



Module 7 : Storage

Date: 2081/05/05

AWS EBS

AWS EBS stands for Amazon Elastic Block Store. Amazon EBS provides persistent block storage volumes for use with Amazon EC2 instances.

Persistent storage is any data storage device that retains data after power to that device is shut off. It is also called non-volatile storage. Each Amazon EBS volume is automatically replicated within its availability zone to protect you from component failure.

Amazon EBS volumes provide **the consistent and low latency performance** that is needed to run your workloads. With Amazon EBS, you can scale your usage up or down within minutes while paying a low price for only what you provision



Persistent means: lasting for long time.

Latency : the delay in network communication

Provision : supply

usage: the action of using something (noun)

AWS Storage options: Block storage VS object storage

What if you want to change one character in a 1-GB file?

Block Storage	Object Storage
Change one block (piece of file) that contains the character	Entire file must be updated
Block storage solutions are typically faster and use less bandwidth	Object Storage solutions are slow and use more bandwidth
They cost more	Less cost

More about Amazon EBS

Amazon EBS enables you to create individual storage volumes and attach them to an Amazon EC2 instance:

Amazon EBS offers block-level storage

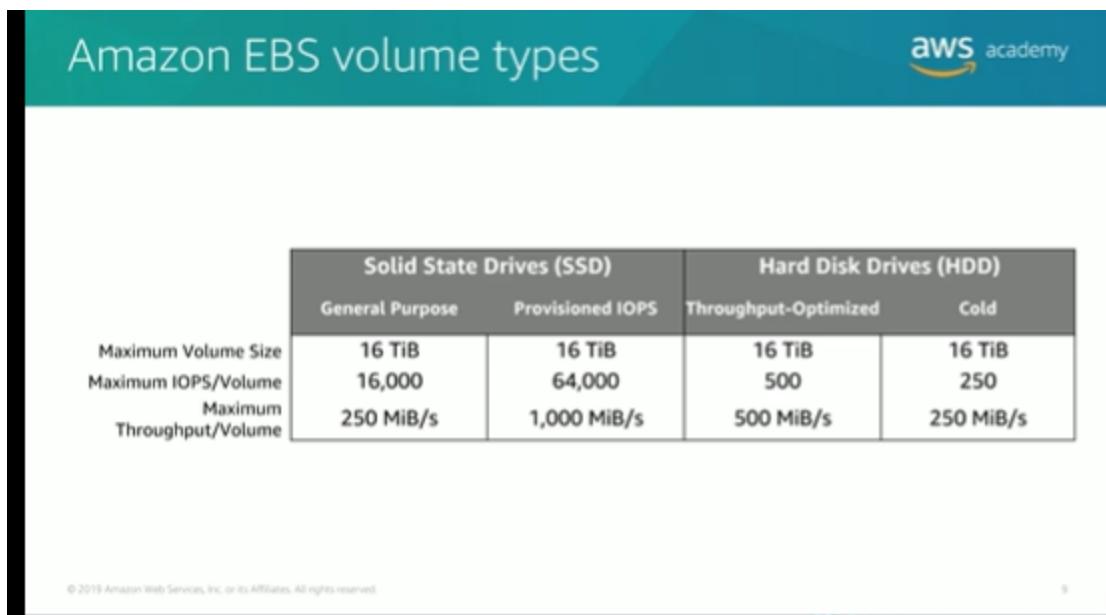
- Volumes are automatically replicated within it's Availability Zone.
- A backup of an Amazon EBS volume is called a snapshot. The first snapshot is called the Baseline snapshot, any other snapshot after the Baseline captures only what is different from the previous snapshot.
- Use of Amazon EBS volumes are:
 - Boot volumes and storage for Amazon EC2 instances
 - file system for data storage
 - database hosts
 - storage for enterprise applications.

Amazon EBS volume types

You can reduce your storage cost by selecting the type of storage that is the best match for your workload.

Provisioned IOPS SSD backed Amazon EBS volumes give you the highest performance. However, if your application doesn't require or won't use performance that high, general-purpose SSD might be sufficient.

You can use Amazon EBS volumes as primary storage for data that requires frequent updates such as the system drive for an instance or storage for a database application. You can also use them for throughput-intensive applications that perform continuous disc scans. Only SSDs can be used as boot volumes for EC2 instances.



The slide is titled "Amazon EBS volume types" and features the AWS Academy logo. It contains a table comparing Solid State Drives (SSD) and Hard Disk Drives (HDD) across four categories: General Purpose, Provisioned IOPS, Throughput-Optimized, and Cold. The table provides maximum volume size, maximum IOPS per volume, and maximum throughput per volume for each category.

	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

© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

Snapshots

To provide an even higher level of data durability with Amazon EBS, you can create point-in-time snapshots of your volumes and you can recreate a new volume from a snapshot at any time.

You can also share snapshots or even copy snapshots to different AWS regions for even greater Disaster Recovery protection.

Encryption

You can also encrypt Amazon EBS volumes at no additional cost. So the data is encrypted as it is in transit.

Elasticity

As your company grows the amount of data stored on your Amazon EBS volumes is also likely to grow. Amazon EBS volumes can increase capacity and change to different types. So you can change from the hard disk drives to solid state drives or increase from a 50 gigabyte volume to a 16 terabyte volume.

Amazon EBS: Volumes, IOPS, and pricing

1. Volumes -

Storage for all Amazon EBS volume types is charged by the amount you provision in gigabytes per month until you release the storage.

2. IOPS -

a. General Purpose SSD

Charge by the amount that you provision in GB per month until storage is released. Input output is included in the price of general purpose SSD volumes.

b. Magnetic

Charged by the number of request that you make to your volume.

c. Provisioned IOPS SSD:

You are charge by the amount you provision in IOPS multiplied by the percentage of days that you provision for the month.

Amazon EBS: Snapshots and data transfer

3. Snapshots:

If you opt for Amazon EBS snapshots, the cost of the snapshot is per gigabyte for each month the data is stored.

4. Data transfer:

Inbound data transfer is free.

Outbound data transfer across Regions incurs charges.

If you copy snapshots to another region , standard Amazon EBS snapshot charges apply for the storage in the destination region.

AWS S3

Companies need the ability to simply and securely collect, store, and analyze their data on a

massive scale. Amazon S3 is object storage that is built to store and retrieve any amount of data

from anywhere: websites and mobile apps, corporate applications, and data from Internet of

Things (IoT) sensors or devices.

Amazon S3 is object-level storage, which means that if you want to change a part of a file, you must make the change and then re-upload the entire modified file. Amazon S3 stores data as objects within resources that are called buckets

Strong and Useful Points:

1. **Unlimited Storage Capacity:** Amazon S3 provides virtually unlimited storage, allowing users to store as many objects as needed. This is crucial for businesses that generate large amounts of data over time.
2. **High Durability:** Amazon S3 is designed to offer "11 9s" of durability (99.99999999% durability), which means that the risk of losing your data is extremely low. This high level of durability is achieved by storing data redundantly across multiple facilities.
3. **Scalability:** S3 can handle trillions of objects and millions of requests per second, ensuring that it can scale with your storage needs without any performance degradation.
4. **Ease of Management:** Amazon S3 is fully managed, meaning that users do not need to manage any infrastructure. This reduces the complexity and time needed to maintain storage systems.
5. **Data Access and Security:** Users have fine-grained control over who can access their data using AWS Identity and Access Management (IAM), bucket policies, and object-level access control lists (ACLs). Additionally, data can be encrypted both in transit and at rest.

6. **Multi-Access Options:** S3 data can be accessed through multiple methods, including the AWS Management Console, APIs, SDKs, or third-party solutions. This flexibility allows integration with various applications and services.
7. **Event Notifications and Automation:** S3 can automatically trigger notifications or processes, such as AWS Lambda functions, when certain events occur (e.g., object uploads or deletions). This automation helps in streamlining workflows.
8. **Storage Class Analysis:** S3 provides tools to analyze storage access patterns, helping you transition data to the appropriate storage class, like S3 Standard-Infrequent Access (S3 Standard-IA), which optimizes cost-efficiency.
9. **Data Analytics Integration:** Users can analyze storage usage and export data to other services, like Amazon QuickSight, to gain insights and make informed decisions about data management.

Example to Understand:

Imagine you run an online photo storage service, where users upload and store millions of photos. These photos are stored in Amazon S3 buckets. Here's how Amazon S3 makes managing this service easy:

- **Unlimited Storage:** Whether you have 100 photos or 100 million photos, Amazon S3 can store them all without worrying about running out of space.
- **High Durability:** Your users trust that their photos will never be lost. With S3's "11 9s" durability, their photos are safe, even if something happens to the hardware in one facility.
- **Scalability:** As your service becomes popular, millions of users might upload photos simultaneously. S3 can handle this traffic without slowing down.
- **Ease of Management:** You don't need to worry about managing servers or storage devices. S3 takes care of all the infrastructure, so you can focus on improving your service.
- **Data Access and Security:** You can control who has access to the photos. By default, photos aren't shared publicly unless you explicitly allow it. You can also encrypt photos to keep them safe during transfer and storage.

- **Event Notifications and Automation:** When a user uploads a photo, you can set up an automatic notification or trigger a process (like generating a thumbnail) without manual intervention.
- **Storage Class Analysis:** If you notice that certain photos aren't accessed frequently, you can move them to a cheaper storage class like S3 Standard-IA, saving costs while still keeping the photos available when needed.

Explanation of New Terms:

- **Bucket:** A container in S3 where you store objects (files). Each bucket must have a unique name globally.
- **Object:** A file stored in a bucket. It can be anything like an image, video, or document.
- **Durability:** The likelihood that your data will remain intact and not be lost over time.
- **IAM (Identity and Access Management):** A service that controls who has access to resources in your AWS account, like S3 buckets.
- **Access Control Lists (ACLs):** Rules that define who can access an individual object within a bucket.
- **AWS Lambda:** A serverless computing service that runs code in response to events, like an object being uploaded to S3.
- **Storage Class:** A way to categorize your data based on how frequently it is accessed, with different pricing for each class.
- **Encryption:** The process of encoding data to protect it from unauthorized access, ensuring it can only be read by someone with the correct key.

Table 1: First Three Amazon S3 Storage Classes

Storage Class	Description	Key Points
Amazon S3 Standard	High durability, availability, and performance for frequently accessed data.	- Suitable for dynamic websites, big data analytics, and cloud applications. - Low latency and high throughput. - Ideal for frequently accessed data.

Amazon S3 Intelligent-Tiering	Automatically optimizes costs by moving data between frequent and infrequent access tiers.	- Great for data with unpredictable access patterns. - No retrieval fees. - Automatically moves data between tiers based on access patterns.
Amazon S3 Standard-Infrequent Access (S3 Standard-IA)	Low-cost storage for data accessed less frequently, but still needs quick access.	- Combines low cost with high performance. - Best for long-term storage and backups. - Retrieval fee per GB.

Table 2: Last Three Amazon S3 Storage Classes

Storage Class	Description	Key Points
Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA)	Low-cost storage for infrequently accessed data in a single Availability Zone.	- Stored in a single Availability Zone. - Cheaper than S3 Standard-IA. - Suitable for secondary backups and easily re-creatable data.
Amazon S3 Glacier	Secure, durable, and low-cost storage class for data archiving.	- Ideal for long-term data archiving. - Low storage cost. - Retrieval times range from minutes to hours.
Amazon S3 Glacier Deep Archive	Lowest-cost storage class designed for long-term retention and digital preservation.	- Designed for data accessed once or twice a year. - Perfect for regulatory compliance data. - Retrieval time up to 12 hours.

Table 3: Comparison of All Six Storage Classes

Storage Class	Use Case	Cost & Performance
Amazon S3 Standard	Frequently accessed data for high-performance applications.	High cost, high performance, low latency.
Amazon S3 Intelligent-Tiering	Data with unpredictable access patterns needing cost optimization.	Moderate cost, automatic tiering, no retrieval fees.
Amazon S3 Standard-Infrequent Access (S3 Standard-IA)	Data accessed less frequently but still requires quick access.	Lower cost than Standard, per-GB retrieval fee, high

Standard-IA)		durability.
Amazon S3 One Zone-Infrequent Access (S3 One Zone-IA)	Secondary backups or easily re-creatable data that don't need the redundancy of multiple zones.	Lower cost than Standard-IA, stored in a single Availability Zone, retrieval fee.
Amazon S3 Glacier	Long-term data archiving where data access is infrequent and can tolerate retrieval delays.	Low storage cost, various retrieval options, high durability.
Amazon S3 Glacier Deep Archive	Data that is rarely accessed, typically for regulatory compliance, with long retrieval times.	Lowest cost, long retrieval time (up to 12 hours), designed for long-term preservation.

Explanation of Terms:

- **Durability:** The likelihood that your data will not be lost. Amazon S3 offers "11 9s" durability, which means a very high level of data protection.
- **Latency:** The time it takes to retrieve or access data. Lower latency means faster access.
- **Availability Zone:** A distinct location within an AWS region that is engineered to be isolated from failures in other Availability Zones.
- **Tiering:** The process of moving data between storage classes to optimize costs.
- **Retrieval Fee:** A fee charged when accessing data stored in certain S3 storage classes, like S3 Standard-IA or S3 One Zone-IA.

Important Points about Using Amazon S3

1. Buckets:

- Buckets are logical containers for storing data in Amazon S3.
- Each bucket must have a unique name across all of Amazon S3 globally.
- Buckets act as the prefix for a set of files (objects).

2. Object Storage:

- Amazon S3 refers to files as "objects."

- Objects consist of the data being stored and metadata that describes the file.
- You can store almost any number of objects in a bucket.

3. Access Control:

- You can control who can create, delete, and list objects in a bucket.
- Permissions can be set at the bucket level and on individual objects.
- Access logs for buckets and objects can be viewed.

4. Geographical Region:

- When creating a bucket, you choose the AWS Region where the bucket and its contents will be stored.
- The Region is identified by a specific Region code (e.g., Tokyo Region: ap-northeast-1).

5. URL Structure:

- There are two URL styles to access buckets and objects in Amazon S3:

1. **Path-style URL:** `https://s3.[Region code].amazonaws.com/[bucket-name]/[object-name]`

2. **Virtual hosted-style URL:** `https://[bucket-name].s3-[Region code].amazonaws.com/[object-name]`

- The URL includes the bucket name, Region code, and object name.

6. Uploading Data:

- To upload data to Amazon S3, create a bucket in your chosen AWS Region.
- Upload almost any number of objects to the bucket, setting permissions on the data and metadata as needed.

Data Redundancy and Durability

- **Redundant Storage:** Data stored in an Amazon S3 bucket is redundantly stored across multiple AWS facilities within the selected Region.
- **High Durability:** Amazon S3 is designed to maintain data durability even in the event of concurrent data loss in two AWS facilities.

Scalability and Management

- **Seamless Scaling:** Amazon S3 automatically manages storage growth, scaling to handle high volumes of requests without the need for manual provisioning.
- **Billing:** You are billed only for the storage and requests you use, with no need to provision storage or throughput.

Access Methods

- **Flexible Access:** You can access data in Amazon S3 through various methods including the AWS Management Console, AWS CLI, SDKs, or REST-based endpoints.
- **Global Uniqueness:** Bucket names must be globally unique and DNS-compliant, and object keys should use URL-safe characters.

Common Use Cases

- **Application Data Storage:** S3 buckets can store application data, accessible by instances across EC2 and other platforms.
- **Static Web Hosting:** S3 can serve static website content like HTML, CSS, and JavaScript files.
- **Backup and Disaster Recovery:** S3's high durability makes it ideal for backups. Cross-Region replication can enhance availability and disaster recovery.
- **Media Hosting and Software Delivery:** S3 supports scalable and highly available infrastructure for media hosting and software delivery.

Pricing Considerations

- **Pay for Usage:** Costs vary based on storage used, data transfer, and the number and type of requests (e.g., GET, PUT, COPY).
- **Free Transfers:** Data transfers into S3 and within the same Region (e.g., to CloudFront or EC2) are free, but charges apply for transfers out of the Region.

Storage Classes and Cost Management

- **Storage Class Types:**

- **Standard Storage:** Offers 11 9s of durability and four 9s of availability.
- **Standard-Infrequent Access (S-IA):** Also provides 11 9s of durability but with slightly lower availability (three 9s) for less frequently accessed data.
- **Considerations for Cost Estimation:**
 - **Storage Class Type:** Different classes have different rates.
 - **Amount of Storage:** Costs are affected by the number and size of stored objects.
 - **Requests and Data Transfer:** Different rates apply to various request types, and charges are incurred for data transferred out of the S3 Region.

Key Takeaways

- **Fully Managed Service:** Amazon S3 provides a fully managed, scalable cloud storage solution.
- **Virtually Unlimited Storage:** You can store an unlimited number of objects.
- **Accessible Anytime, Anywhere:** Data in S3 can be accessed from anywhere at any time through a URL.
- **Rich Security Controls:** Amazon S3 offers robust security features to protect stored data.

AWS EFS

- **Amazon Elastic File System (Amazon EFS)** provides **simple, scalable, elastic file storage** for use with AWS services and on-premises resources.
- **Amazon EFS** offers a simple interface that enables quick and easy creation and configuration of file systems.
- **Scalability:** Amazon EFS automatically scales on demand without disrupting applications. It **grows and shrinks automatically** as files are added or removed, ensuring that your applications have the storage they need when they need it.
- **Fully Managed Service:** Amazon EFS is a fully managed service, making it easy to set up and scale file storage in the AWS Cloud. **No minimum fee or**

setup costs are required, and you pay only for the storage you use.

- **Supports High Durability and Availability:** Amazon EFS is designed to be highly durable and available. It can automatically scale from gigabytes to petabytes of data, with consistent performance for thousands of Amazon EC2 instances simultaneously accessing the file system.
- **Multiple Use Cases:** Amazon EFS is ideal for big data and analytics, media processing workflows, content management, web serving, and home directories.
- **Network File System (NFS) Compatibility:** Amazon EFS supports NFS versions 4.0 and 4.1 (NFSv4) and is compatible with all Linux-based AMIs for Amazon EC2.
- **Mount Targets:** To access your file system, you must create **mount targets** in your VPC. These mount targets allow Amazon EC2 instances in different Availability Zones to access and share a common data source.

Amazon EFS Architecture and Implementation

- **Amazon EFS Architecture** involves multiple **Availability Zones** (AZs) within a **VPC** (Virtual Private Cloud), with each AZ containing one or more **private subnets**. These subnets house **mount targets**, which are essential for accessing the Elastic File System (EFS).
- **Steps to Create and Use Amazon EFS:**
 1. **Create Amazon EC2 Resources:** Before launching and connecting to an EC2 instance, **create a key pair** unless you already have one.
 2. **Create Amazon EFS File System:** Set up your file system within the desired VPC.
 3. **Create Mount Targets:** In the appropriate subnets of your VPC, **create mount targets** that link to your EFS.
 4. **Connect EC2 Instances:** Mount the Amazon EFS file system to your EC2 instances by connecting them to the mount targets.
 5. **Cleanup and Protection:** Finally, **clean up your resources** and ensure that your AWS account is protected.

Amazon EFS resources

- **File System:** The primary resource in Amazon EFS, with properties like **ID**, **creation token**, **creation time**, **size**, and **state**.
- **Mount Targets:** These are essential for accessing your file system and come with properties such as **mount target ID**, **subnet ID**, **file system ID**, **IP address**, and **state**.
- **Tags:** Tags are **key-value pairs** that help organize your file systems. They do not exist unless associated with a file system.

New Terms Defined:

- **Mount Target:** A network endpoint in your VPC that allows you to access your Amazon EFS file system from an Amazon EC2 instance.
- **Petabyte:** A unit of data storage equal to 1,024 terabytes (TB) or approximately 1 million gigabytes (GB).
- **NFS (Network File System):** A distributed file system protocol that allows users on client computers to access files over a network as if they were on local storage.

Difference Between Amazon S3 and Amazon EFS:

Feature	Amazon S3	Amazon EFS
Type of Storage	Object storage for unstructured data (e.g., images, videos, backups).	File storage for structured data, ideal for use with AWS services and on-premises resources.
Scalability	Automatically scales for virtually unlimited data storage.	Automatically scales up or down as files are added or removed.
Access	Accessible via web interface, REST API, AWS CLI, or SDK.	Accessible via NFS interface, with support for full file system semantics (e.g., file locking).
Use Cases	Backup and storage, static web hosting, big data analytics.	Big data analytics, media processing workflows, content management, web serving, home directories.

Example Use Case	Hosting a static website with HTML, CSS, and JavaScript files.	Providing a shared file system for multiple EC2 instances in a big data processing workflow.
Durability & Availability	Highly durable (11 9s durability) with cross-Region replication options.	Highly durable and available, with consistent performance for multiple EC2 instances.

AWS S3 Glacier

Overview of Amazon S3 Glacier:

Amazon S3 Glacier is a data archiving service designed for **security, durability**, and **extremely low cost**. It's specifically tailored for long-term data storage needs where data retrieval is infrequent.

- **Key Features:**

- **Durability:** Designed to provide **11 nines (99.99999999%)** durability for stored objects, ensuring data integrity over time.
- **Security:** Supports encryption of data both in transit and at rest using **SSL** (Secure Sockets Layer) or **TLS** (Transport Layer Security).
- **Vault Lock Feature:** Enforces compliance by allowing the creation of immutable policies, ensuring that once a policy is locked, it cannot be changed.
- **Cost-Efficiency:** The service is optimized for **low-cost** storage, making it ideal for long-term archiving.

- **Storage Concepts:**

- **Archive:** The fundamental unit of storage in S3 Glacier, representing any stored object like a photo, video, or document. Each archive has a unique ID and can include a description.
- **Vault:** A container for storing archives. When creating a vault, you define its name and the region where it will reside.
- **Vault Access Policy:** Determines access permissions for the vault, specifying who can access the data and what operations they can

perform. It can also include a **Vault Lock Policy** to ensure the vault cannot be altered after the policy is locked.

- **Retrieval Options:**

- **Expedited Retrieval:** Typically takes **1-5 minutes** to access data, making it the quickest and most expensive option.
- **Standard Retrieval:** Takes **3-5 hours** to complete, offering a balance between speed and cost.
- **Bulk Retrieval:** The most cost-effective option, taking **5-12 hours** to retrieve data.

Practical Use Case:

- You can configure lifecycle policies to automatically archive data from an **Amazon S3 bucket** to S3 Glacier. For example, you might archive data after 30 days and set it to be deleted after 5 years.

Key Takeaways:

- **Amazon S3 Glacier** is ideal for organizations that need to store large amounts of data at a low cost but do not require frequent or immediate access to that data.
- The service offers multiple retrieval options to suit different needs, whether quick access is necessary or if cost savings are prioritized.
- **Vault Access Policies** provide robust control over who can access your archived data, adding an additional layer of security and compliance.

Amazon S3 Glacier: Use Cases and Lifecycle Management

Use Cases:

Amazon S3 Glacier is an ideal solution for various types of data archiving needs, including:

- **Media Asset Archiving:** Storing media files like videos and photos that are rarely accessed but need to be preserved for the long term.
- **Healthcare Information Archiving:** Storing sensitive healthcare data that must be kept secure and accessible for compliance purposes.

- **Regulatory and Compliance Archiving:** Ensuring data is stored securely and meets regulatory requirements.
- **Scientific Data Archiving:** Preserving large datasets from scientific research that need to be retained but are infrequently accessed.
- **Digital Preservation:** Archiving digital assets, such as historical documents, that require long-term storage.
- **Magnetic Tape Replacement:** Transitioning from traditional magnetic tape storage to a more reliable and accessible cloud-based solution.

Accessing Data in Amazon S3 Glacier:

To store and access data in Amazon S3 Glacier, you can use several tools:

- **AWS Management Console:** Allows you to perform basic operations like creating and deleting vaults and managing archive policies.
- **REST APIs, AWS SDKs (Java or .NET), and AWS CLI:** These tools provide more extensive capabilities for interacting with Amazon S3 Glacier, allowing for a broader range of operations.

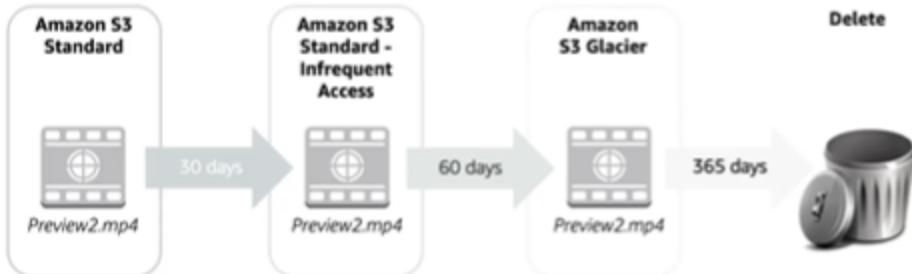
Lifecycle Management:

Automating data lifecycle management in Amazon S3 is crucial for optimizing storage costs:

Lifecycle policies



Amazon S3 lifecycle policies enable you to delete or move objects based on age.



© 2019 Amazon Web Services, Inc. or its Affiliates. All rights reserved.

- **Lifecycle Policies:** These policies enable you to automatically transition data between different Amazon S3 storage classes based on its age and access patterns.
 - Example:
 - **Amazon S3 Standard:** Initial storage for data that is frequently accessed.
 - **Amazon S3 Standard – Infrequent Access:** After 30 days, data is moved here if it's no longer frequently accessed.
 - **Amazon S3 Glacier:** After an additional 30 days, data is moved to Glacier for long-term storage.
 - **Deletion:** After a year, the data is automatically deleted.

Comparison of Amazon S3 and Amazon S3 Glacier:

- **Data Volume:** Both services allow for unlimited data storage.
- **Latency:** Amazon S3 offers millisecond latency for frequent access, while Glacier provides retrieval times ranging from minutes to hours.
- **Item Size:** Amazon S3 supports up to 5 TB per item, while Glacier can store items up to 40 TB.

- **Cost:** Glacier is more cost-effective for long-term storage, while S3 is priced higher due to its low-latency access.
- **Billing:** Amazon S3 charges for operations like PUT, COPY, POST, LIST, and GET, while Glacier charges for UPLOAD and retrieval operations.
- **Retrieval Pricing:** Glacier has a higher cost for data retrieval compared to S3.

Security and Encryption:

- **Server-Side Encryption:**
 - **Amazon S3 Glacier:** Data is encrypted by default using AES-256 encryption.
 - **Amazon S3:** Server-side encryption must be enabled by the application, with options like SSE-S3, SSE-C, and AWS KMS.

Access Control:

- **IAM Policies:** You can control access to your data in Amazon S3 Glacier using AWS Identity and Access Management (IAM) policies, which specify user permissions for accessing and managing data.

Summary:

Amazon S3 Glacier is a powerful and cost-effective solution for long-term data archiving. It is particularly suited for scenarios where data retrieval is infrequent but the data needs to be securely stored for extended periods. With automated lifecycle policies, you can efficiently manage data storage and optimize costs. Additionally, Amazon S3 Glacier provides robust security features, including encryption and IAM-based access control, making it a reliable choice for archiving sensitive data.

Thank you 