Module 7: Storage

Monday, September 15, 2025 1:03 PM

A. Learning Outcomes (Los) and Topics

- Topics:
  + Amazon Elastic Block Store (Amazon EBS)
  + Amazon Simple Storage Service
  + Amazon Elastic File System (Amazon EFS)
  + Amazon Simple Storage Service Glacier
- Module Objectives:
  + Identify the different types of storage
  + Explain Amazon S3
  + Identify the functionality in Amazon S3
  + Explain Amazon EBS
  + Identify the functionality in Amazon EBS
  + Perform functions in Amazon EBS to build an Amazon EC2 storage solution
  + Explain Amazon EFS
  + Identify the functionality in Amazon EFS
  + Explain Amazon S3 Glacier
  + Identify the functionality in Amazon S3 Glacier
  + Difference between Amazon EBS, Amazon S3, Amazon EFS, and Amazon S3 Glacier

B. Amazon Elastic Block Store (Amazon EBS)

- Amazon EBS provides persistent block stirage volumes for uses in EC2 instances
- AWS Storage options: Block storage versus object storage (What if user wanna change one character in a 1-GB file?)
  + Block storage: Change one block (piece of the file) that contains the character
  + Object storage: Entire file must be updated
- Amazon EBS enables user to create individual storage volumes and attach them to an Amazon EC2 instances:
  + Amazon EBS offers block-level storage
  + Volumes are automatically replicated within its Availability Zone
  + It can be backed up automatically to Amazon S3 through snapshots
  + Uses include:
    * Boot volumes and storage for Amazon Elastic Compute Cloud (Amazon EC2) instances
    * Data storage with a file system.
    * Database hosts
    * Enterprise applications

	Solid State Drives (SSD)		Hard Disk Drives (HDD)	
	General Purpose	Provisioned IOPS	Throughput-Optimized	Cold
Maximum Volume Size	16 TiB	16 TiB	16 TiB	16 TiB
Maximum IOPS/Volume	16,000	64,000	500	250
Maximum Throughput/Volume	250 MiB/s	1,000 MiB/s	500 MiB/s	250 MiB/s

- Amazon EBS:
  + Snapshots:
    * Point-in-time snapshots: mean that capture the exact state of user data so user can use them later to restore a volume to that specific state.
    * Recreate a new volume at any time
  + Encryption:
    * Encrypted Amazon EBS volumes
    * No additional cost
  + Elasticity:
    * Increase capacity
    * Change to different styles
- Amazon EBS: Volumes, IOPS, and pricing
  + Volumes:
    * Amazon EBS volumes persist independently from the instance.
    * All volume types are charged by the amount that is provisioned per month
  + IOPS:
    * General Purpose SSD: Charged by the amount that user provision in GB per month until storage is released
    * Magnetic: Charged by the number of requests to the volume
    * Provisioned IOPS SSD: Charged by the amount that user provision in IOPS (multiplied by the percentage of days that user provision

for the month)
- Amazon EBS: Snapshots and data transfer
  + Snapshots: Added cost of Amazon EBS snapshots to Amazon S3 is per GB-month of data stored
  + Data Transfer:
    * Inbound data transfer is free
    * Outbound data transfer across Regions incurs charges.

C. Amazon Simple Storage Service (Amazon S3)

- Data is stored as objects in buckets
- Virtually unlimited storage: Single object is limited to 5 TB
- Designed for 11 9s of durability
- Granular access to bucket and objects
- Amazon S3 offers a range of object-level storage classes that are designed for different use cases:
  + Amazon S3 Standard:
    * It was designed for high availability, high durability, and performance for freqiemtly accessed data
    * Due to its low latency and high throughput, the Amazon S3 is appropriate for variety of uses including content distributions and big data analysis
  + Amazon S3 Intelligent-Tiering: It is designed for optimize costs by automatically moving data to the most cost-effective access year
  + Amazon S3 Standard-Infrequemt Access (Amazon S3 Standard-IA)
    * It is used for data that is accessed less frequently but requires rapid access when needed
    * It is designed to provide high durability, high throughput, and low latency of Amazon S3 standard with a low per gigabyte storage price and per gigabyte retrieval fee.
    * Due to combination of low cost and high performance, it is suitable for long-term storage and back-ups

NOTE: Other Amazon S3 classes allows to store data in a minimum of three AZ

  + Amazon S3 One Zone-Infrequent Access (Amazon S3 One Zone-IA)
    * Stores data in a single availability zone and costs less than Amazon S3 Standard-Infrequent Access
  + Amazon S3 Glacier: It is secure, durable, and low costs storage class for data archiving
  + Amazon S3 Glacier Deep Archive:
    * the lowest cost storage class
    * Supports long-term retention and digital preservation for data that might be accessed once or twice within a year
    * It is designed ofr highly regulated industries, such as financial services, healthcare, and public sectors
    * It can be used for backup and disaster recovery use cases
- Amazon S3 bucket URLs (2 styles):
  +To upload user data:
    * Create a bucket in an AWS Region
    * Upload almost any number of objects to the bucket.
  + Bucket path-style URL endpoint (https://s3.ap-northeast-1.amazonaws.com/bucket-name): It is normally used when need to access objects
  + Bucket virtual-hosted-style URL endpoint (https://bucket-namr.s3-ap-northeast-1.amazonaws.com): It is used when using bucket as a website for static data.
- Access data anywhere including AWS Management Console, AWS Command Line Interface, and SDK
- Amazon S3 common scenarios:
  + Backup and storage
  + Application hosting
  + Media hosting
  + Software delivery
- Amazon S3 pricing:
  + Pay only for what user use, including GBs per month, Transfer OUT to other Regions, PUT, COPY, LIST, and GET requests.
  + User do not pay for transfer IN to Amazon S3, Transfers OUT from Amazon S3 to Amazon CloudFront or Amazon EC2 in the same Region
- Amazon S3 Storage pricing
  + Requests:
    * The number of requests (GET, PUT, COPY)
    * Type of requests: Different rates of GET requests than other requests
  + Data Transfer:
    * Pricing is based on the amount of data that is transferred out of the Amazon S3 Region
    8 Data transfer in is free, but user incur charges for data that is transferred out

D. Amazon Elastic File System (Amazon EFS)

- Amazon EFS implement storage for EC2 instances that multiple virtual machines can access at the same time.
- It is implemented as a shared file system that uses the Network File System
- Amazon EFS feature:
  + File Storage in the AWS Cloud
  + Works well for big data and analytics, media processing workflows, content management, web serving, and home directories
  + Petabyte scale, low-latency file system

  + Shared storage
  + Elastic capacity
  + Supports Network File System (NFS) versions 4.0 and 4.1 (NFSv4)
  + Compatible with all Linux-based AMIs for Amazon EC2
- Amazon EFS implementation
  + 1. Create Amazon EC2 resources and launch Amazon EC2 instance
  + 2. Create Amazon EFS file system
  + 3. Generate mount targets in the aporpriate subnets
  + 4. Connect Amazon EC2 instances to mount targets.
  + 5. Verify the resources and protection of AWS account
- Amazon EFS resources => File system
  + Mount target:
    * Subnet ID
    * Security groups
    * One or more per file system
    * Create in a VPC subnet
    * One per AZ
    + Must be in the same VPC
  + Tags: Key-value pairs
- Features of Amazon EFS:
  + Authentication and Access Control:

Amazon EFS offers multiple ways to control access to your file systems. Network access is managed through **VPC security groups** and **Network ACLs**. For a more granular level of control, you can use **AWS Identity and Access Management (IAM)** policies and **EFS Access Points**. IAM policies define which AWS users or roles can perform actions on the EFS file system, while access points enforce a specific user and group identity for all NFS (Network File System) client connections. Additionally, Amazon EFS supports **encryption for data both at rest and in transit** for enhanced security

  + Data Consistency:

Amazon EFS provides the **close-to-open consistency** model, which is a standard for network file systems. This means that when a file is written and then closed, any subsequent open of that file will see the changes. For applications that require stricter consistency, like those with synchronous writes, Amazon EFS can provide **read-after-write consistency**. The service also supports NFS version 4 file locking to help prevent data corruption from concurrent write operations.

  + Availability and Durability of EFS File Sytems: Amazon EFS is designed to be highly available and durable. It offers two main file system types:

    * **Regional file systems** store data redundantly across multiple Availability Zones (AZs) within a region, providing high availability and durability. This design is built to withstand the loss of an entire AZ without data loss

    * **One Zone file systems** store data redundantly within a single AZ. While they are more cost-effective, they are not resilient to an AZ failure and data loss is possible in such an event. EFS is designed for **11 nines (99.999999999%) of durability** over a given year.

  + Replication:

EFS offers a managed, asynchronous **replication feature** that automatically and transparently copies data from a source file system to a destination file system. This can be used for **disaster recovery** or to create a secondary copy of your data in a different AWS Region or within a different Availability Zone. The service maintains a **Recovery Point Objective (RPO)** of 15 minutes for most file systems.

E. Amazon S3 Glacier

- Amazon S3 Glacier is a secure, durable, and extremely low-cost service for data archieving and long-term backups
- Amazon S3 Glacier is a data archieving service that is designed for security, durability, and an extremely low cost.
- An archive is ay object such as a photo, video, file, or document that user store in the Amazon S3 Glacier
- Each archive has its own unique ID and also have description
- A vault is a container for storing archives
- When creating the Vault, specify the name and Region where to locate the Vault
- Each vault can have one vault access policy and one vault policy attached to it
  + Amazon S3 Glacier is designed ot provide 11 9s of durability for objects.
  + It supports the encryption of data in trasit and at rest through Secure Sockets Layer (SSL) or Transport Layer Security (TLS)
  + The Vault Lock feature enforces compliance through a policy
  + Extremely low-cost design works well for long-term archieving: Provides three options for access to archives - expedited, standard, and bulk-retrieval times range from a few minutes to several hours.
- Storage service for low-cost data archiving and long-term backup
- User can configure lifecycle archiving of Amazon S3 content to Amazon S3 Glacier
- Retrieval options:
  + Standard: 3-5 hours
  + Bulk: 5-12 hours
  + Expedited: 1-5 minutes
- Amazon S3 Glacier use cases:
  + Media asset archiving
  + Healthcare information archiving
  + Regulatory and compliance archiving

+ Scientific data archiving
+ Digital preservation
+ Magnetic tape replacement

- Lifecycle policies: Amazon S3 lifecycle policies enable user to delete or remove objects based on age

Amazon S3 Standard => Amazon S3 Standard-Infrequent Access => Amazon S3 Glacier => Delete
(30 days) (30 days) (1 year)

Storage Class	Features
S3 Standard	Larger or equal to three AZ
S3 Standard-Infrequent Access (IA)	• Retrieval fee is associated with objects • Most suitable for infrequently accessed data
S3 Intelligent-Tiering	• It automatically moves objects between tiers based on access patterns • Larger or equal to 3 AZ
S3 One Zone-IA	• One AZ • Cost less than Amazon S3 Standard-IA
S3 Glacier	• It is not available for real-time access • User must resotre objects before accessing them • Restoring objects can take between 1 minute and 12 hours
S3 Glacier Deep Archive	• Lowest cost storage for long-term retention (7-10 years old) • Larger or equal to three AZ • Retrieval time within 12 hours

- Storage Comparison:

	Amazon S3	Amazon S3 Glacier
Data Volume	No limit	No limit
Average Latency	ms	minutes/hours
Item Size	5 TB maximum	40 TB maximum
Cost/GB per Month	Higher Cost	Lower Cost
Billed Requests	PUT, COPY, POST, LIST and GET	UPLOAD and retrieval
Retrieval Pricing	$ per request	$ per request and per GB

- Security with Amazon S3 Glacier:
+ Control access with IAM
+ Amazon S3 Glacier encrypts user data with AES-256
+ Amazon S3 Glacier manages user's key for user