Day 6/13

Module 6 : Storage

#1 Introduction

Block storage
→ Data divided into pieces called blocks
→ Direct data access without file system layers
→ Best for applications/ databases needing fast frequent updates.
EX- Amazon EC2 instance store, Amazon Elastic Block Store (EBS)

Object Storage
→ object = data + unique ID + metadata
→ Full rewrite required to update an object
→ organized using buckets.
→ Best for large or infrequently changed files
Ex- Amazon Simple Storage Service (S3)

File storage
→ Cloud-based access through shared file systems
→ Straightforward implementation without code changes
→ Best for applications needing shared file access
Ex- Amazon Elastic File System (EFS)
Amazon FSx

Block storage

Date
DELTA Pg No

#2 EC2 Instance Store and Amazon EBS

EBS (Elastic Block Store) → (External hard drive)

EBS is like a hard disk attached to your EC2 instance.

EC2 = computer

EBS = Permanent hard drive of that computer.

→ Data is persistent (saved permanently)

→ If you stop the EC2 instance → data remains.

→ You can take the snapshots (backups)

→ Used for production workloads.

EX- Databases, App server, Important business data

EC2 Instance Store

Instance store is temporary storage directly attached to the physical machine running the EC2 instance.

EC2 = Computer

Instance Store = Temporary storage space

→ Data is not persistent

→ If you stop or terminate the instance - data is lost

→ Very fast storage

→ No backup option
Ex- Cache, Temporary files, scratch Data

#3 AWS EBS Data Lifecycle

Amazon Data Lifecycle Manager
→ Schedule automatic Snapshot creation
→ set retention policies
→ Manage Snapshot lifecycle
→ Apply consistent backup policies

Root volume :- The root volume is the main disk where the OS is installed like C Drive.

Data volume - A data volume is an additional disk attached to the instance for storing application or user data. like D and other drive

Snapshots - EBS Snapshots are point-in-time backups of EBS volume. They are increamental

Increamental backup is a backup method that saves only the data that has changed since the last backup, reducing storage space and backup time.

Amazon Machine Image (AMI) is a template used to launch EC2 instances with a pre-installed operating system and configuration

Backup role: It helps recreate the same instance setup if needed.

Object storage

#4 Amazon Simple Storage Service (S3)

Amazon S3 :- Store and retrieve an unlimited amount of data.

→ Store data as objects
→ Store objects in buckets
→ Upload a maximum object size of 5 TB
→ Create Multiple buckets
→ Version objects.

99.999999999% data durability by automatically storing multiple copies of data across at least three physically separate facilities within an AZ.

Amazon S3 security
→ Private access by default
→ Bucket policies
→ Presigned URLs for temporary access.
→ Amazon S3 access point for access policies.
→ Amazon S3 audit logs for tracking

S3 buckets: An S3 bucket is a container for storing objects in Amazon S3

Note: S3 Block public access setting override bucket policies, preventing public access even when policies allow it.

5 Amazon S3 Storage classes and S3 Lifecycle
→ Designed for different storage needs
→ Multiple storage classes, single bucket.

(i) Amazon S3 Standard - Frequent access, high performance

(ii) S3 Intelligent-Tiering - Automatically moves data to cheaper tiers based on usage.

(iii) S3 Standard-IA - Infrequent access, low cost, quick retrieval.

(iv) S3 One zone-IA - infrequent access, stored in one AZ (cheaper, less resilient).

(v) S3 Glacier Instant Retrieval - Rare access, instant retrieval, low cost.

(vi) S3 Glacier Flexible Retrieval - Archive storage, retrieval in minutes to hours.

(vii) S3 Glacier Deep Archive - Lowest cost, retrieval in hours [long-term archive]

(viii) ~~S3 Glacier Deep Archive~~

(viii) S3 Express one zone - Very High performance single AZ, low latency.

(ix) S3 outposts - S3 Storage on premises with AWS outposts.

S3 Lifecycle is a feature that automatically manages your objects over time to reduce storage cost.

It allows you to:

- Transition objects to cheaper storage classes after a certain number of days

Date / / DELTA Pg No.

Ex- Standard → Standard-IA → Glacier
→ Expire objects and permanently delete them after a defined period.

#6 Amazon Elastic File system (EFS)

→ Multiple instances can access the data in EFS at the same time.

Amazon EBS	Amazon EFS
→ volumes attach to EC2 instances	→ Multiple instances reading and writing simultaneously.
→ Availability zone level resource	→ Linux file system
→ Need to be in the same AZ to attach EC2 instance	→ Regional resource
→ volumes do not automatically scale	→ Automatically scales

Storage Class

Standard (Multi AZ	One zone (single AZ)	Archive
→ Stored across multiple AZ	→ Stored in one AZ only	→ Lowest cost
→ High durability & availability	→ cheaper	→ For rarely accessed data
→ For production workloads.	→ Less resilient.	→ used with lifecycle

Date
DELTA Pg No

EFS Lifecycle Transition

(1) Transition to IA
Standard → standard IA or One zone → One zone IA

(2) Transition to Archive
IA → Archive

(3) Transition to Standard
When archived or IA data is accessed, it moves back to standard automatically.

#7 Amazon FSx
Amazon FSx is a fully managed file system service that provides specialized file system for specific workloads.

EFS → general purpose shared storage
FSx is workload-specific file storage.

Types:
(1) FSx for windows File server
→ Fully managed windows file system
→ Uses SMB protocol
→ Integrated with Active Directory

(II) FSx for lustre
→ High performance storage
→ Used in ML, HPC, Big data.

(III) FSx for NetApp ONTAP
→ Enterprise features.
→ Snapshots, replication, hybrid cloud.

(IV) FSx for OpenZFS
• Linux based workloads
• ZFS features like compression & snapshots.

#8 AWS storage Gateway

AWS storage gateway is a hybrid storage service that connects your on-premises environment to AWS cloud storage.

Types:

(I) File Gateway:
→ Stores files in Amazon S3
→ Uses NFS or SMB protocol
→ Appears like a local file share.

(II) Volume Gateway
→ provides block storage
→ Backed by Amazon S3

→ Snapshots in EBS

Two modes

- Cached volumes (most data in S3)
- Stored volumes (most data on-prem)

Usecases:- Backup + disaster recovery

(iii) Tape Gateway

→ Virtual tape Library

→ Stores data in S3 Glacier

→ Replaces physical tape backup.

AWS Elastic Data Recovery (DRS) is a service that continuously replicates on-premises or cloud servers to AWS so they can be quickly recovered during a disaster.