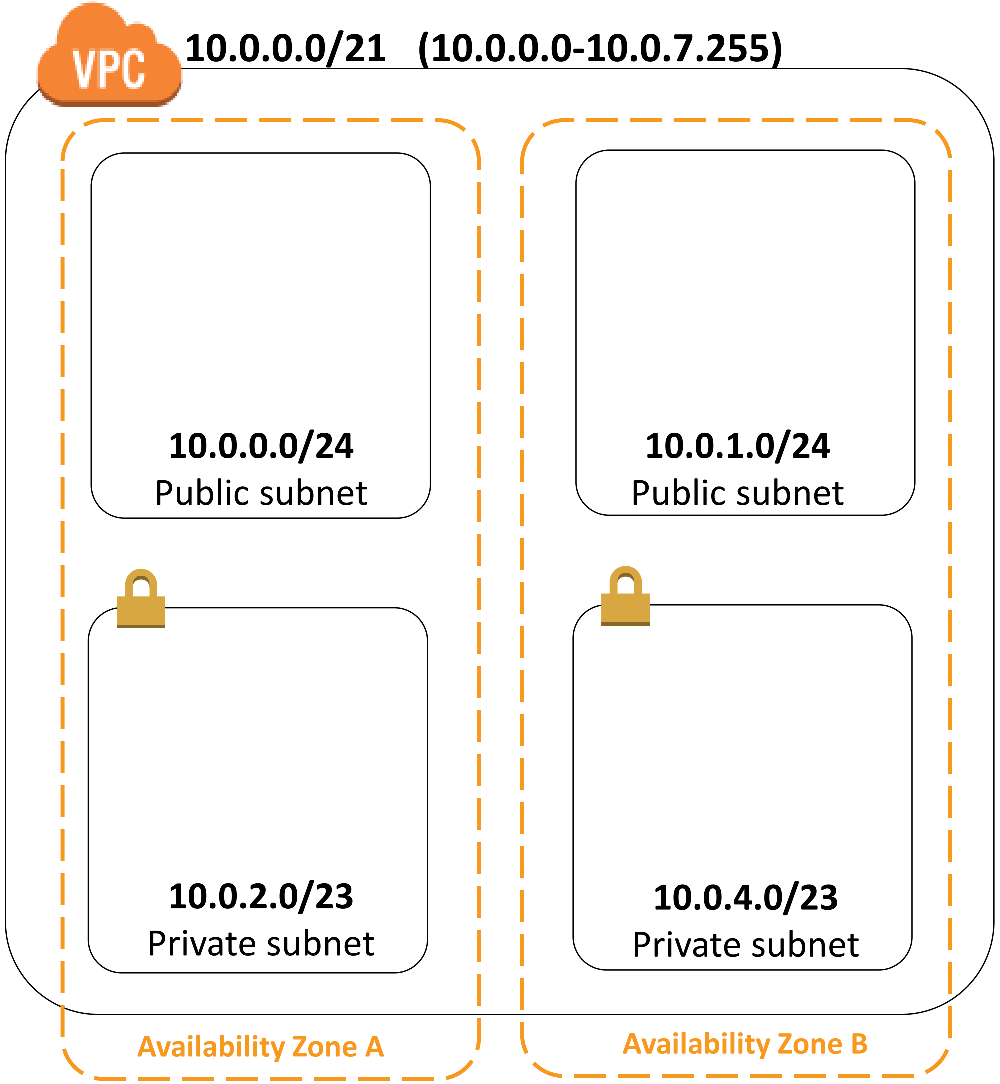
Private subnets

- Do not have a routing table entry to an Internet gateway and are **not directly accessible** from the public Internet.
- Typically use a "jump box" (NAT/proxy/bastion host) to support restricted, **outbound-only** public Internet access.

Subnets

Recommendation:

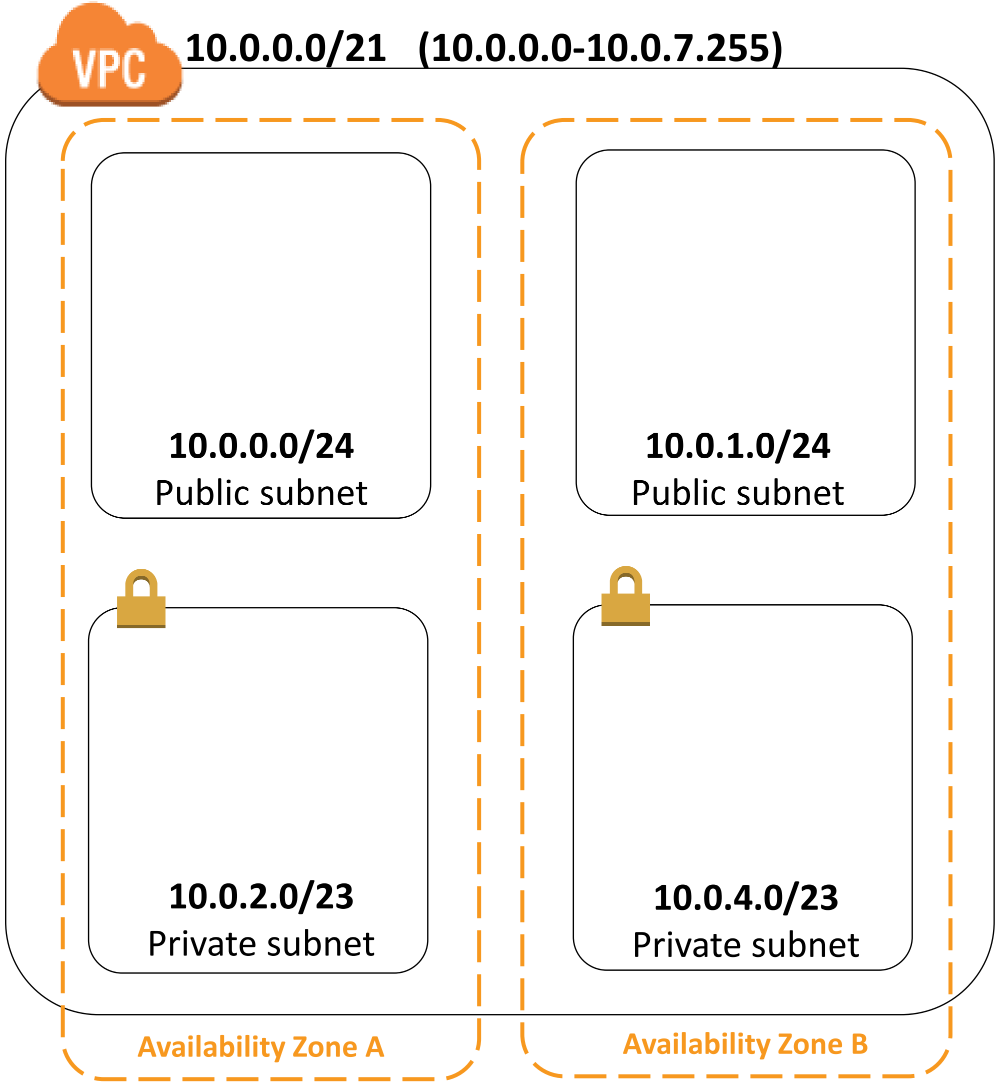
Start with **one public** and **one private** subnet per Availability Zone.



Subnets: Part II

Recommendation:

Allocate substantially **more IP addresses for private subnets** than for public subnets.



Subnet Sizes

Recommendation:

Consider larger subnets over smaller ones (/24 and larger).

Simplifies workload placement:

- Choosing where to place a workload among 10 small subnets is more complicated than with one large subnet.

Less likely to waste or run out of IP addresses:

- If your subnet runs out of available IP addresses, you can't add more to that subnet.
  - Ex.: If you have 251 IP addresses in a subnet that's using only 25, you can't share the unused 226 IP addresses with another subnet that's running out.

Select the Subnet Types

Which subnet type (public or private) should you use for these resources?

Data store instances

- **Private**

Batch processing instances

- **Private**

Backend instances

- **Private**

Web application instances

- **Public or private***

Part 5: Default VPCs and Default Subnets

What is a Default VPC?

Details about default VPCs:

- **Each Region** in your account has a default VPC.
- Default CIDR is **172.31.0.0/16.**
- If you create a VPC-based resource (Amazon EC2, Amazon RDS, Elastic Load Balancing, etc.) but **don't specify a custom VPC**, it will be placed in your default VPC in that region.
- Includes a default **subnet**, **IGW**, main **route table** connecting default subnet to the IGW, default **security group**, and default **NACL**.
- **Configurable** the same as other VPCs; e.g., adding more subnets.

What Is a Default Subnet?

Default subnets in default VPCs:

- Created **within each Availability Zone** for each default VPC.
- **Public** subnet with a CIDR block of **/20** (4,096 IP addresses).
- You can convert it (and any public subnet) into a **private** subnet by removing its route to the IGW.
- When a new Availability Zone is added to a region, your default VPC in that region gets a subnet placed in the new Availability Zone (unless you've made modifications to that VPC).

Default VPCs and Subnets

Recommendation: Use default VPCs and their subnets only for experimenting in your AWS account.

- Default VPCs are a quick start solution.
  - They provide an easy way to test launching instances of your VPC-based resources, without having to set up a new VPC.
- For real-world applications, create your own VPCs and subnets.
  - You'll have greater control and knowledge of their configurations.
  - Possible to re-establish default VPC if accidentally deleted.