Unit 3: Object Storage Services

Traditional Storage, Need to Move to Cloud Storage, Traditional vs. Cloud Storage, Cost Cloud Storage, Different Storage Options Available on AWS, Simple Storage Service (S3) and Its Components, Working of S3, Difference Between S3, EBS and EFS, Bucket Policy, Access Control List (ACL), Versioning, Cross-Region Replication (CRR) and Its Use Case, Amazon S3, Transfer Acceleration, Choice of Storage Classes on S3 Lifecycle.

What is Cloud Storage?

Cloud storage is a cloud computing model that stores data on the Internet through a cloud computing provider who manages and operates data storage as a service. It’s delivered on demand with just-in-time capacity and costs, and eliminates buying and managing your own data storage infrastructure. This gives you agility, global scale and durability, with “anytime, anywhere” data access.

How Does Cloud Storage Work?

Cloud storage is purchased from a third party cloud vendor who owns and operates data storage capacity and delivers it over the Internet in a pay-as-you-go model. These cloud storage vendors manage capacity, security and durability to make data accessible to your applications all around the world.

Applications access cloud storage through traditional storage protocols or directly via an API. Many vendors offer complementary services designed to help collect, manage, secure and analyze data at massive scale.

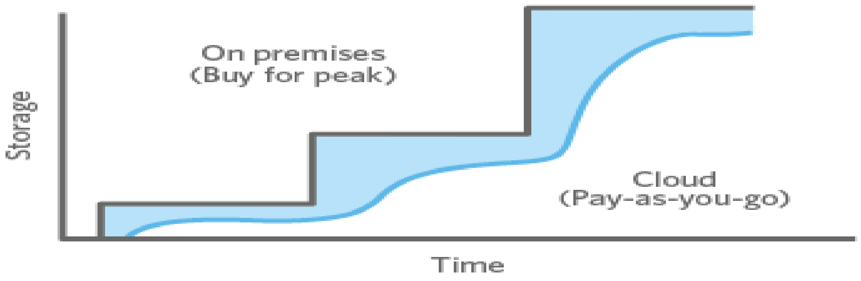
Benefits of Cloud Storage

Storing data in the cloud lets IT departments transform three areas:

1. Total Cost of Ownership. With cloud storage, there is no hardware to purchase, storage to provision, or capital being used for "someday" scenarios. You can add or remove capacity on demand, quickly change performance and retention characteristics, and only pay for storage that you actually use. Less frequently accessed data can even be automatically moved to lower cost tiers in accordance with auditable rules, driving economies of scale.

2. Time to Deployment. When development teams are ready to execute, infrastructure should never slow them down. Cloud storage allows IT to quickly deliver the exact amount of storage needed, right when it's needed. This allows IT to focus on solving complex application problems instead of having to manage storage systems.

3. Information Management. Centralizing storage in the cloud creates a tremendous leverage point for new use cases. By using cloud storage lifecycle management policies, you can perform powerful information management tasks including automated tiering or locking down data in support of compliance requirements.



Traditional vs. Cloud Storage

1. Cloud Storage :

Cloud storage is the storage option in which we use remote drives to store the data at the cloud location used by the client. It also uses the network to store the data to an off-site server which is owned by the service provider. User uses this storage options for capacity, bandwidth, and remote access.

Features of Cloud Storage:

- Cloud storage offers a variety of data security options.
- These storage options are easily accessible with every internet-connected device.
- Faults are easily traceable in cloud-based storage.
- Setting up and cloud storage is way more effective and easy than traditional ones.

2. Traditional Storage:

Traditional storage is the storage option in which we use local physical drives to store the data at the primary location of the client. User generally uses the disk-based hardware to store data and these are used for copying, managing, and integrating the data to software.

Features of Traditional Storage:

- Traditional storage is fast as they do not rely on internet speeds.
- Security can be manually set up by the user in traditional storage options.
- Users can recover the data anytime without having accessibility issues.
- On-site backup and modification are easy.

Differences between Cloud Storage and Traditional Storage:

Parameters	Cloud Storage	Traditional Storage
Performance	Cloud storage perform better due to using NoSQL.	Traditional storage perform a bit slow as compared to cloud.
Maintenance	This type of storage options are easy to maintain as you use and service provider takes care of maintenance.	This storage are heavy to manage as you need to manually run through maintenance tools.
Reliability	Cloud storage are highly reliable as it takes less time to get under functioning.	Traditional storage requires high initial effort and is less reliable.

Parameters	Cloud Storage	Traditional Storage
File Sharing	Cloud storage supports file sharing dynamically as it can be shared anywhere with network access.	Traditional storage requires physical drives to share data and network is to established between both.
File access time	In this system file access time is dependent on the network speed.	This system has fast access time as compared to cloud storage.
Security	Cloud storage are more secure as it integrates with many security tools.	Traditional storage are secure with they can get attacked easily through virus and malwares.
Applications	Amazon Drive, Dropbox, AutoSync are some applications of cloud storage.	HHD, SSD and Pendrives are some applications of traditional storage.

Need to Move to Cloud Storage

With Information Technology becoming more and more Cloud based nowadays (due to industry demanding reliability and scalability in their infrastructure), the Cloud storage system has become a very feasible solution. Various organizations are migrating their data to cloud storage, due to a few simple reasons. They want data to be easily accessible, cost effective and reliable.

How is Cloud storage better than any traditional data storage and why we should move towards cloud storage

- Performance: We are using NoSQL for Identity storage, NoSQL storage brings powerful read/write performance. We are maintaining low latency SSD for storage, this is why performance of NoSQL storage is continually progressing ahead of traditional HDD storage.
- Maintenance: Doing everything in-house is not ideal for businesses especially when you are a start-up or small to mid-sized business. Maintaining in-house traditional databases is very painful, you lose focus from you main application/feature, this is why SaaS(Software as a Service) solutions are more feasible. They allow you to outsource this nasty upkeep to those who know best and allow you to focus on your strengths.
- Support: Probably the most important thing that every single customer wants from providers. This is also one of biggest reasons to go for paid solution instead of Open source or free. Cloud storage has an advantage in this regard, support for these paid cloud storage solutions is very quick, accurate and efficient when compared with non-existent support for in-house solutions.
- Reliability: If you have any doubt on cloud storage's reliability then I can assure you that the cloud is built for reliability and up time. The architecture of cloud storage makes it reliable, no miracle

or magic behind this. Organizations from start-ups to established institutions, everyone requires up-time because the less up-time the less reliable your application is, and will impact your business.

- Security: Security is more subjective and anyone can argue both sides when you store data on the cloud, but if you find a vendor that is trusted then I don't think data on the cloud is any more or less secure than traditional storage. It is becoming more and more secure everyday and Cloud storage solutions have already captured a large market share of storage solutions.

Types of Cloud Storage available in AWS

AWS solutions typically fall into one of two categories: unmanaged or managed.

Unmanaged services are typically provisioned in discrete portions as specified by the user. You must manage how the service responds to changes in load, errors, and situations where resources become unavailable. Say that you launch a web server on an Amazon Elastic Compute Cloud (Amazon EC2) instance. Because Amazon EC2 is an unmanaged solution, that web server will not scale to handle increased traffic load or replace unhealthy instances with healthy ones unless you specify that it use a scaling solution, such as AWS Automatic Scaling. The benefit to using an unmanaged service is that you have more fine-tuned control over how your solution handles changes in load, errors, and situations where resources become unavailable.

Managed services require the user to configure them. For example, you create an Amazon Simple Storage Service (Amazon S3) bucket and then set permissions for it. However, managed services typically require less configuration. Say that you have a static website that you host in a cloud-based storage solution, such as Amazon S3. The static website does not have a web server. However, because Amazon S3 is a managed solution, features such as scaling, fault-tolerance, and availability would be handled automatically and internally by Amazon S3.

There are three types of cloud data storage: object storage, file storage, and block storage. Each offers their own advantages and have their own use cases:

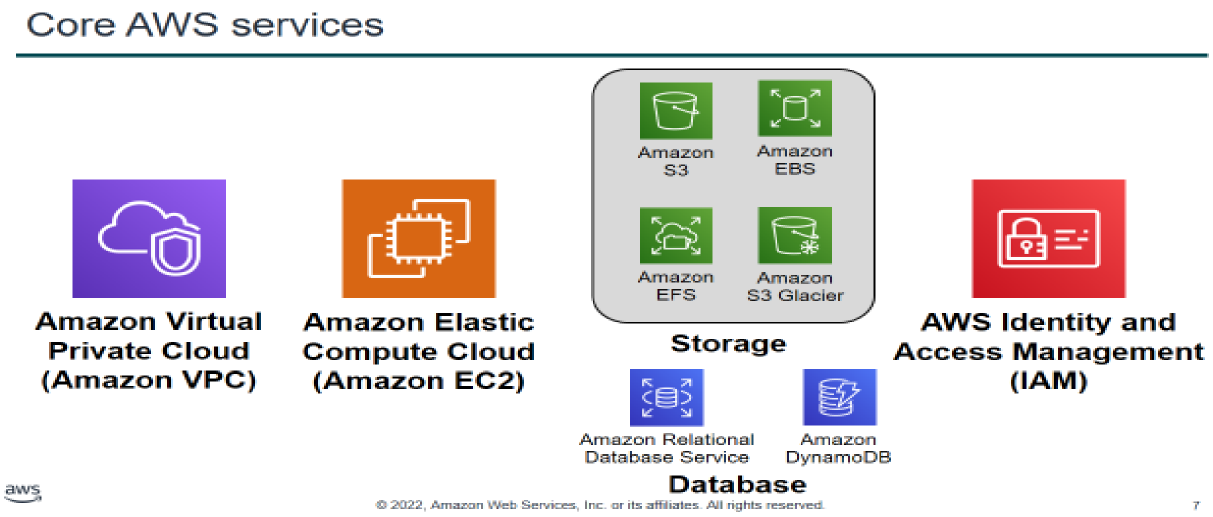
- Object Storage - Applications developed in the cloud often take advantage of object storage's vast scalablity and metadata characteristics. Object storage solutions like Amazon Simple Storage Service (S3) are ideal for building modern applications from scratch that require scale and flexibility, and can also be used to import existing data stores for analytics, backup, or archive.

- File Storage - Some applications need to access shared files and require a file system. This type of storage is often supported with a Network Attached Storage (NAS) server. File storage solutions like Amazon Elastic File System (EFS) are ideal for use cases like large content repositories, development environments, media stores, or user home directories.

- Block Storage - Other enterprise applications like databases or ERP systems often require dedicated, low latency storage for each host. This is analagous to direct-attached storage (DAS) or a Storage Area Network (SAN). Block-based cloud storage solutions like Amazon Elastic Block Store (EBS) are provisioned with each virtual server and offer the ultra low latency required for high performance workloads.

Storage is another AWS core service category. Some broad categories of storage include: instance store (ephemeral storage), Amazon EBS, Amazon EFS, Amazon S3, and Amazon S3 Glacier.

- Instance store, or ephemeral storage, is **temporary storage** that is added to your Amazon EC2 instance.
- Amazon EBS is **persistent, mountable storage** that can be mounted as a device to an Amazon EC2 instance. Amazon EBS can be mounted to an Amazon EC2 instance only within the same Availability Zone. Only one Amazon EC2 instance at a time can mount an Amazon EBS volume.
- Amazon EFS is a shared file system that multiple Amazon EC2 instances can mount at the same time.
- Amazon S3 is persistent storage where each file becomes an object and is available through a Uniform Resource Locator (URL); it can be accessed from anywhere.
- Amazon S3 Glacier is for cold storage for data that is not accessed frequently (for example, when you need long-term data storage for archival or compliance reasons).



Core AWS services

Amazon Virtual Private Cloud (Amazon VPC)

Amazon Elastic Compute Cloud (Amazon EC2)

Amazon S3

Amazon EBS

Amazon EFS

Amazon S3 Glacier

Storage

Amazon Relational Database Service

Amazon DynamoDB

Database

AWS Identity and Access Management (IAM)

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

7

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 sometimes called non-volatile storage.

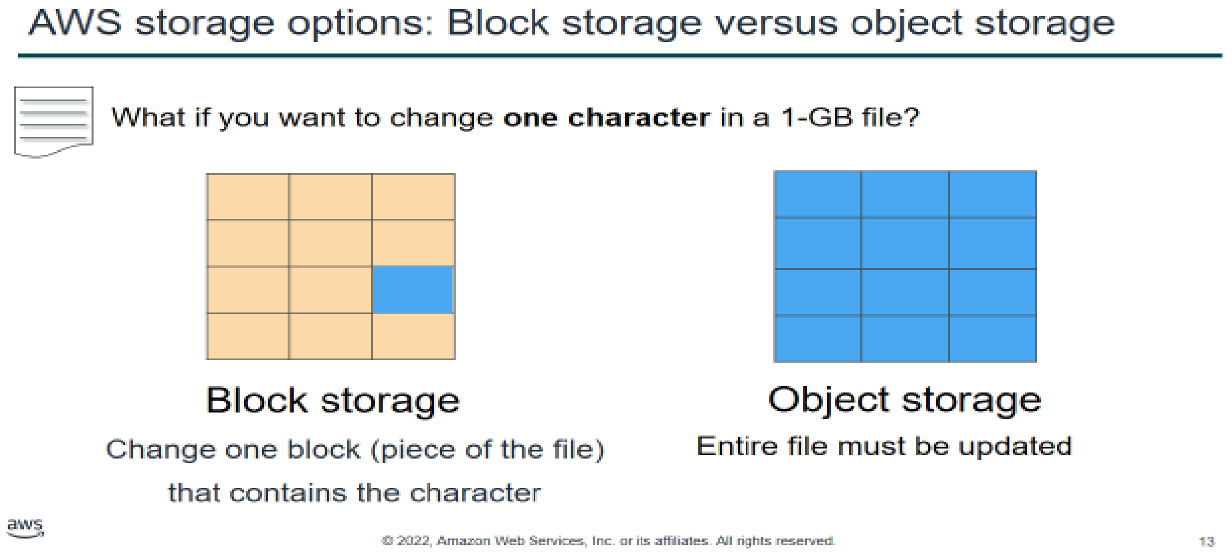
Each Amazon EBS volume is automatically replicated within its Availability Zone to protect you from component failure. It is designed for high availability and durability. Amazon EBS volumes provide the consistent and low-latency performance that is needed to run your workloads.

With Amazon EBS, you can scale your usage up or down within minutes, while paying a low price for only what you provision.

What happens if you want to change one character in a 1-GB file? With block storage, you change only the block that contains the character. With object storage, the entire file must be updated.

One critical difference between some storage types is whether they offer block-level storage or object-level storage.

This difference has a major effect on the throughput, latency, and cost of your storage solution. Block storage solutions are typically faster and use less bandwidth, but they can cost more than object-level storage.



AWS storage options: Block storage versus object storage

What if you want to 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

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. 13

Amazon EBS enables you to create individual storage volumes and attach them to an Amazon EC2 instance. Amazon EBS offers block-level storage, where its volumes are automatically replicated within its Availability Zone. Amazon EBS is designed to provide durable, detachable, block-level storage (which is like an external hard drive) for your Amazon EC2 instances. Because they are directly attached to the instances, they can provide low latency between where the data is stored and where it might be used on the instance.

For this reason, they can be used to run a database with an Amazon EC2 instance. Amazon EBS volumes are included as part of the backup of your instances into Amazon Machine Images (or AMIs). AMIs are stored in Amazon S3 and can be reused to create new Amazon EC2 instances later.

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.

Amazon EBS volumes uses include:

- Boot volumes and storage for Amazon EC2 instances
- Data storage with a file system
- Database hosts
- Enterprise applications

Amazon EBS enables you to create **point-in-time** snapshots of your volumes, and you can re-create 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 (DR) protection**. For example, you can encrypt and share your snapshots from Virginia in the US to Tokyo, Japan.

You can also have encrypted Amazon EBS volumes at no additional cost, so the data that moves between the EC2 instance and the EBS volume inside AWS data centers is encrypted in transit.

As your company grows, the amount of data that is 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 hard disk drives (HDDs) to solid state drives (SSDs) or increase from a 50-GB volume to a 16-TB volume. For example, you can do this resize operation dynamically without needing to stop the instances.

Amazon EBS provides block-level storage volumes for use with Amazon EC2 instances. Amazon EBS volumes are off-instance storage that persists independently from the life of an instance. They are analogous to virtual disks in the cloud. Amazon EBS provides three volume types: General Purpose SSD, Provisioned IOPS SSD, and magnetic.

The three volume types differ in performance characteristics and cost, so you can choose the right storage performance and price for the needs of your applications.

Additional benefits include replication in the same Availability Zone, easy and transparent encryption, elastic volumes, and backup by using snapshots.

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**.

Amazon S3 is a managed cloud storage solution that is designed to scale seamlessly and provide 11 9s of durability. You can store virtually as many objects as you want in a bucket, and you can write, read, and delete objects in your bucket. Bucket names are universal and must be unique across all existing bucket names in Amazon S3. Objects can be up to 5 TB in size. By default, data in Amazon S3 is stored redundantly across multiple facilities and multiple devices in each facility.

The data that you store in Amazon S3 is not associated with any particular server, and you do not need manage any infrastructure yourself. You can put as many objects into Amazon S3 as you want. Amazon S3 holds trillions of objects and regularly peaks at millions of requests per second.

Objects can be almost any data file, such as images, videos, or server logs. Because Amazon S3 supports objects as large as several terabytes in size, you can even store database snapshots as objects. Amazon S3 also provides low-latency access to the data over the internet by Hypertext Transfer Protocol (HTTP) or Secure HTTP (HTTPS), so you can retrieve data anytime from anywhere. You can also access Amazon S3 privately through a virtual private cloud (VPC) endpoint. You get fine-grained control over who can access your data by using AWS Identity and Access Management (IAM) policies, Amazon S3 bucket policies, and even per-object access control lists.

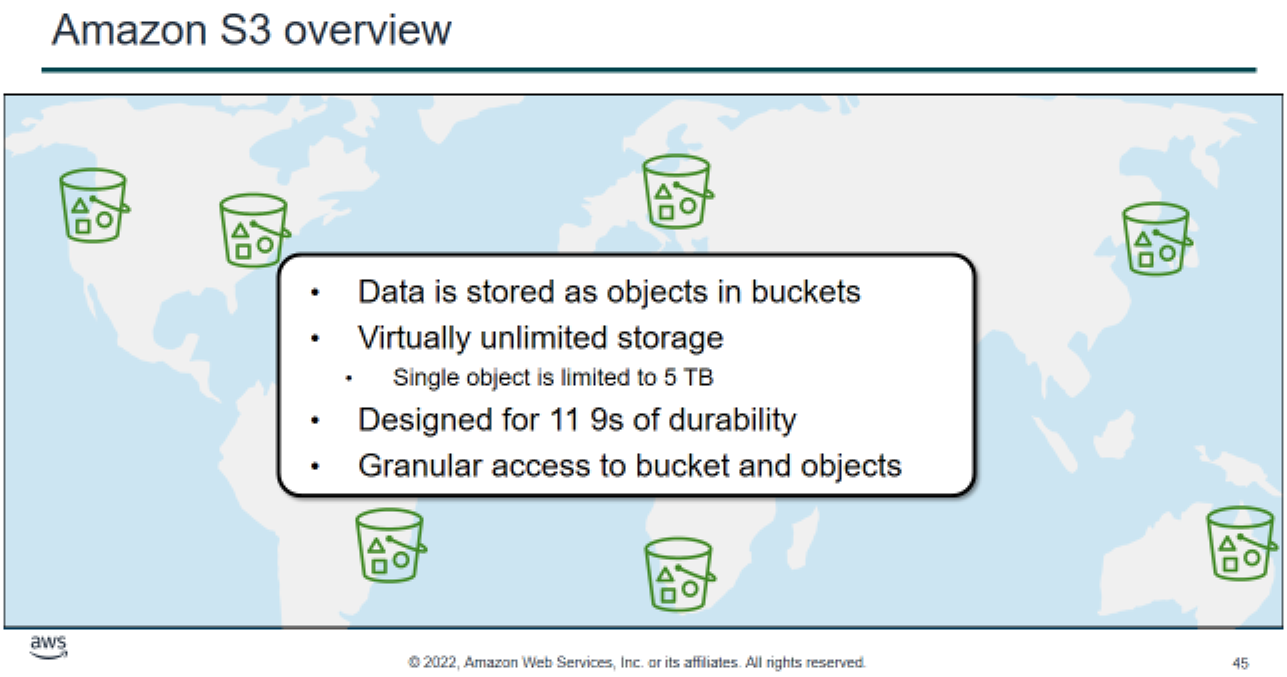
By default, none of your data is shared publicly. You can also encrypt your data in transit and choose to enable server-side encryption on your objects.

You can access Amazon S3 through the web-based AWS Management Console; programmatically through the API and SDKs; or with third-party solutions, which use the API or the SDKs.

Amazon S3 includes event notifications that enable you to set up automatic notifications when certain events occur, such as when an object is uploaded to a bucket or deleted from a specific bucket. Those notifications can be sent to you, or they can be used to trigger other processes, such as AWS Lambda functions.

With storage class analysis, you can analyze storage access patterns and transition the right data to the right storage class. The Amazon S3 Analytics feature automatically identifies the optimal lifecycle policy to transition less frequently accessed storage to Amazon S3 Standard – Infrequent Access (Amazon S3 Standard-IA). You can configure a storage class analysis policy to monitor an entire bucket, a prefix, or an object tag.

When an infrequent access pattern is observed, you can easily create a new lifecycle age policy that is based on the results. Storage class analysis also provides daily visualizations of your storage usage in the AWS Management Console. You can export them to an Amazon S3 bucket to analyze by using the business intelligence (BI) tools of your choice, such as Amazon QuickSight.



Amazon S3 overview

- 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

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

45

Amazon S3 offers a range of object-level storage classes that are designed for different use cases. These classes include:

- **Amazon S3 Standard** – Amazon S3 Standard is designed for high durability, availability, and performance object storage for frequently accessed data. Because it delivers low latency and high throughput, Amazon S3 Standard is appropriate for a variety of use cases, including cloud applications, dynamic websites, content distribution, mobile and gaming applications, and big data analytics.
- **Amazon S3 Intelligent-Tiering** – The Amazon S3 Intelligent-Tiering storage class is designed to optimize costs by automatically moving data to the most cost-effective access tier, without performance impact or operational overhead. For a small monthly monitoring and automation fee per object, Amazon S3 monitors access patterns of the objects in Amazon S3 Intelligent-Tiering, and moves the objects that have not been accessed for 30 consecutive days to the infrequent access tier. If an object in the infrequent access tier is accessed, it is automatically moved back to the frequent access tier. There are no retrieval fees when you use the Amazon S3 Intelligent-Tiering storage class, and no additional fees when objects are moved between access tiers. It works well for long-lived data with access patterns that are unknown or unpredictable.

- **Amazon S3 Standard-Infrequent Access (Amazon S3 Standard-IA)** – The Amazon S3 Standard-IA storage class is used for data that is accessed less frequently, but requires rapid access when needed. Amazon S3 Standard-IA is designed to provide the high durability, high throughput, and low latency of Amazon S3 Standard, with a low per-GB storage price and per-GB retrieval fee. This combination of low cost and high performance makes Amazon S3 Standard-IA good for long-term storage and backups, and as a data store for disaster recovery files.

- **Amazon S3 One Zone-Infrequent Access (Amazon S3 One Zone-IA)** – Amazon S3 One Zone-IA is for data that is accessed less frequently, but requires rapid access when needed. Unlike other Amazon S3 storage classes, which store data in a minimum of three Availability Zones, Amazon S3 One Zone-IA stores data in a single Availability Zone and it costs less than Amazon S3 Standard-IA. Amazon S3 One Zone-IA works well for customers who want a lower-cost option for infrequently accessed data, but do not require the availability and resilience of Amazon S3 Standard or Amazon S3 Standard-IA. It is a good choice for storing secondary backup copies of on-premises data or easily re-creatable data. You can also use it as cost-effective storage for data that is replicated from another AWS Region by using Amazon S3 Cross-Region Replication.

- **Amazon S3 Glacier** – Amazon S3 Glacier is a secure, durable, and low-cost storage class for data archiving. You can reliably store any amount of data at costs that are competitive with—or cheaper than—on-premises solutions. To keep costs low yet suitable for varying needs, Amazon S3 Glacier provides three retrieval options that range from a few minutes to hours. You can upload objects directly to Amazon S3 Glacier, or use Amazon S3 lifecycle policies to transfer data between any of the Amazon S3 storage classes for active data (Amazon S3 Standard, Amazon S3 Intelligent-Tiering, Amazon S3 Standard-IA, and Amazon S3 One Zone-IA) and Amazon S3 Glacier.

- **Amazon S3 Glacier Deep Archive** – Amazon S3 Glacier Deep Archive is the lowest-cost storage class for Amazon S3. It supports long-term retention and digital preservation for data that might be accessed once or twice in a year. It is designed for customers — particularly customers in highly regulated industries, such as financial services, healthcare, and public sectors — that retain datasets for 7–10 years (or more) to meet regulatory compliance requirements. Amazon S3 Glacier Deep Archive can also be used for backup and disaster recovery use cases. It is a cost-effective and easy-to-manage alternative to magnetic tape systems, whether these tape systems are on-premises libraries or off-premises services. Amazon S3 Glacier Deep Archive complements Amazon S3 Glacier, and it is also designed to provide 11 9s of durability. All objects that are stored in Amazon S3 Glacier Deep Archive are replicated and stored across at least three geographically dispersed Availability Zones, and these objects can be restored within 12 hours.

- To use Amazon S3 effectively, you must understand a few simple concepts. First, Amazon S3 stores data inside **buckets**. Buckets are essentially the prefix for a set of files, and must be uniquely named across all of Amazon S3 globally. Buckets are logical containers for objects. You can have one or more buckets in your account. You can control access for each bucket—who can create, delete, and list objects in the bucket. You can also view access logs for the bucket and its objects, and choose the geographical region where Amazon S3 stores the bucket and its contents.

- To upload your data (such as photos, videos, or documents), create a bucket in an AWS Region, and then upload almost any number of objects to the bucket.

- In the example, Amazon S3 was used to create a bucket in the Tokyo Region, which is identified within AWS formally by its Region code: ap-northeast-1

- The URL for a bucket is structured like the examples. You can use two different URL styles to refer to buckets.

- Amazon S3 refers to files as objects. As soon as you have a bucket, you can store almost any number of objects inside it. An object is composed of data and any metadata that describes that file, including a URL. To store an object in Amazon S3, you upload the file that you want to store to a bucket.

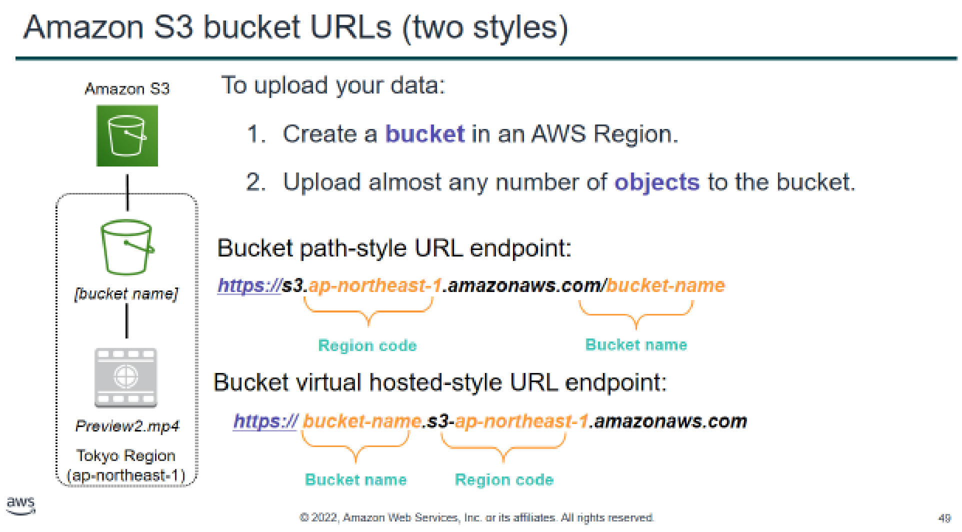
- When you upload a file, you can set permissions on the data and any metadata.

- In this example, the object Preview2.mp4 is stored inside the bucket. The URL for the file includes the object name at the end.

- When you create a bucket in Amazon S3, it is associated with a specific AWS Region. When you store data in the bucket, it is redundantly stored across multiple AWS facilities within your selected Region.

- Amazon S3 automatically manages the storage behind your bucket while your data grows. You can get started immediately, and your data storage will grow with your application needs.

- Amazon S3 also scales to handle a high volume of requests. You do not need to provision the storage or throughput, and you are billed only for what you use



Amazon S3 bucket URLs (two styles)

To upload your data:

1. Create a **bucket** in an AWS Region.
2. Upload almost any number of **objects** to the bucket.

Bucket path-style URL endpoint:

https://s3.ap-northeast-1.amazonaws.com/bucket-name

Bucket virtual hosted-style URL endpoint:

https:// bucket-name.s3-ap-northeast-1.amazonaws.com

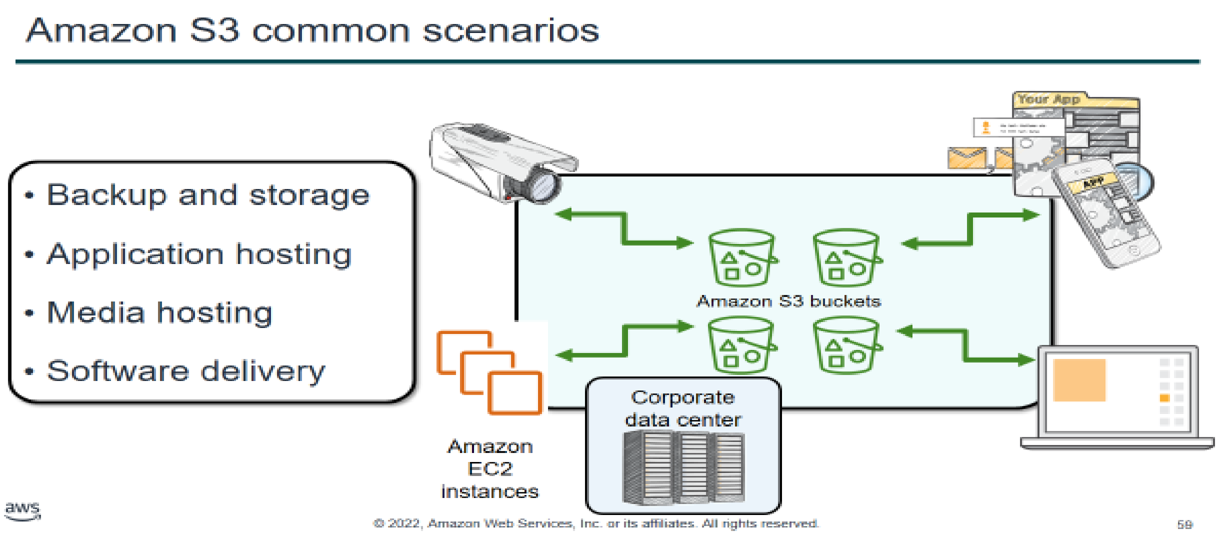
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved. 49

This flexibility to store a virtually unlimited amount of data—and to access that data from anywhere—means that Amazon S3 is suitable for a variety of scenarios. You will now consider some use cases for Amazon S3:

- As a location for any application data, Amazon S3 buckets provide a shared location for storing objects that any instances of your application can access—including applications on Amazon EC2 or even traditional servers. This feature can be useful for user-generated media files, server logs, or other files that your application must store in a common location. Also, because the content can be fetched directly over the internet, you can offload serving that content from your application and enable clients to directly fetch the data from Amazon S3 themselves.

- For static web hosting, Amazon S3 buckets can serve the static contents of your website, including HTML, CSS, JavaScript, and other files.

- The high durability of Amazon S3 makes it a good candidate for storing backups of your data. For greater availability and disaster recovery capability, Amazon S3 can even be configured to support cross-Region replication so that data in an Amazon S3 bucket in one Region can be automatically replicated to another Amazon S3 Region.
- **Backup and storage** – Provide data backup and storage services for others
- **Application hosting** – Provide services that deploy, install, and manage web applications
- **Media hosting** – Build a redundant, scalable, and highly available infrastructure that hosts video, photo, or music uploads and downloads
- **Software delivery** – Host your software applications that customers can download



Amazon S3 common scenarios

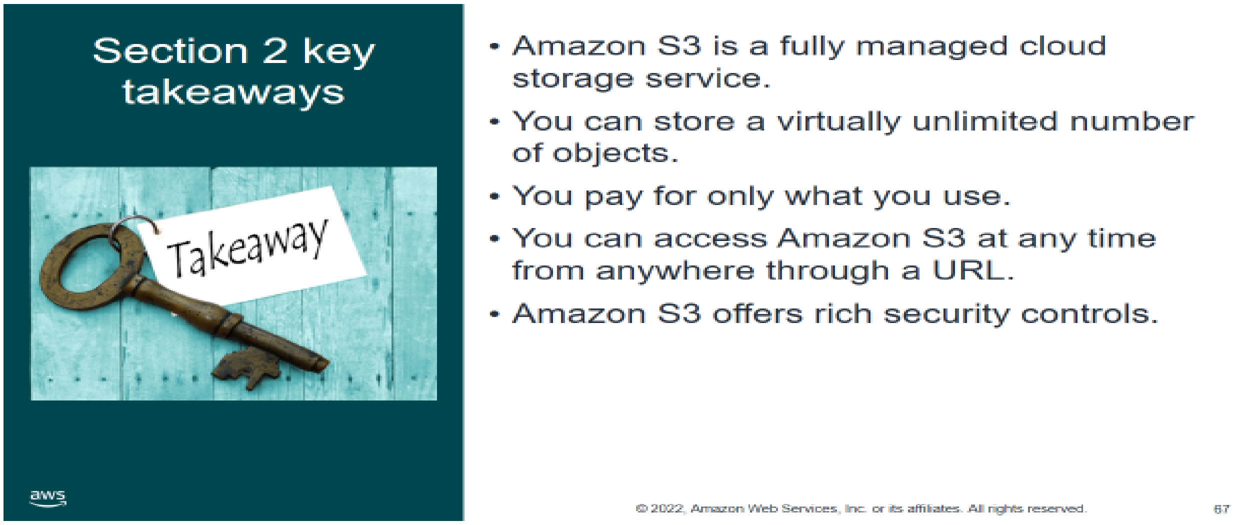
- Backup and storage
- Application hosting
- Media hosting
- Software delivery

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

59

With Amazon S3, specific costs vary depending on the Region and the specific requests that were made. You pay only for what you use, including gigabytes per month; transfer out of other Regions; and PUT, COPY, POST, LIST, and GET requests.

As a general rule, you pay only for transfers that cross the boundary of your Region, which means you do not pay for transfers in to Amazon S3 or transfers out from Amazon S3 to Amazon CloudFront edge locations within that same Region.



Section 2 key takeaways

Takeaway
aws

- Amazon S3 is a fully managed cloud storage service.
- You can store a virtually unlimited number of objects.
- You pay for only what you use.
- You can access Amazon S3 at any time from anywhere through a URL.
- Amazon S3 offers rich security controls.

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

67

Amazon Elastic File System (Amazon EFS)

Amazon EFS implements 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 (NFS) protocol.

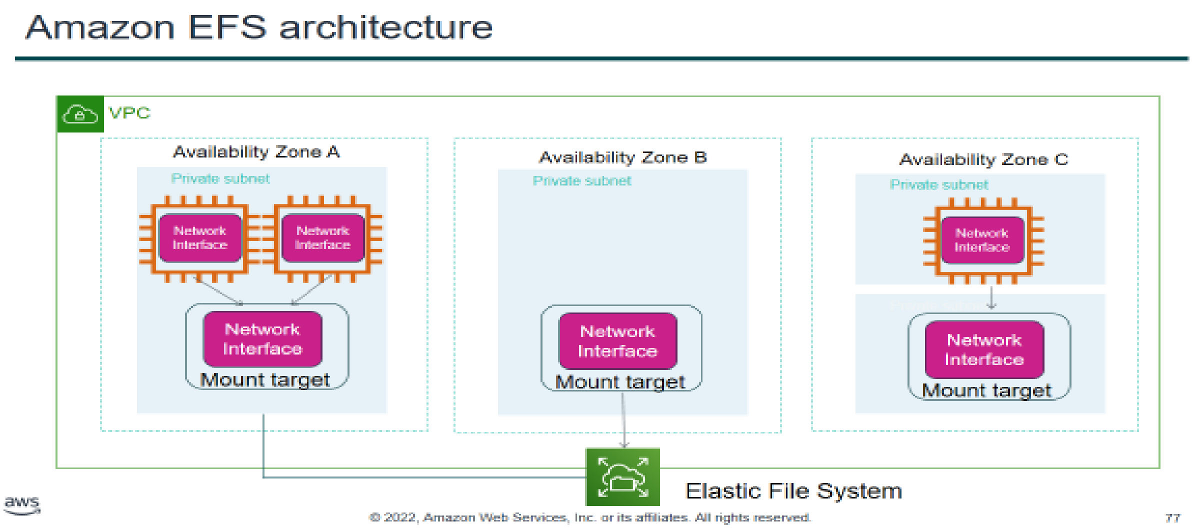
Amazon Elastic File System (Amazon EFS) provides simple, scalable, elastic file storage for use with AWS services and on-premises resources. It offers a simple interface that enables you to create and configure file systems quickly and easily.

Amazon EFS is built to dynamically scale on demand without disrupting applications—it will grow and shrink automatically as you add and remove files. It is designed so that your applications have the storage they need, when they need it.

Amazon EFS is a fully managed service that makes it easy to set up and scale file storage in the AWS Cloud. You can use Amazon EFS to build a file system for big data and analytics, media processing workflows, content management, web serving, and home directories.

You can create file systems that are accessible to Amazon EC2 instances through a file system interface (using standard operating system file I/O APIs). These file systems support full file system access semantics, such as strong consistency and file locking.

Amazon EFS file systems can automatically scale from gigabytes to petabytes of data without the need to provision storage. Thousands of Amazon EC2 instances can access an Amazon EFS file system at the same time, and Amazon EFS is designed to provide consistent performance to each Amazon EC2 instance. Amazon EFS is also designed to be highly durable and highly available. Amazon EFS requires no minimum fee or setup costs, and you pay only for the storage that you use.



Amazon EFS architecture

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon EFS provides file storage in the cloud. With Amazon EFS, you can create a file system, mount the file system on an Amazon EC2 instance, and then read and write data from to and from your file system. You can mount an Amazon EFS file system in your VPC, through NFS versions 4.0 and 4.1 (NFSv4).

You can access your Amazon EFS file system concurrently from Amazon EC2 instances in your VPC, so applications that scale beyond a single connection can access a file system. Amazon EC2 instances that run in multiple Availability Zones within the same AWS Region can access the file system, so many users can access and share a common data source.

In the diagram, the VPC has three Availability Zones, and each Availability Zone has one mount target that was created in it. We recommend that you access the file system from a mount target within the same Availability Zone. One of the Availability Zones has two subnets. However, a mount target is created in only one of the subnets.

In Amazon EFS, a file system is the primary resource. Each file system has properties such as:

- ID
- Creation token
- Creation time
- File system size in bytes
- Number of mount targets that are created for the file system
- File system state

Amazon EFS also supports other resources to configure the primary resource. These resources include mount targets and tags.

Mount target: To access your file system, you must create mount targets in your VPC.

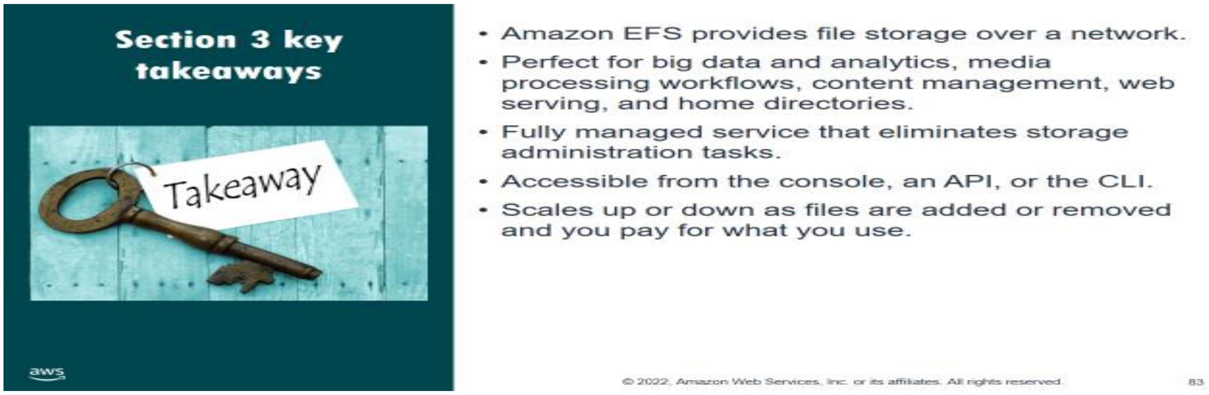
Each mount target has the following properties:

- The mount target ID
- The subnet ID for the subnet where it was created
- The file system ID for the file system where it was created
- An IP address where the file system can be mounted
- The mount target state

You can use the IP address or the Domain Name System (DNS) name in your mount command.

Tags: To help organize your file systems, you can assign your own metadata to each of the file systems that you create. Each tag is a key-value pair.

Think of mount targets and tags as sub resources that do not exist unless they are associated with a file system.


Takeaway
aws

Section 3 key takeaways

- Amazon EFS provides file storage over a network.
- Perfect for big data and analytics, media processing workflows, content management, web serving, and home directories.
- Fully managed service that eliminates storage administration tasks.
- Accessible from the console, an API, or the CLI.
- Scales up or down as files are added or removed and you pay for what you use.

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon S3 Glacier is a secure, durable, and extremely low-cost cloud storage service for data archiving and long-term backup.

When you use Amazon S3 Glacier to archive data, you can store your data at an extremely low cost (even in comparison to Amazon S3), but you cannot retrieve your data immediately when you want it.

Data that is stored in Amazon S3 Glacier can take several hours to retrieve, which is why it works well for archiving.

There are three key Amazon S3 Glacier terms you should be familiar with:

- **Archive** – Any object (such as a photo, video, file, or document) that you store in Amazon S3 Glacier. It is the base unit of storage in Amazon S3 Glacier. Each archive has its own unique ID and it can also have a description.
- **Vault** – A container for storing archives. When you create a vault, you specify the vault name and the Region where you want to locate the vault.
- **Vault access policy** – Determine who can and cannot access the data that is stored in the vault, and what operations users can and cannot perform. One vault access permissions policy can be created for each vault to manage access permissions for that vault. You can also use a vault lock policy to make sure that a vault cannot be altered. Each vault can have one vault access policy and one vault lock policy that are attached to it.

You have three options for retrieving data, each with varying access times and cost:

- **Expedited** retrievals are typically made available within 1–5 minutes (highest cost).
- **Standard** retrievals typically complete within 3–5 hours (less time than expedited, more time than bulk).
- **Bulk** retrievals typically complete within 5–12 hours (lowest cost).

You might compare these options to choosing the cost for shipping a package by using the most economical method for your needs.

Amazon S3 Glacier's data archiving means that although you can store your data at an extremely low cost (even in comparison to Amazon S3), you cannot retrieve your data immediately when you want it.

Data stored in Amazon S3 Glacier can take several hours to retrieve.

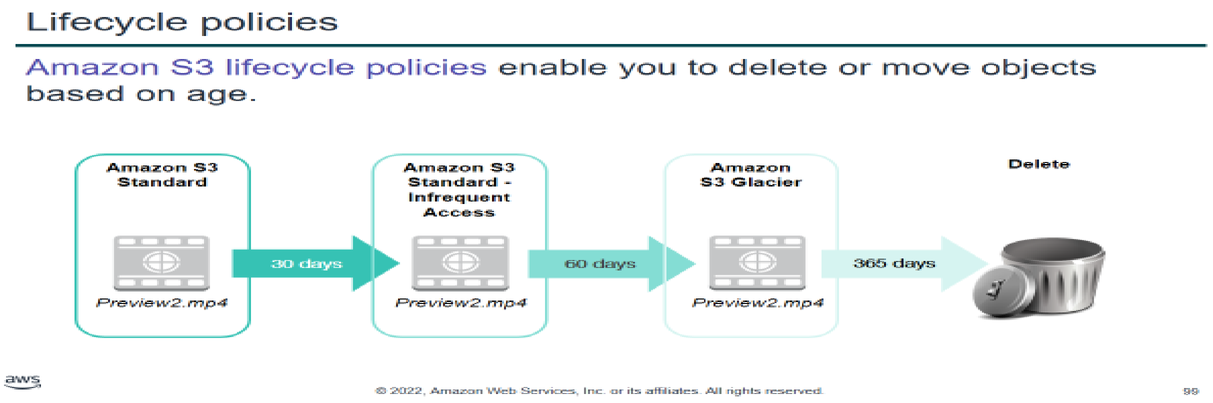
You should be familiar with three key Amazon S3 Glacier terms:

- **Archive:** Any object such as a photo, video, file, or document that you store in Amazon S3 Glacier. It is the base unit of storage in Amazon S3 Glacier. Each archive has its own unique ID and can also have a description.
- **Vault:** A container for storing archives. When you create a vault, you specify the vault name and the region in which you would like to locate the vault.
- **Vault Access Policy:** Determine who can and cannot access the data stored in the vault and what operations users can and cannot perform. One vault access policy can be created for each vault to manage access permissions for that vault. You can also use a vault lock policy to make sure a vault

cannot be altered. Each vault can have one vault access policy and one vault lock policy that is attached to it.

Three options are available for retrieving data with varying access times and cost: expedited, standard, and bulk retrievals. They are listed as follows:

- **Expedited** retrievals are typically made available within 1 – 5 minutes (highest cost).
- **Standard** retrievals typically complete within 3 – 5 hours (less than expedited, more than bulk).
- **Bulk** retrievals typically complete within 5 – 12 hours (lowest cost).
- To store and access data in Amazon S3 Glacier, you can use the AWS Management Console. However, only a few operations—such as creating and deleting vaults, and creating and managing archive policies—are available in the console.
- For almost all other operations and interactions with Amazon S3 Glacier, you must use either the Amazon S3 Glacier REST APIs, the AWS Java or .NET SDKs, or the AWS CLI.
- You can also use lifecycle policies to archive data into Amazon S3 Glacier. Next, you will learn about lifecycle policies.
- You should automate the lifecycle of the data that you store in Amazon S3. By using lifecycle policies, you can cycle data at regular intervals between different Amazon S3 storage types. This automation reduces your overall cost, because you pay less for data as it becomes less important with time.
- In addition to setting lifecycle rules per object, you can also set lifecycle rules per bucket.
- Consider an example of a lifecycle policy that moves data as it ages from **Amazon S3 Standard** to **Amazon S3 Standard – Infrequent Access**, and finally, into **Amazon S3 Glacier** before it is deleted. Suppose that a user uploads a video to your application and your application generates a thumbnail preview of the video. This video preview is stored to Amazon S3 Standard, because it is likely that the user wants to access it right away.
- Your usage data indicates that most thumbnail previews are not accessed after 30 days. Your lifecycle policy takes these previews and moves them to Amazon S3 – Infrequent Access after 30 days. After another 30 days elapse, the preview is unlikely to be accessed again. The preview is then moved to Amazon S3 Glacier, where it remains for 1 year. After 1 year, the preview is deleted. The important thing is that the lifecycle policy manages all this movement automatically.



Lifecycle policies

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

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

While **Amazon S3** and **Amazon S3 Glacier** are both object storage solutions that enable you to store a virtually unlimited amount of data, they have some critical differences between them. The chart outlines some of these differences.

1. Be careful when you decide which storage solution is correct for your needs. These two services serve very different storage needs. Amazon S3 is designed for frequent, low-latency access to your data, but Amazon S3 Glacier is designed for low-cost, long-term storage of infrequently accessed data.
2. The maximum item size in Amazon S3 is 5 TB, but Amazon S3 Glacier can store items that are up to 40 TB.
3. Because Amazon S3 gives you faster access to your data, the storage cost per gigabyte is higher than it is with Amazon S3 Glacier.
4. While both services have per-request charges, Amazon S3 charges for **PUT, COPY, POST, LIST, GET** operations. In contrast, Amazon S3 Glacier charges for **UPLOAD** and **retrieval** operations.
5. Because Amazon S3 Glacier was designed for less-frequent access to data, it costs more for each retrieval request than Amazon S3.

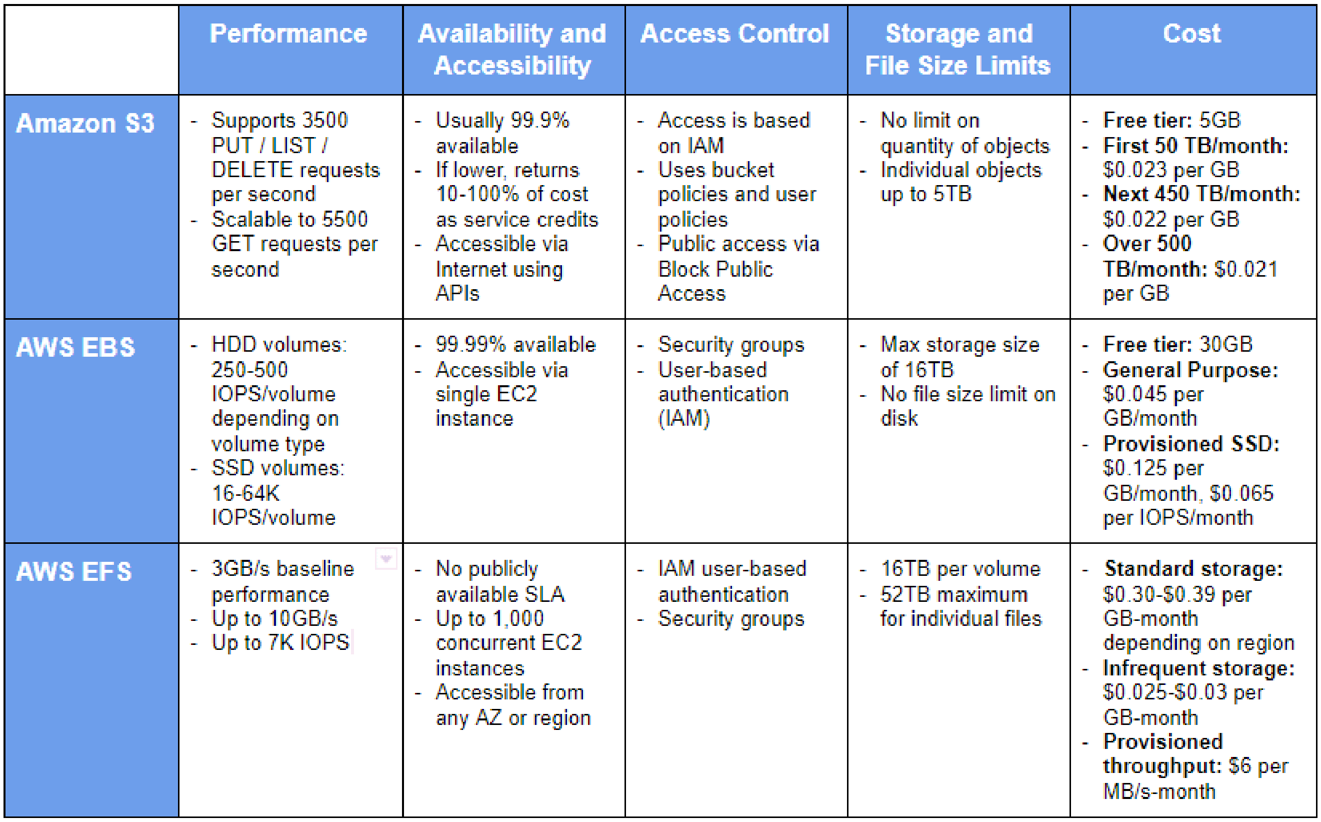
Another important difference between Amazon S3 and Amazon S3 Glacier is how data is encrypted. Server-side encryption is focused on protecting data at rest. With both solutions, you can securely transfer your data over HTTPS. Any data that is archived in Amazon S3 Glacier is encrypted by default. With Amazon S3, your application must initiate server-side encryption. You can accomplish server-side encryption in Amazon S3 in several ways:

- Server-side encryption with **Amazon S3-managed encryption keys (SSE-S3)** employs strong multi-factor encryption. Amazon S3 encrypts each object with a unique key. As an additional safeguard, it encrypts the key with a main key that it regularly rotates. Amazon S3 server-side encryption uses one of the strongest block ciphers available, 256-bit Advanced Encryption Standard (AES-256), to encrypt your data.
- Using server-side encryption **with Customer-provided Encryption Keys (SSE-C)** enables you to set your own encryption keys. You include the encryption key as part of your request, and Amazon S3 manages both encryption (as it writes to disks), and decryption (when you access your objects).
- Using server-side encryption with **AWS Key Management Service (AWS KMS)** is a service that combines secure, highly available hardware and software to provide a key management system that is scaled for the cloud. AWS KMS uses **Customer Master Keys (CMKs)** to encrypt your Amazon S3 objects. You use AWS KMS through the **Encryption Keys** section in the IAM console. You can also access AWS KMS through the API to centrally create encryption keys, define the policies that control how keys can be used, and audit key usage to prove that they are being used correctly. You can use these keys to protect your data in Amazon S3 buckets.

Difference between EFS, EBS, and S3

- Now let’s take a look into the differences between Elastic File System, Elastic Block Storage, and Simple Storage Service
- **Accessibility:** S3 is publically accessible and you do not need a server to access it while EBS can only be accessed via EC2 machine and EFS can be accessed via several EC2 machines and AWS Services.
- **Interface:** S3 provides you a web interface while EBS provides a file system interface and EFS provides a web and file system interface.
- **Storage Type:** AWS S3 is object storage while Amazon EBS is block storage and Amazon EFS is file storage.

- **Scalability:** Amazon S3 and EFS are highly scalable while EBS is hardly scalable.
- **Speed:** Amazon S3 is slowest among the three and comes under tier 3 while Amazon EBS is fastest among the all and comes under tier 0 and Amazon EFS is faster than S3 and slower than EBS and comes under tier 1. So based on the performance hardware will be used in the backend of the storage services
- **Backup:** Amazon S3 is good for storing backups while EBS is meant to be an EC2 drive and since EFS is accessible over multiple servers so it's good for shareable applications and workloads.



	Performance	Availability and Accessibility	Access Control	Storage and File Size Limits	Cost
Amazon S3	- Supports 3500 PUT / LIST / DELETE requests per second - Scalable to 5500 GET requests per second	- Usually 99.9% available - If lower, returns 10-100% of cost as service credits - Accessible via Internet using APIs	- Access is based on IAM - Uses bucket policies and user policies - Public access via Block Public Access	- No limit on quantity of objects - Individual objects up to 5TB	- **Free tier:** 5GB - **First 50 TB/month:** $0.023 per GB - **Next 450 TB/month:** $0.022 per GB - **Over 500 TB/month:** $0.021 per GB
AWS EBS	- HDD volumes: 250-500 IOPS/volume depending on volume type - SSD volumes: 16-64K IOPS/volume	- 99.99% available - Accessible via single EC2 instance	- Security groups - User-based authentication (IAM)	- Max storage size of 16TB - No file size limit on disk	- **Free tier:** 30GB - **General Purpose:** $0.045 per GB/month - **Provisioned SSD:** $0.125 per GB/month, $0.065 per IOPS/month
AWS EFS	- 3GB/s baseline performance - Up to 10GB/s - Up to 7K IOPS	- No publicly available SLA - Up to 1,000 concurrent EC2 instances - Accessible from any AZ or region	- IAM user-based authentication - Security groups	- 16TB per volume - 52TB maximum for individual files	- **Standard storage:** $0.30-$0.39 per GB-month depending on region - **Infrequent storage:** $0.025-$0.03 per GB-month - **Provisioned throughput:** $6 per MB/s-month