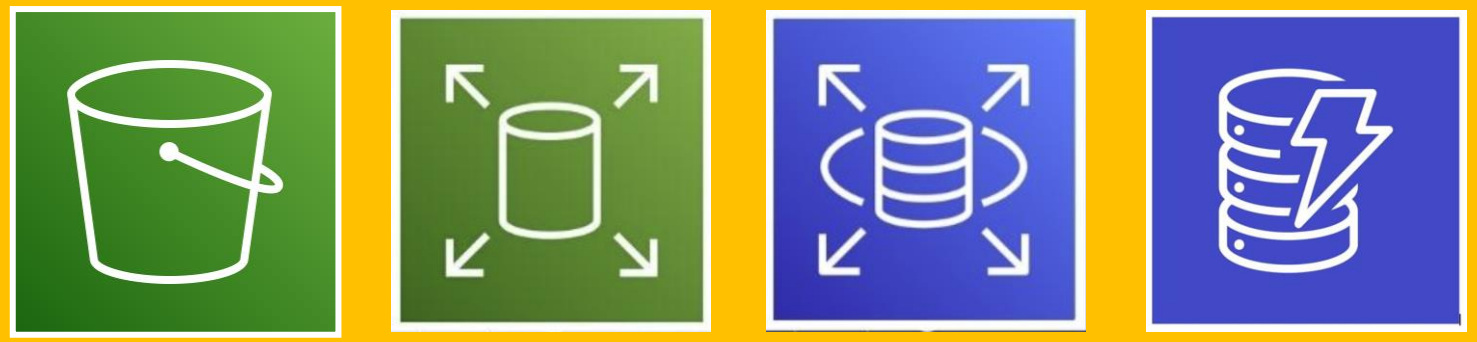
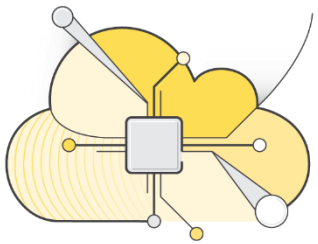
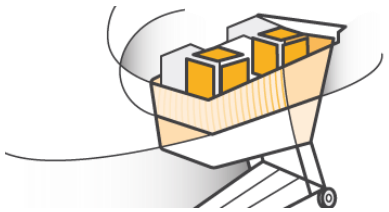
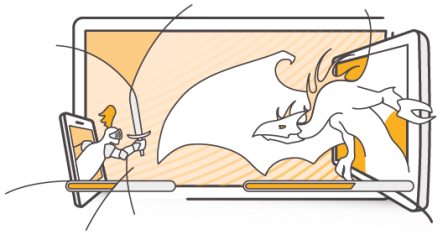
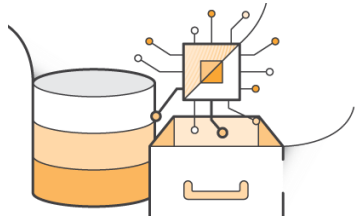
Week 3 – Storage and Databases on AWS







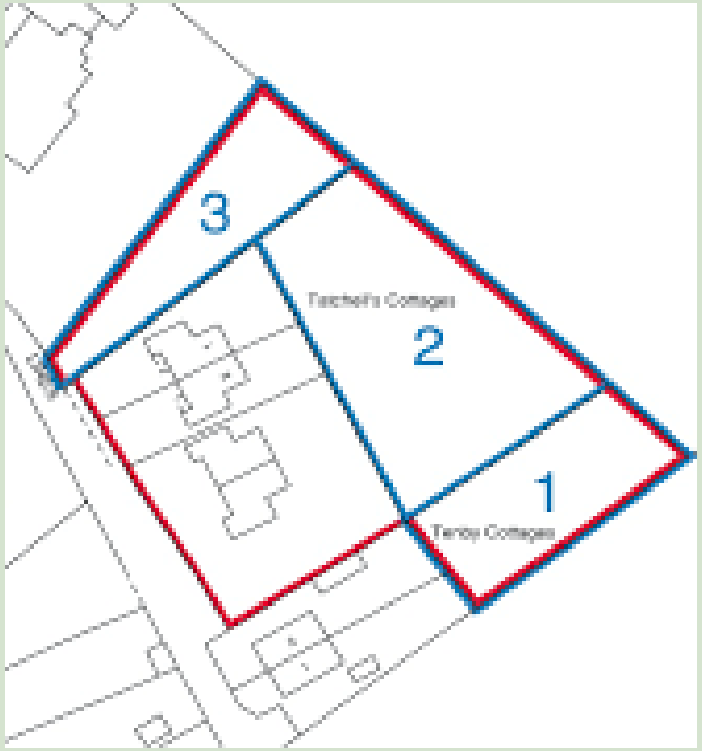


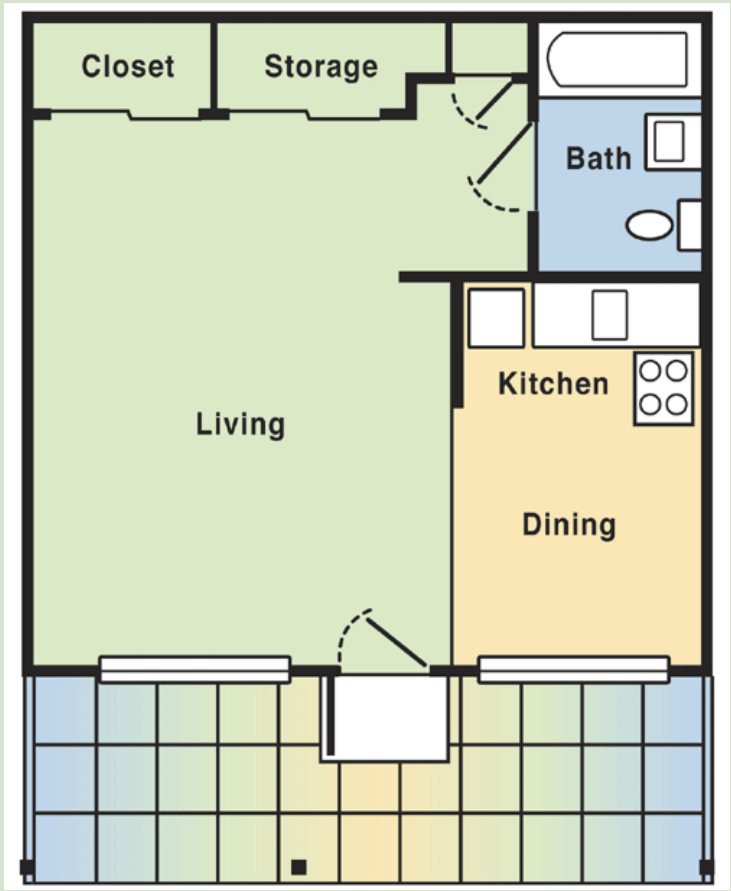


Understanding Storage on AWS

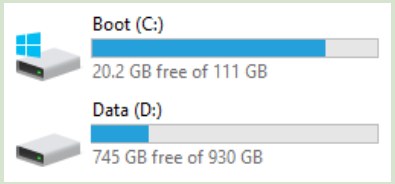
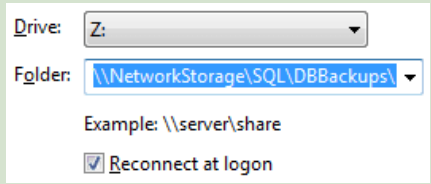
Storage Types

Block Storage	File Storage	Object Storage
A Piece of Land	An Apartment	Storage Unit

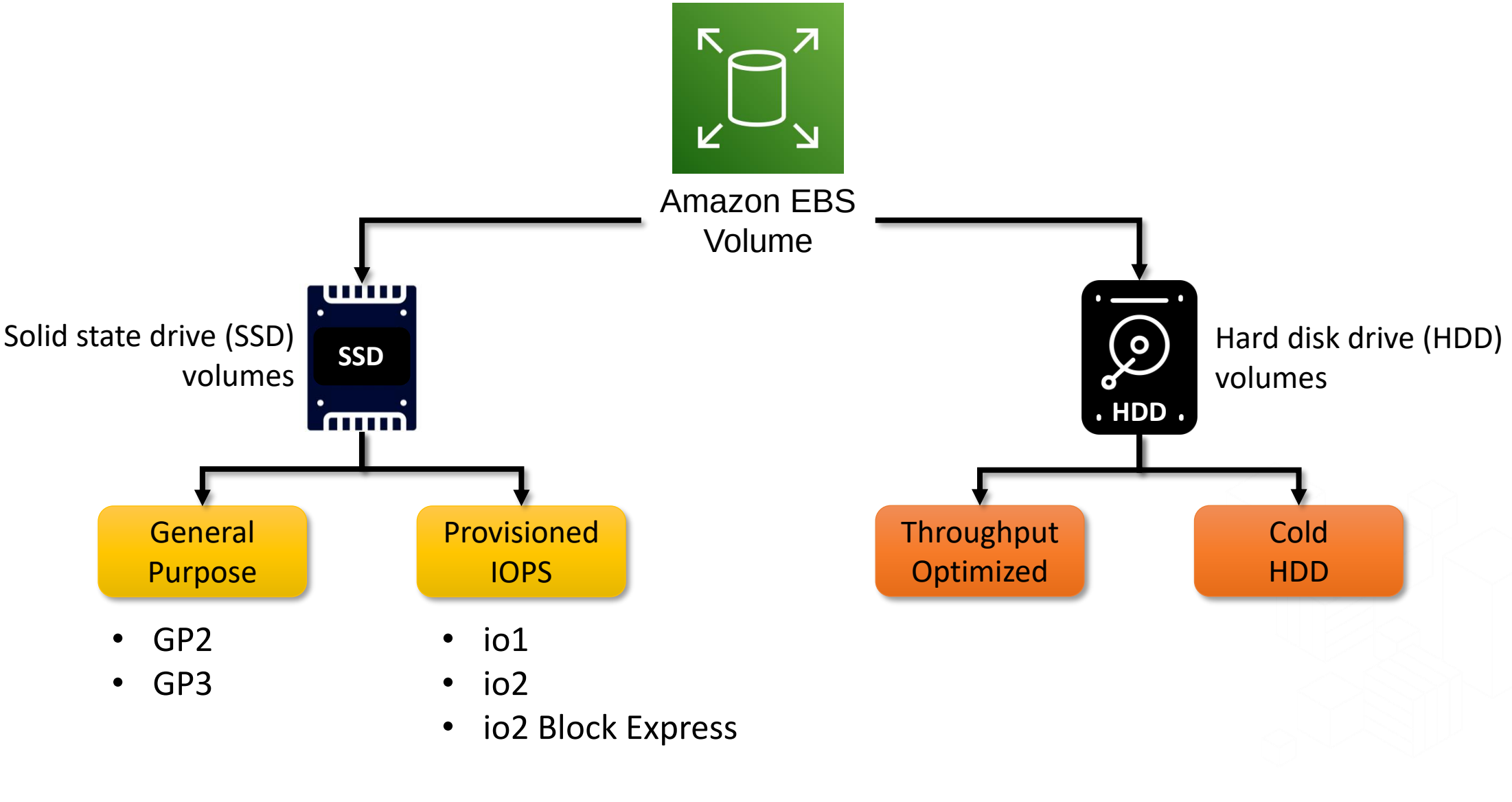




Storage Types

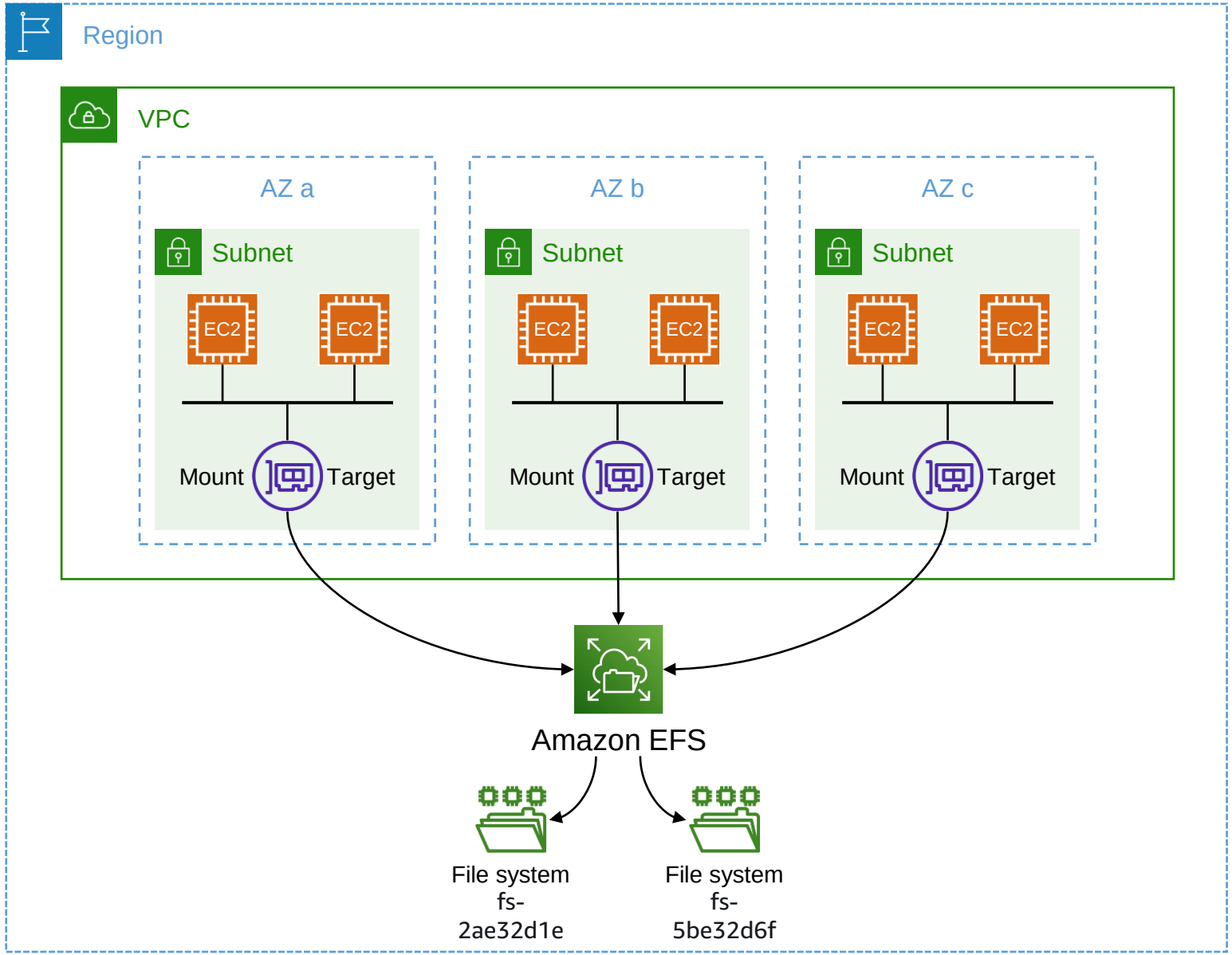
	Block Storage	File Storage	Object Storage
Unit of Transaction	Blocks	Files	Objects (files with metadata)
Example	Laptop Disk 	Windows Share 	OneDrive / Google Drive / Dropbox 
How can you update?	You can directly update the file	You can directly update the file	You cannot update the object directly. You create a new version of the object and replace the existing one or keep multiple versions of the same object
Protocols	SCSI, Fiber Channel, SATA	SMB, CIFS, NFS	REST/SOAP over HTTP/HTTPs
Support for metadata	No metadata support it stores only file system attributes	No metadata support it stores only file system attributes	Supports custom metadata
AWS Services	Amazon EBS Amazon Instance Store	Amazon EFS Amazon FSX	Amazon S3 Amazon Glacier

• Amazon EBS Volume Types



Amazon EFS

Amazon EFS Architecture



- Fully managed
  - Highly available and durable
- Dynamic elasticity
  - Grow/Shrink
- Storage classes and lifecycle management
- NFS v4.0 and v4.1
- Can be accessed across
  - VPC
  - Regions
  - Accounts
  - On-Prem

Amazon FSx

Amazon FSx

Amazon FSx for
Windows File Server

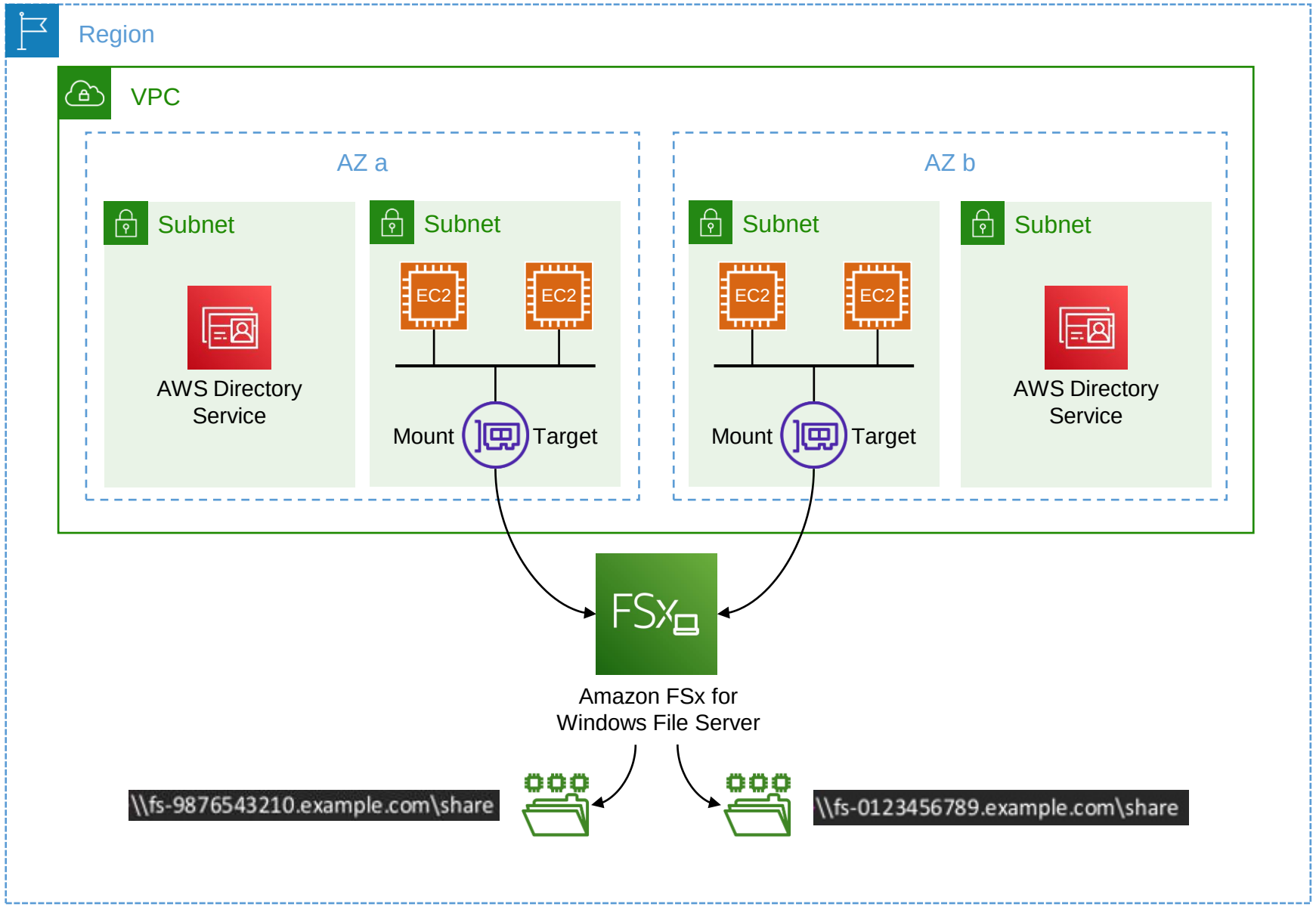
Fully managed file storage built on Windows Server

Amazon FSx
for Lustre

Fast and scalable shared storage to power your HPC workloads

- Use cases
  - Home directories, Highly available Microsoft SQL Server deployments, Lift-and-shift Windows applications
- Built on Windows Server
  - integrates with your on-premises Microsoft Active Directory (AD) as well as with AWS Microsoft Managed AD.
- Fully managed
  - AWS manages updates and patches, failover and failback, backups

Amazon FSx for Windows Architecture



- Multi-AZ availability and durability
- SMB 2.0 to 3.1.1
- DFS Namespace and Replication
- Can be accessed across
  - VPC
  - Regions
  - Accounts
  - On-Prem

Amazon Simple Storage Service (S3)

Bucket and Objects

- S3 Name Space – Global
- Bucket – Regional
- Durability – 99.999999999 %
- Availability – 99.9x %
- Max object size 5 terabytes
- Can host a static website

Files (Objects)

Objects

Bucket

S3 Pricing	
Compute	Number of requests
Storage	Capacity used
Network	Data transfer out

S3 Features

Versioning

Replication

Encryption

S3 Storage Class

- Amazon S3 offers a range of storage classes designed for different use cases.
  - https://aws.amazon.com/s3/storage-classes/

S3 Glacier

- A storage service optimized for infrequently used data, or "cold data"
- Suitable for data archiving and backup
- Archive Retrieval Options

Retrieval Mode	Retrieval Time	Retrieval Cost
Expedited	1 – 5 minutes	$$$
Standard	3 – 5 hours	$$
Bulk	5 – 12 hours	$

S3	Glacier
Bucket	Vault
Object	Archive
Object Size Max 5 TB	Object Max 40 TB

AWS Snow Family

- Physical Device for Data Transfer

Snowcone

8 TB

Snowball

50/80 TB

Snowmobile

100 PB

Additional Resources

- Bucket Policy Examples
  - https://docs.aws.amazon.com/AmazonS3/latest/userguide/example-bucket-policies.html
- Amazon S3 Charges
  - https://aws.amazon.com/s3/pricing/
- Amazon S3 Multipart Upload
  - https://docs.aws.amazon.com/AmazonS3/latest/userguide/mpuoverview.html
- Amazon S3 Storage Classes
  - https://aws.amazon.com/s3/storage-classes/
- Amazon S3 Glacier
  - https://aws.amazon.com/glacier/
- Amazon S3 Life Cycle Policies
  - https://docs.aws.amazon.com/AmazonS3/latest/userguide/object-lifecycle-mgmt.html
- Amazon Snow Family
  - https://aws.amazon.com/snow/

Types of Databases

SQL vs. NoSQL Databases

	SQL (Optimized for Storage)	NoSQL (Optimized for performance)
Data Storage	Rows and Columns	Key-value, document, wide-column, graph
Schemas	Fixed	Dynamic
Querying	Using SQL	Focused on collection of documents
Scalability	Vertical	Horizontal
Transactions	Supported	Support varies

A — **Atomicity**
Transitions are all or nothing

C — **Consistency**
Only valid data is saved

I — **Isolation**
Transactions do not affect each other

D — **Durability**
Written data won't be lost

BA — **Basically Available**
System does guarantee availability

S — **Soft state**
System may change over time

E — **Eventual consistency**
system will become consistent over time

Databases on AWS

Databases on AWS – Relational DB – RDS

RDS or Relational Database Service is an AWS offering which makes it easy to setup, operate and maintain relational Database in the cloud environment. RDS is a managed AWS service which helps you with management of time-consuming DB tasks such as:

- ✓ Backup and restore
- ✓ Scalability
- ✓ High availability
- ✓ Patching and upgrade

Databases on AWS – Relational DB – RDS

RDS supports the following DB engines:

- ✓ Amazon Aurora
- ✓ MySQL
- ✓ Oracle,
- ✓ MS SQL Server
- ✓ PostgreSQL
- ✓ MariaDB
- ✓ IBM Db2

Databases on AWS – Non-Relational Databases

AWS supports all major non-relational database workloads such as Key-value, in-memory, document etc.

Important Nonrelational DB services:

- ✓ Amazon DynamoDB
- ✓ Amazon ElastiCache
  - ✓ Amazon ElastiCache for Redis
  - ✓ Amazon ElastiCache for Memcached
- ✓ Amazon DocumentDB (with MongoDB compatibility)
- ✓ Amazon Keyspaces (for Apache Cassandra)

Databases on AWS – Other Purpose-built Databases

Apart from the Relational and Non-relational Database engines, AWS also supports the following purpose built, fully managed Databases:

- ✓ Amazon Neptune
- ✓ Amazon Redshift
- ✓ Amazon Timestream
- ✓ Amazon QLDB

Databases on AWS – Summary

Database type	Use cases	AWS service
Relational	Traditional applications, ERP, CRM, e-commerce	Amazon Aurora, Amazon RDS, Amazon Redshift
Key-value	High-traffic web apps, e-commerce systems, gaming applications	Amazon DynamoDB
In-memory	Caching, session management, gaming leaderboards, geospatial applications	Amazon ElastiCache for Memcached Amazon ElastiCache for Redis
Document	Content management, catalogs, user profiles	Amazon DocumentDB (with MongoDB compatibility)
Wide column	High scale industrial apps for equipment maintenance, fleet management, and route optimization	Amazon Keyspaces (for Apache Cassandra)
Graph	Fraud detection, social networking, recommendation engines	Amazon Neptune
Time series	IoT applications, DevOps, industrial telemetry	Amazon Timestream
Ledger	Systems of record, supply chain, registrations, banking transactions	Amazon QLDB

https://aws.amazon.com/products/databases/

Additional Resources

- Amazon Databases [Product Page]
  - https://aws.amazon.com/products/databases/
- Getting started tutorials [Databases]
  - https://aws.amazon.com/products/databases/learn/#Getting_started_tutorials
- Purpose-built databases on AWS
  - https://aws.amazon.com/blogs/publicsector/purpose-built-databases-model-building-applications-cloud/
- AWS DynamoDB Features
  - https://aws.amazon.com/dynamodb/features/
- AWS Purpose-built Databases workshop
  - https://catalog.us-east-1.prod.workshops.aws/workshops/93f64257-52be-4c12-a95b-c0a1ff3b7e2b/en-US

Questions?