



Amazon Cloud Storage Options

Table of Contents

| | |
|---|----|
| 1. Overview of AWS Storage Options..... | 02 |
| 2. Why you should use the AWS Storage..... | 02 |
| 3. How to get Data into the AWS..... | 03 |
| 4. Types of AWS Storage Options..... | 03 |
| 5. Object Storage: Amazon Simple Storage Service (S3) | |
| Benefits..... | 04 |
| 6. Archival Storage: Amazon Glacier | |
| Benefits..... | 06 |
| 7. Block Storage: Elastic Block Storage (EBS) | |
| Benefits..... | 07 |
| 8. Gateway Storage | |
| Benefits..... | 09 |
| 9. File Storage: Amazon Elastic File System | |
| Benefits..... | 10 |
| 10. Conclusion..... | 11 |

AWS Storage Options

Amazon Web Services (AWS) is a flexible, cost-effective, easy-to-use cloud computing platform. This paper is developed to understand the primary data storage options available in the AWS cloud. It provides an overview of each storage option, its basic working principle and benefits.

The Amazon Cloud computing global infrastructure is available to the users across 11 regions, 28 availability zones, and 52 edge locations and is continuously expanding. Under this vast foundation, Storage is a big part of this evolving infrastructure.

A traditional, on-premises IT infrastructure have numerous storage options, which includes memory storage like RAM disks, caches etc, Network storage, Databases, Messaging Queues, Backup and Archives. Using a combination of these storage options, the applications and businesses select the one which satisfies their needs and requirements.

Why you should use the AWS Storage?

Along with the obvious necessity to store your growing data, there are additional benefits which will be applicable to your businesses when using the AWS Storage.

Reduce Costs -It reduces both your CAPEX and OPEX, while simultaneously increasing your scalability, thereby eliminating the need for secondary backup sites.

Reduce On-Premises costs - It allows you to eliminate some of the on-premises equipment and maintenance costs required for storage. You can consolidate the on-premise costs by effectively using the Cloud services.

Change Processes - Sometimes it is necessary to change processes due to your growing business requirements. This requires prior capacity / storage planning and provisioning for peak demand needs. With AWS, you eliminate both these factors. You use what you want and only pay for what you use.

Eliminate/remove aging technologies - The fast evolving technologies will make even the current storage methods redundant. You can eliminate or remove the aging technologies such as old tape archives or disk arrays, and move the data to the more efficient and effective AWS storage options such as S3.

How to get Data into the AWS?

The next question that comes up is how to transfer your data to the Amazon Cloud. There are three ways in which this can be done:



- **Internet** is the obvious choice when you want to transfer data to the AWS. This should be always done over secure and encrypted channels when you are using the public Internet.
- **AWS Direct Connect** is the next choice when your files size constraints or the slow Internet connections prevent you from storing to the AWS. It allows you to have a dedicated bandwidth between your on-premise site and the AWS.
- **AWS Import/Export** is the ideal choice when you want to physically transfer your huge data from your site to the AWS.

Types of AWS Storage Options

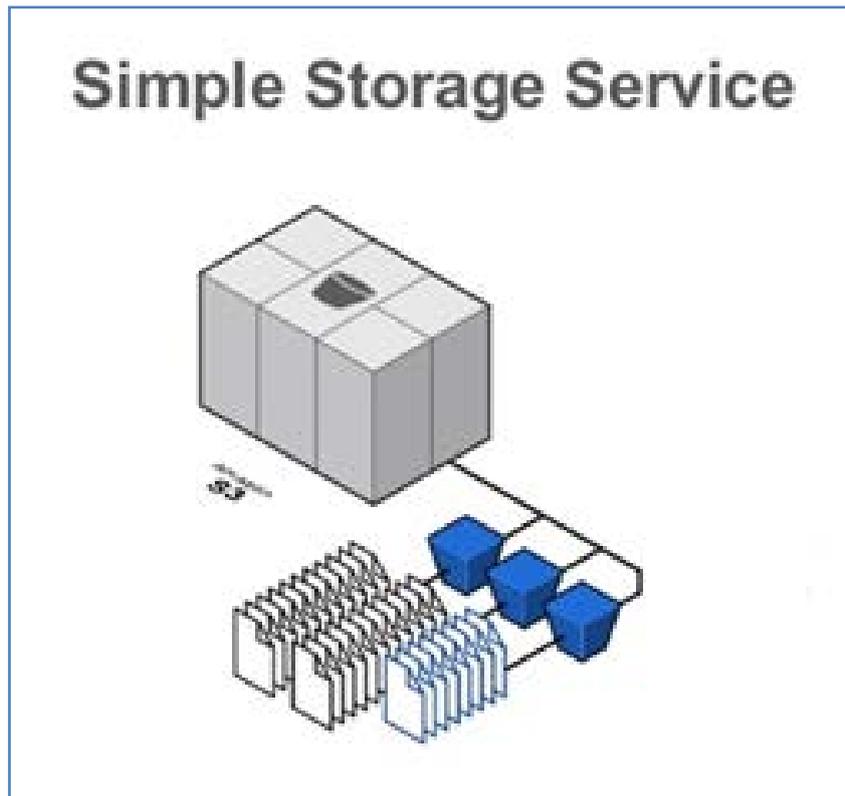
Amazon Cloud Service offers the following storage options to its users

- Object Storage
- Archive Storage
- Block Storage
- Gateway Storage
- File Storage

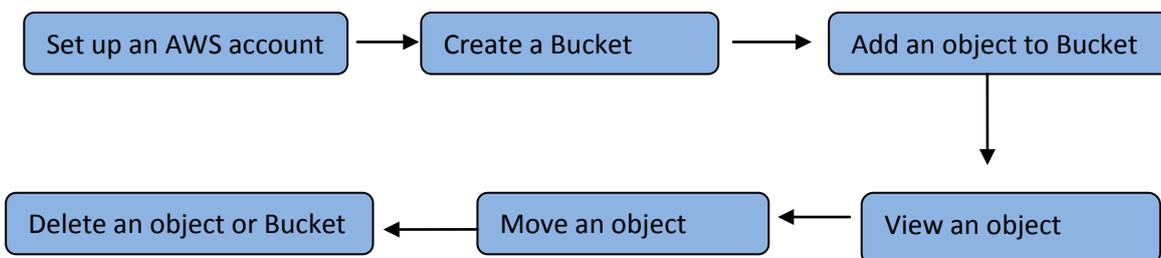
Object Storage: Amazon Simple Storage Service (S3)

Amazon S3 stores data in the form of objects, which are in turn stored in buckets. Each bucket is a container which holds thousands of objects, which represents a file and optionally a description of the file. The action of uploading an object means to store it in the bucket. Unless you specify, only the AWS account user can access the object in the S3 environment. Additionally you can set permissions to limit who can access the objects (to add, delete, modify and update).

Using the AWS Management Console, for each bucket, you can group the objects in folders, for easy management and storage of the objects.



The typical work flow of using AWS Simple Storage Service is as below:



- When creating a bucket, you have to specify the location or Region for storage, which is optimized for latency, minimal costs and which also meets your business requirements. Typically a unique name is assigned to a bucket to differentiate among the thousands of other buckets stored in the Amazon Cloud.

Note: An object stored in a region cannot be moved to another region unless it is explicitly done.

- Adding an object to the bucket is the next step. An object can be any text file, video, image or any other type of file. Using the Management Console, a single or multiple objects can be uploaded to the bucket, in a single action.

Note: When an object is added to a bucket, you can set permissions to prevent unauthorized access to the file.

- Once you have added an object to the bucket, you can either view or move the object to another bucket. Accessing the Amazon S3 console, you can perform the required action.
- Deleting an object or a bucket, is possible only if you have the required permissions. This is permanent action and will require your confirmation. When a bucket is deleted, it will delete all the objects within it.

Benefits of Amazon S3

As a result of its object based storage concept, Amazon S3 is used by companies and users who want to store unstructured data. Each bucket has a defined storage limit of 5 TB, with capability of storing thousands of objects. Each region can store a huge number of buckets and which can be accessed from anywhere in the world.

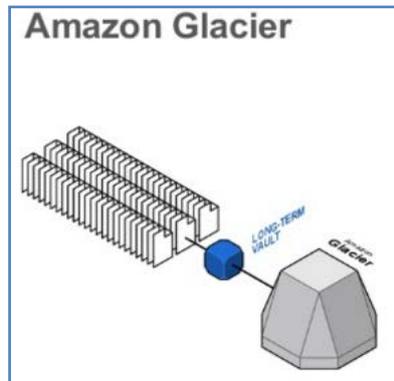
- Unlimited storage capacity
- Scalability
- Performance
- Used for storing static data
- High Availability
- Allows versioning and controlled access with 256-bit encryption
- Ideal for Static assets

When to use Amazon S3

- Web-scale storage capacity and performance for web applications.
- High data durability
- Storage for log files
- Storage for active archives and backup
- Single origin store with delivery through Content Delivery Networks such as Amazon Cloud Front
- Ingestion Points for Big Data applications

Archival Storage: Amazon Glacier

Amazon Glacier is an archival storage option that is designed for long term backup and archival storage. It is a service which is utilized for storage infrequently used data, or "cold data" such as log files, which may be required for auditing and compliance purposes.



The Amazon Glacier service provides a durable and extremely low-cost storage with security features, for data archiving and backup. You can use the AWS Management console to create or delete a vault. This logically represents the storage area inside the AWS environment.

The data can be stored cost effectively for months, years, or even decades. It allows you to eliminate capacity planning, hardware provisioning, data replication, hardware failure detection and recovery, or other time-consuming hardware migrations. It allows you to concentrate on the actual application development without worrying about the administrative burdens of operating and scaling storage to AWS.

Glacier storage is designed for long-term backup and archival storage. As a result, it doesn't require fast data retrieval times which are usually about three to five hours. Due to this, the cost per gigabyte is much less than that of Amazon S3 (It is \$0.01 per gigabyte in the U.S. East region). Glacier pricing is also dependent on the region but it is as much as 90% cheaper than S3. There is no cost for data transferred into Glacier, but the fee for data transferred out is \$0.120 per gigabyte.

There is a link between the Amazon S3 and Amazon Glacier. As Amazon S3 Lifecycle policies allows you to delete or move objects based on time; you can set rules per S3 bucket, regarding moving your objects to the Glacier vault after certain period of time.

For example, the objects can be moved to the Glacier after 30 days or can be deleted after 365 days.

Benefits of Amazon Glacier

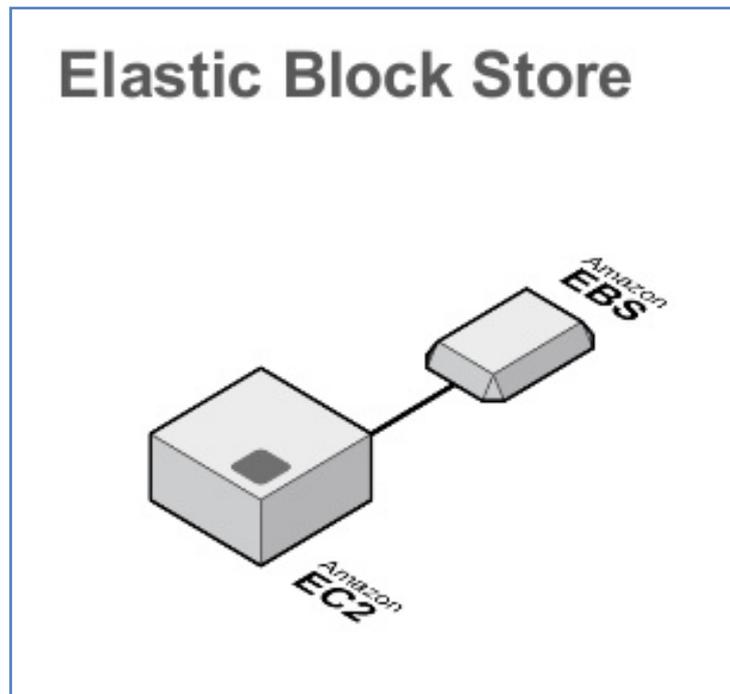
- Low cost per gigabyte which allows storage of large data at a very low cost.
- Durable storage infrastructure where data is stored across multiple facilities and multiple devices in each facility.
- Secure data transfer over SSL and automatic encryption during the process. It also enforces access control using the AWS Identity and Access Management (IAM) service.
- Simple and flexible setup allowing unlimited data storage depending on the type of business requirements and needs.

Block Storage: Elastic Block Storage (EBS)

Amazon Elastic Block Store (Amazon EBS) provides block level storage volumes for use with Amazon EC2 instances in the AWS Cloud. These storage volumes are highly available and reliable, that can be attached to any running instance that is in the same Availability Zone.

The EBS volumes that are attached to an EC2 instance can be used as storage volumes that persist independently from the life of the instance. You can use attach multiple Amazon EBS volumes to a single Amazon EC2 instance. However, at any point in time, only a single Amazon EBS volume may be attached to only one Amazon EC2 instance.

Each EBS volume is automatically replicated within its Availability Zone, to offer high availability and durability in case of any component failure. It offers scalability and consistency of performance, while you pay only for what you provision for or use to store data.



There are three types of Elastic Block Storage volumes.

- **General Purpose (SSD) volumes** are the default EBS volume type for Amazon EC2 instances. They are backed by Solid-State Drives (SSDs) and are suitable for a wide variety of small to medium-sized databases, development and test environments, and boot volumes. It supports up to 10,000 IOPS and 160 MB/s of throughput per volume. I/O is included in the price of General Purpose (SSD) volumes, so you pay only for each GB of storage you provision.
- **Provisioned IOPS (SSD) volumes** are backed by Solid-State Drives (SSDs) and suitable for applications with I/O-intensive workloads such as databases. It supports a consistent baseline performance of up to 30 IOPS/GB to a maximum of 20,000 IOPS, and provide up to 320 MBps of throughput per volume. In addition, you can couple multiple volumes together to achieve up to 48,000 IOPS or 800MBps when attached to larger EC2 instances.

- **Magnetic volumes** provide the lowest cost per GB of all EBS volume types. They are backed by magnetic drives and are ideal for workloads where data access is infrequent and low the storage cost is important. The maximum IOPS per volume is 40 - 200.

* **Note: IOPS = I/O per second for up to 256KB blocks.**

EBS Snapshots

Amazon EBS provides you the ability to take snapshots of EBS volumes to save in the Amazon S3. These snapshots can be used to create new EBS volumes across the AWS regions. They are saved incrementally, meaning only the latest block changes will be saved and you are billed only for the changed blocks.

Since the EBS volumes are created in a specific Availability Zone, they can only be attached to any instance in that same availability zone. EBS snapshots can be used to make a volume to be available outside of the Availability Zone. You can copy snapshots to other regions and restore them to the new volumes.

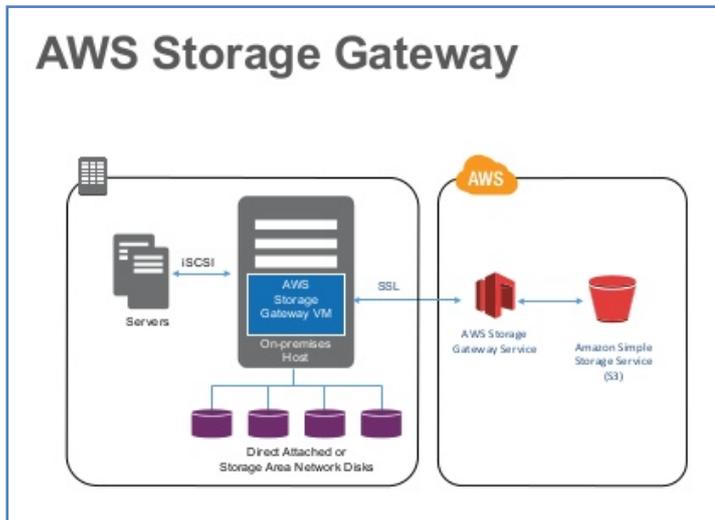
This facilitates easier data migration and disaster recovery and also making it easier for multiple AWS regions for geographical expansion. A large repository of public data set snapshots can be restored to EBS volumes and seamlessly integrated into AWS cloud-based applications.

Benefits of Elastic Block Storage:

- Reliable and secure storage is provided by automatic data replication within its Availability Zone to protect you from component failure.
- Flexible access control and encryption which offers a strong in-depth security strategy for your data.
- Provide consistent and low-latency performance through I/O performance scaled to the needs of your application. You can use multiple volumes together to achieve even higher I/O performance.
- Snapshots (or backups) of any EBS volume and write a copy of the data into the volume to Amazon S3, where it is stored redundantly in multiple Availability Zones. You can boost your business by using Amazon EBS snapshots to create new EC2 instance.

Gateway Storage:

AWS Storage Gateway connects your on-premise software application with cloud-based storage to provide seamless integration with data security features between your on-premises IT environment and the Amazon Web Services (AWS) storage infrastructure. This service provides a cost-effective cloud data storage along with scalability and data security. AWS Storage Gateway offers both volume-based and tape-based storage solutions:



AWS Gateway Solution basically gives you a gateway where you can get access to the Direct attached or storage area network Disks. This data is then encrypted using the SSL encryption methods to be used by the AWS Storage Gateway service, which then transports it to the Amazon S3.

The Gateway Solution supports three different configurations:

- **Gateway Cached Volumes** - In here the data is stored in the Amazon S3 but a local cached copy of the frequently accessed files is maintained on-premises.
- **Gateway-Stored Volumes** - Sometimes you might want to store your primary data locally, while asynchronously backing up that data to AWS. These volumes provide you durable, off-site backups while the on-premises applications have low-latency access to their entire datasets. You can create storage volumes up to 1 TB in size and mount them as iSCSI devices from your on-premises application servers. Data written to your gateway-stored volumes is stored on your on-premises storage hardware, and asynchronously backed up to Amazon S3 in the form of Amazon EBS snapshots.
- **Gateway-Virtual Tape Library (VTL)** - It exposes an industry standard virtual tape library to your preferred backup application. Some of the preferred backup applications might actually write to the tape in your backup data centers. If you write your data to the tape, then you are actually writing to the Storage Gateway service, which implies that the VTL is writing to the Amazon S3 storage.

You can cost-effectively and durably archive backup data in Amazon Glacier. Gateway-VTL provides a virtual tape infrastructure that scales seamlessly with your business needs and eliminates the operational and administrative burden of provisioning, scaling, and maintaining a physical tape infrastructure.

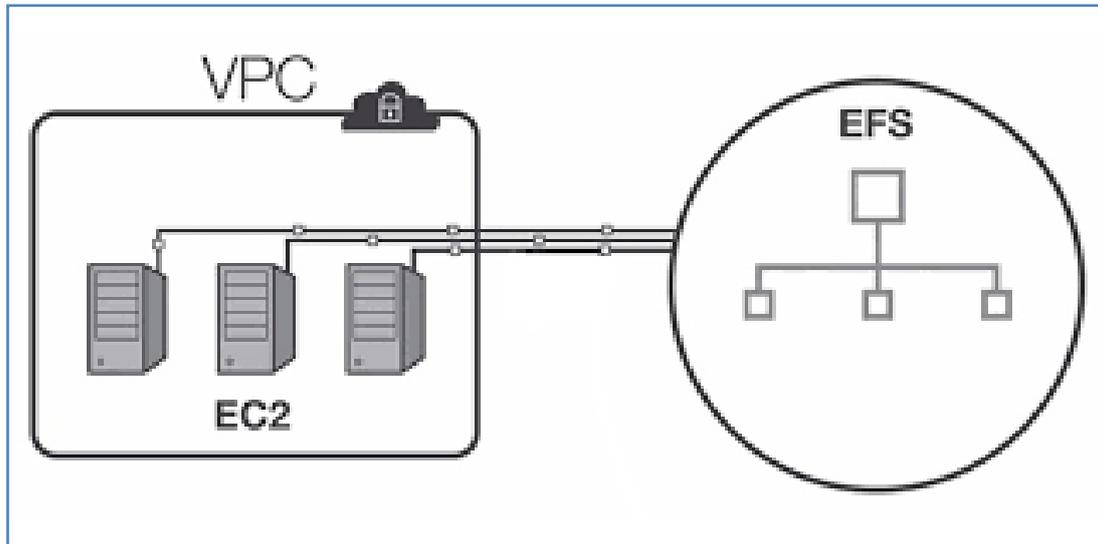
Benefits of Storage Gateway:

- Backup data to Amazon S3.
- Disaster Recovery of applications to EC2.
- Archive data to Amazon Glacier.

File Storage: Amazon Elastic File System

Amazon Elastic File System (EFS) is a cloud storage system where data is stored as a file system. It provides a flexible solid-state disk (SSD)-based storage on demand for Amazon EC2 instances, scaling up and down automatically as your storage needs change. It is shared by multiple EC2 instances.

EFS eliminates the need for administrators to make provision for storage and monitor capacity, but it requires constant review of usage reports to prevent uncontrolled storage expenses when multiple EC2 instances are using the storage.



Elastic File System is simple and easy to setup, requiring no hardware or network layer. It supports the Network File System version 4 (NFSv4) protocol, so that the applications and tools you use integrates seamlessly with the EFS. This protocol is a widely adopted, industry standard protocol that allows a server to share directories and files with clients over a network.

Data is stored in multiple availability zones for high availability, automatic scalability and durability. It provides consistent low latencies, multi-tenancy with high-performance solid-state disks (SSD)

As EFS supports thousands of concurrent EC2 client connections with consistent performance, it is ideal for large data projects, development environment, big data applications and content repositories.

The EFS can be set up and managed through the AWS Management Console, Command Line Interface (CLI) or the Software Development Kits (SDK). Using these ways you can create a file system, create and manage target mounts, delete or tag a file system.

EFS security measures include controlled administrative access, network traffic, file and directory access.

Benefits of Elastic File System

- Seamless integration with existing apps and tools using the NFSv4 protocol.
- Automatic scalability as files are stored/deleted in the system.
- Fully managed services, which manages the file storage infrastructure for you, while avoiding the complexity of deploying, patching, and maintaining the complex file system.
- Consistent and scalable performances through SSDs.
- High availability, cost effective, secure and durable across multiple availability zones and regions.

Conclusion

This paper presents an overview of the AWS storage options in the cloud environment. It highlights their basic working and benefits. Each of these storage options differ in their performance, cost, elasticity, scalability and availability. Based on your business needs and requirements, you can opt for the one which is cost effective, scalable and applicable for your organization.

As with the traditional IT infrastructure, you can use these different cloud storage options together to build a strong data storage infrastructure.