AWS Storage Services

Block vs File vs Object Storage

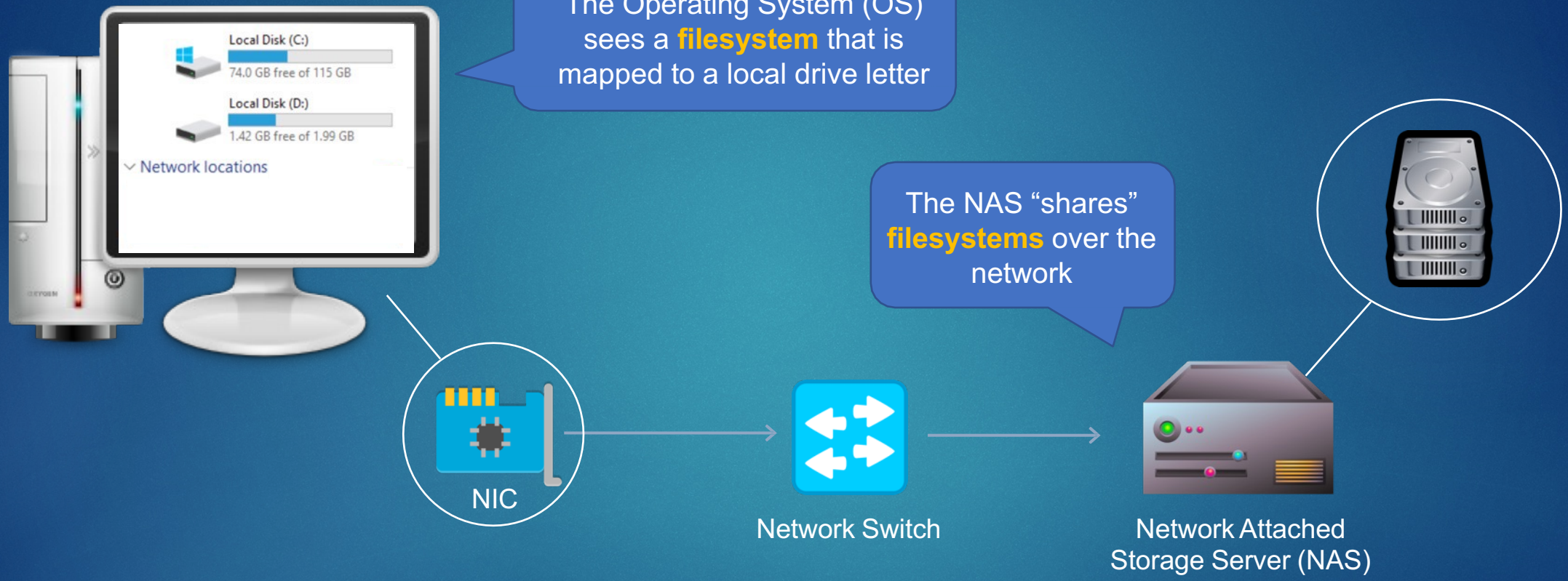
Hard Drives

Hard drives are **block-based** storage systems

The Operating System (OS) can be used to create **volumes**. A volume can be partitioned and formatted

Hard drives are block-based storage systems

Network Attached Storage



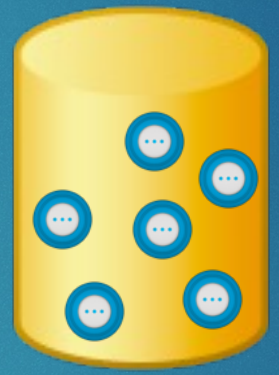
NAS devices are file-based storage systems

Object Storage Systems

User uploads **objects** using a **web browser**

There is no **hierarchy** of objects in the container

The **HTTP protocol** is used with a REST API (e.g. GET, PUT, POST, SELECT, DELETE)



Object Storage Container

Objects can be files, videos, images etc.

Block, File, and Object Storage

The OS sees **volumes** that can be partitioned and formatted

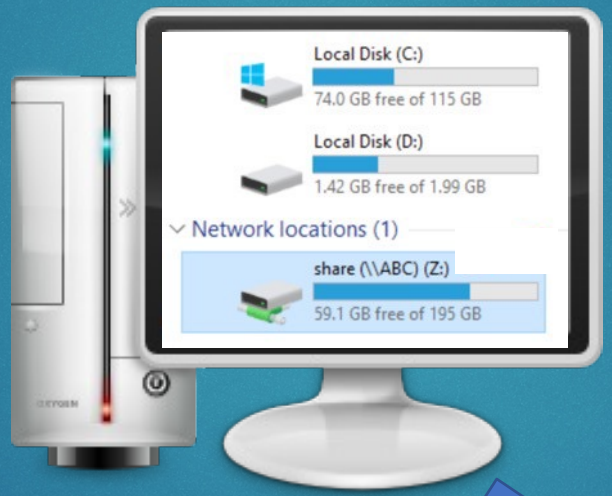
A **filesystem** can be shared by many users/computers

Massively scalable, low cost

Block Storage

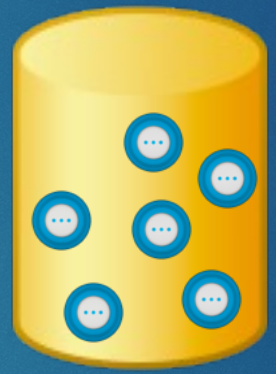
The OS reads/writes at the **block level**. Disks can be internal, or network attached

File Storage



A **filesystem** is "mounted" to the OS using a **network share**

Object Storage



There is **no hierarchy** of objects in the container

Object Storage Container

Uses a **REST API**

AWS Storage Services

Block Storage

Amazon Elastic Block Store

File Storage

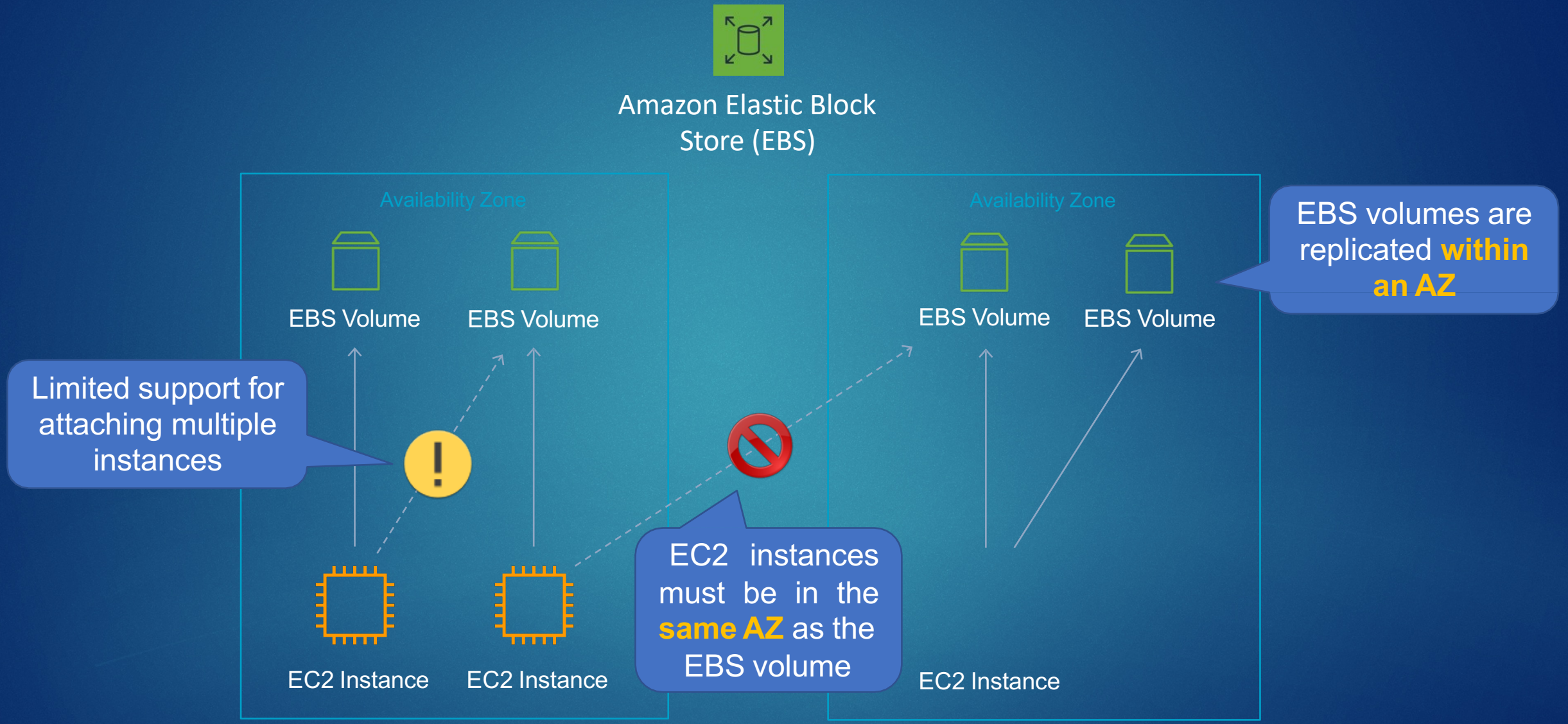
Amazon Elastic File System

Object Storage

Amazon Simple Storage Service (S3)

Amazon Elastic Block Store (EBS)

Amazon EBS


Amazon Elastic Block Store (EBS)
Availability Zone
EBS Volume
EBS Volume
EC2 Instance
EC2 Instance
Availability Zone
EBS Volume
EBS Volume
EC2 Instance
Limited support for attaching multiple instances
EC2 instances must be in the same AZ as the EBS volume
EBS volumes are replicated within an AZ

Amazon EBS SSD-Backed Volumes

New and **not** on the exam yet

	General Purpose SSD		Provisioned IOPS SSD		
Volume type	gp3	gp2	io2 Block Express ‡	io2	io1
Durability	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.999% durability (0.001% annual failure rate)		99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)
Use cases	• Low-latency interactive apps • Development and test environments		Workloads that require sub-millisecond latency, and sustained IOPS performance or more than 64,000 IOPS or 1,000 MiB/s of throughput	• Workloads that require sustained IOPS performance or more than 16,000 IOPS • I/O-intensive database workloads	
Volume size	1 GiB - 16 TiB		4 GiB - 64 TiB	4 GiB - 16 TiB	
Max IOPS per volume (16 KiB I/O)	16,000		256,000	64,000 †	
Max throughput per volume	1,000 MiB/s	250 MiB/s *	4,000 MiB/s	1,000 MiB/s †	
Amazon EBS Multi-attach	Not supported		Not supported	Supported	
Boot volume	Supported				

New and **not** on the exam yet

Amazon EBS HDD-Backed Volumes

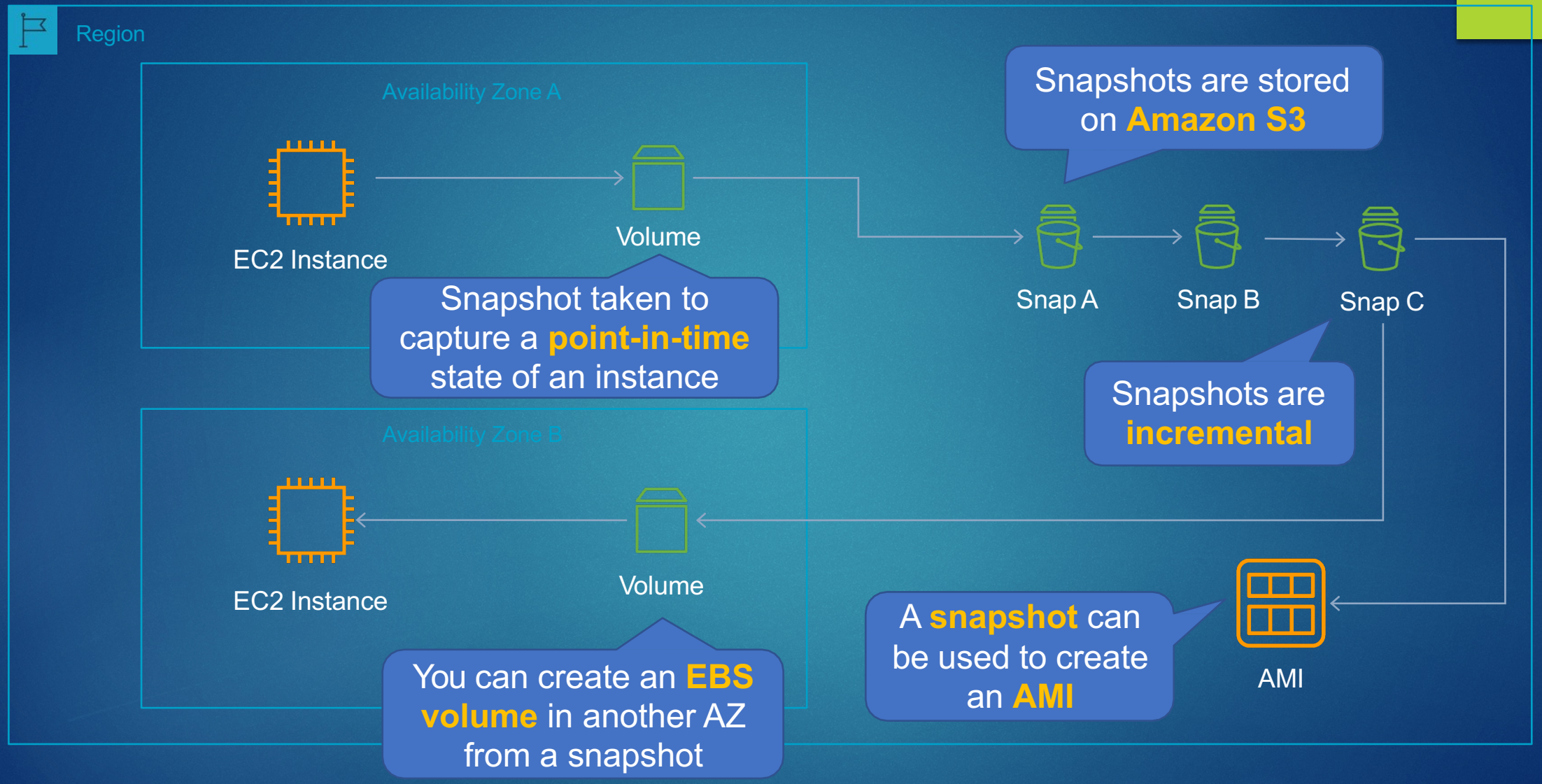
	Throughput Optimized HDD	Cold HDD
Volume type	`st1`	`sc1`
Durability	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)	99.8% - 99.9% durability (0.1% - 0.2% annual failure rate)
Use cases	• Big data • Data warehouses • Log processing	• Throughput-oriented storage for data that is infrequently accessed • Scenarios where the lowest storage cost is important
Volume size	125 GiB - 16 TiB	125 GiB - 16 TiB
Max IOPS per volume (1 MiB I/O)	500	250
Max throughput per volume	500 MiB/s	250 MiB/s
Amazon EBS Multi-attach	Not supported	Not supported
Boot volume	Not supported	Not supported

Amazon EBS

- EBS volume data persists **independently** of the life of the instance
- EBS volumes do not need to be attached to an instance
- You can attach multiple EBS volumes to an instance
- You can use multi-attach to attach a volume to multiple instances but with some constraints
- EBS volumes must be in the **same AZ** as the instances they are attached to
- Root EBS volumes **are deleted** on termination by default
- Extra non-boot volumes **are not deleted** on termination by default

Amazon EBS Snapshots and DLM

Amazon EBS Snapshots



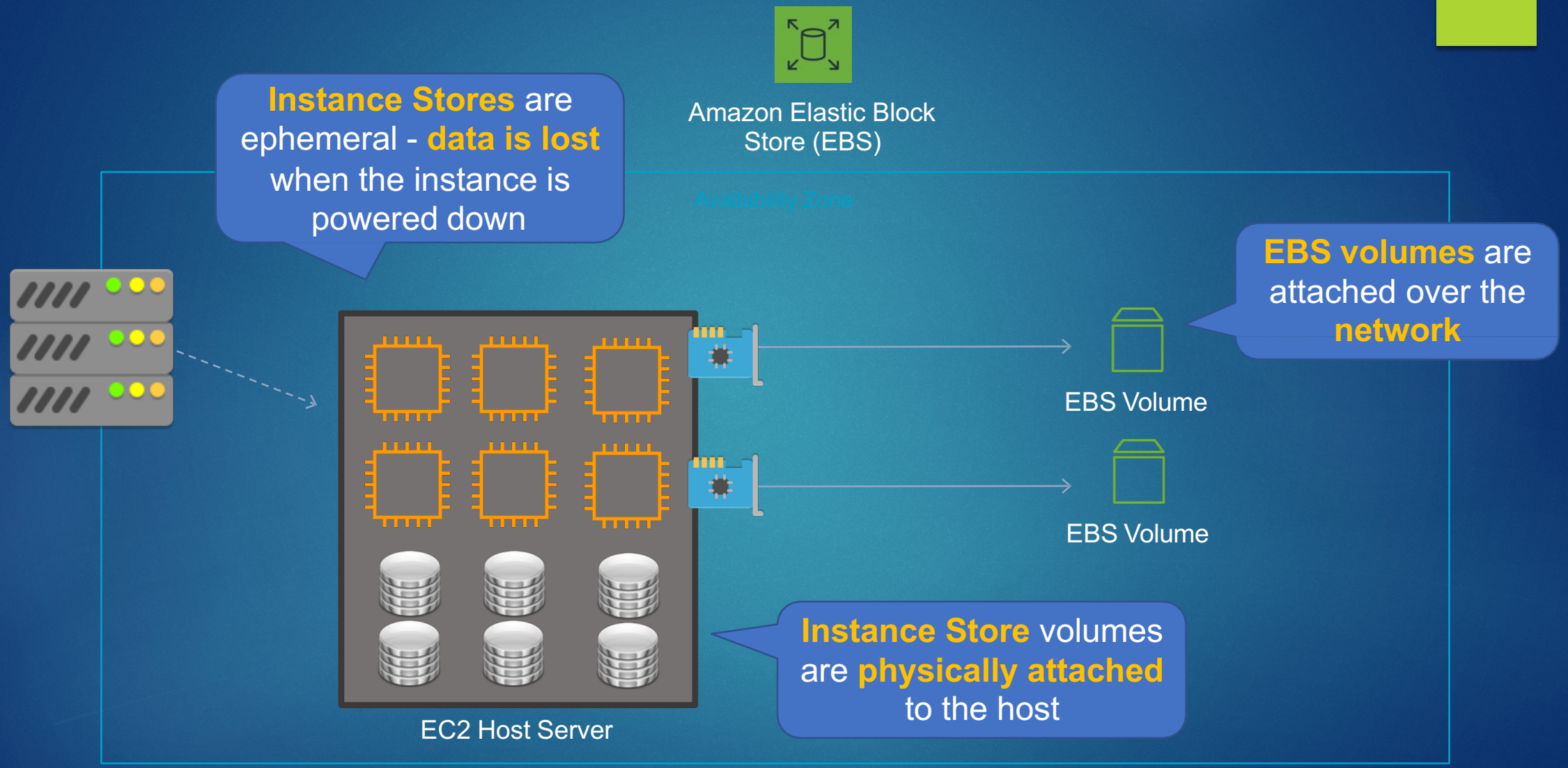
Amazon Data Lifecycle Manager (DLM)

- DLM automates the creation, retention, and deletion of EBS snapshots and EBS-backed AMIs

- DLM helps with the following:
  - Protects valuable data by enforcing a regular backup schedule
  - Create standardized AMIs that can be refreshed at regular intervals
  - Retain backups as required by auditors or internal compliance
  - Reduce storage costs by deleting outdated backups
  - Create disaster recovery backup policies that back up data to isolated accounts

EC2 Instance Store Volumes

EBS vs instance store



EBS Volumes and Snapshots

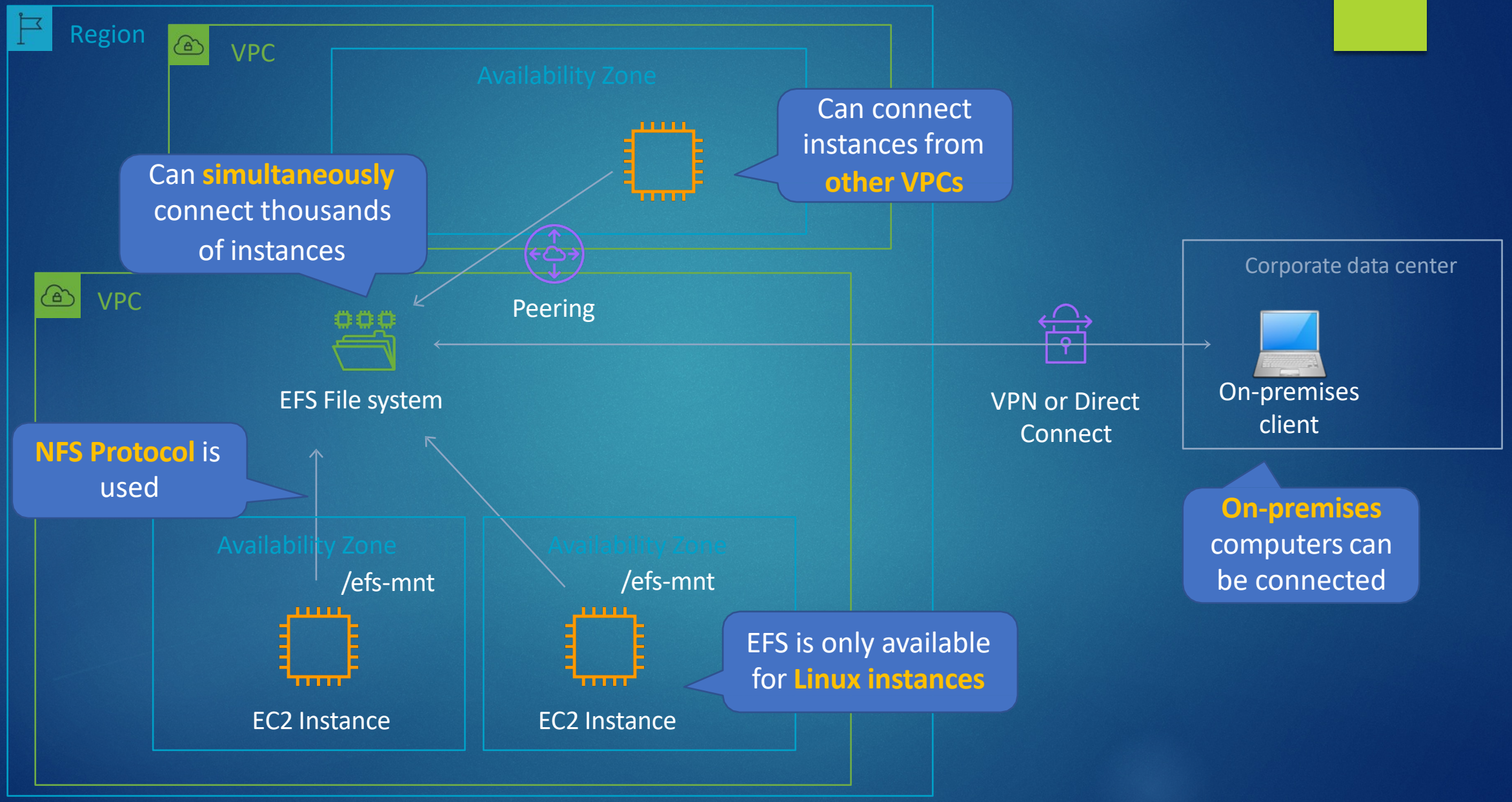
Amazon Machine Images (AMI)

Amazon Machine Images (AMIs)

- An **Amazon Machine Image** (AMI) provides the information required to launch an instance
- An AMI includes the following:
  - One or more EBS snapshots, or, for instance-store-backed AMIs, a template for the root volume of the instance (for example, an operating system, an application server, and applications)
  - Launch permissions that control which AWS accounts can use the AMI to launch instances
  - A block device mapping that specifies the volumes to attach to the instance when it's launched
- AMIs come in three main categories:
  - **Community AMIs** - free to use, generally you just select the operating system you want
  - **AWS Marketplace AMIs** - pay to use, generally come packaged with additional, licensed software
  - **My AMIs** - AMIs that you create yourself

Amazon Elastic File System (EFS)

Amazon EFS


Region
VPC
Availability Zone
Can connect instances from other VPCs
Can simultaneously connect thousands of instances
VPC
Peering
Corporate data center
EFS File system
VPN or Direct Connect
On-premises client
NFS Protocol is used
On-premises computers can be connected
Availability Zone
/efs-mnt
Availability Zone
/efs-mnt
EFS is only available for Linux instances
EC2 Instance
EC2 Instance

Amazon Simple Storage Service (S3)

Amazon S3

Bucket

A **bucket** is a container for objects

http://*bucket*.s3.*aws-region*.amazonaws.com

http://s3.*aws-region*.amazonaws.com/*bucket*

Object

An objects consists of:

- Key (name of objects)
- Version ID
- Value (actual data)
- Metadata
- Subresources
- Access control information

Amazon S3

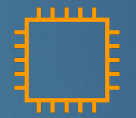
EC2 instances connect using **private** addresses

Private Connection

S3 Gateway Endpoint

VPC

Public subnet



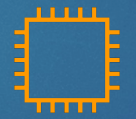
EC2 Instance

Internet gateway

Public Internet

EC2 instances connect using **public** addresses

Private subnet



EC2 Instance

Internet Client

Amazon S3

- You can store any type of file in S3
- Files can be anywhere from 0 bytes to 5 TB
- There is unlimited storage available
- S3 is a universal namespace so bucket names must be unique globally
- However, you create your buckets within a REGION
- It is a best practice to create buckets in regions that are physically closest to your users to reduce latency

Amazon S3 – Additional Features

S3 Capability	What it Does
Transfer Acceleration	Speed up data uploads using CloudFront in reverse
Requester Pays	The requester rather than the bucket owner pays for requests and data transfer
Events	Trigger notifications to SNS, SQS, or Lambda when certain events happen in your bucket
Static Web Hosting	Simple and massively scalable static website hosting
Versioning and Replication	Retain versions of objects and replicate objects within and across AWS Regions

Amazon S3 Storage Classes

Amazon S3 Availability and Durability

- Durability

- Measures the likelihood of data loss
- All storage classes offer 99.999999999% durability
- This means that if you store 100 billion objects in S3, you

- will lose one object at most

Availability

- Measures how readily available the service is
- Measured as a percentage
- S3 availability SLA varies between storage classes

Amazon S3 Storage Classes

	S3 Standard	S3 Intelligent Tiering	S3 Standard-IA	S3 One Zone-IA	S3 Glacier	S3 Glacier Deep Archive
Designed for durability	99.999999999%	99.999999999%	99.999999999%	99.999999999%	99.999999999%	99.999999999%
Designed for availability	99.99%	99.9%	99.9%	99.5%	99.99%	99.99%
Availability SLA	99.9%	99%	99%	99%	99.9%	99.9%
Availability Zones	3	3	3	1	3	3
Minimum capacity charge per object	N/A	N/A	128KB	128KB	40KB	40KB
Minimum storage duration charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval fee	N/A	N/A	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved
First byte latency	milliseconds	milliseconds	milliseconds	milliseconds	select minutes or hours	select hours
Storage type	Object	Object	Object	Object	Object	Object
Lifecycle transitions	Yes	Yes	Yes	Yes	Yes	Yes

Create Amazon S3 Bucket

S3 Versioning, Replication and Lifecycle Rules

Amazon S3 Versioning

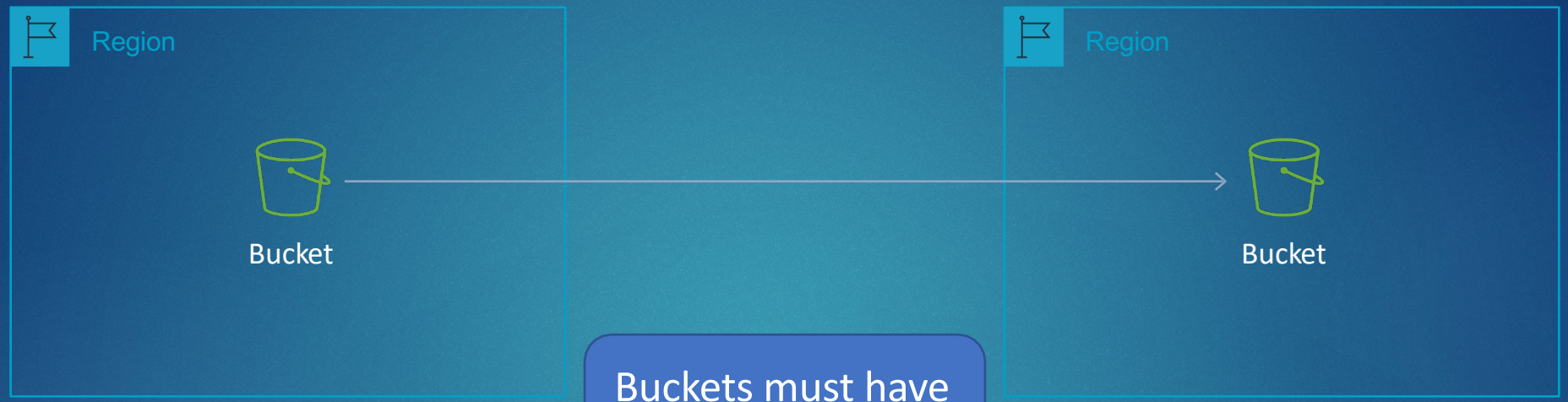
- Versioning is a means of keeping **multiple variants** of an **object** in the same bucket

- Use versioning to preserve, retrieve, and restore every version of every object stored in your Amazon S3 bucket

- Versioning-enabled buckets enable you to recover objects from accidental deletion or overwrite

Amazon S3 Replication

Cross-Region Replication (CRR)



Buckets must have **versioning** enabled

Configure Replication and Lifecycle

Configure S3 Static Website

S3 Permissions and Bucket Policies

Archiving with S3 Glacier

Amazon S3 Glacier

- Extremely low cost and you pay only for what you need with no commitments of upfront fees
- Two classes **Glacier** and **Glacier Deep Archive**
- Three options for access to archives, listed in the table below:

	Expedited	Standard	Bulk
Data access time (Glacier)	1-5 minutes	3-5 hours	5-12 hours
Data access time (Deep Archive)	N/A	12 hours	48 hours

Object Lock and Glacier Vault Lock

S3 Object Lock

- Store objects using a write-once-read-many (WORM) model
- Prevent objects from being deleted or overwritten for a fixed time or indefinitely

S3 Glacier Vault Lock

- Also used to enforce a WORM model
- Can apply a policy and lock the policy from future edits
- Use for compliance objectives and data retention

AWS Storage Gateway

AWS Storage Gateway

- Hybrid cloud storage service
- Access cloud storage from on-premises applications
- Enables access to proprietary object storage (S3) using standard protocols
- Use cases:
  - Moving backups to the cloud
  - Using on-premises file shares backed by cloud storage
  - Low latency access to data in AWS for on-premises applications
  - Disaster recovery

AWS Storage Gateway