AWS – Storage Service

Poornima S
ASP Sr/ SCOPE

Simple Storage Service(S3)

- Amazon S3 has a simple web services interface that you can use to **store and retrieve any amount of data,** at any time, from anywhere on the web.
- Amazon Simple Storage Service is **storage** for the Internet. It is designed to make **web-scale computing** easier for developers.
- It gives any developer **access** to the same highly **scalable, reliable, fast, inexpensive data storage infrastructure** that Amazon uses to run its own global network of web sites.
- The service **aims to maximize benefits of scale** and to pass those benefits on to developers.

Simple Storage Service(S3)

Data is spread across multiple devices and facilities

- **Think about S3 to store your photos or files.**
- Object Base Storage
- Unlimited Storage
- Files are Stored in Buckets/Folders
- Names must be unique globally
- Every time you have a successful upload you get a http 200 code back

S3 is primarily used for:-

- Store and Backup
- Application File Hosting
- Media Hosting
- Software Delivery
- Storing AMI's and Snapshots

Simple Storage Service(S3)

Objects Consists of the following

- **Key** – this is simply the file name of the object
- **Value** – the data and it is made up of a sequence of bytes.
- **Versioning** – which version of the object is this
- **Meta Data -** Data about data, the data about the data file you are storing.
- Think, if you are storing a music track/song. This would have metadata like the information of the singer, the year it was released, the name of the album etc.

Features of S3

- **Storage classes**
- **Storage management**
  - Lifecycle
  - Object lock
  - Replication
  - Batch operations
- **Access management and security**
  - Block public access
  - IAM
  - Bucket policies
  - Access points
  - Access control lists
  - Object ownership
  - Access analyser

- **Data processing**
  - Object lambda
  - Event notifications

- **Storage logging and monitoring**
  - Automated tools – *cloudwatch, cloudtrail*
  - Manual tools – server access logging, trusted advisor

- **Analytics and insights**
  - Storage lens
  - Class analysis
  - Inventory

- **Strong consistency –** write then read

Simple Storage Service(S3)

S3 Storage Classes

- **S3** (Durable, immediately available and frequently accessed)
- **S3 – IA** (durable, immediately available, infrequently accessed)
- **S3 Reduced Redundancy Storage** (Used for data that is easily reproducible, such thumbnails)
- **S3 Express one zone** – latest – 10 times faster access with reduced cost

- **S3 – IA (Infrequently Accessed)** For data that is accessed less frequently but requires rapid access when needed. This here costs lesser than S3 but you are charged for the retrieval of the data.

- **S3 – RRS (Reduced Redundancy Storage)** Basically less durability with the same level of availability.

e.g. This about data you could potentially regenerate like a tax calculation or a payslip. This is cheaper than the original SS. Suppose you create thumbnails for all your pictures. If you lose a thumbnail you could always regenerate it.

- When **deciding which storage to use think** about the various storage options, their advantages vs disadvantages. Are you optimizing for **durability, the frequency of retrieval or availability**?

How Amazon S3 works?

- Amazon S3 is an **object storage service**, which differs from other types of cloud computing storage types, such as block and file storage.
- Each object is **stored as a file with its metadata** included.
- The object is also given an **ID number**. Applications use this ID number to access objects.
- This is unlike file and block cloud storage, where a developer can **access an object via a representational state transfer (REST) API.**
- The S3 object storage cloud service gives a **subscriber access** to the same systems that Amazon uses to run its own websites. S3 **enables customers to upload, store and download** practically any file or object that is **up to 5 terabytes** (TB) in size -- with the largest single upload capped at 5 gigabytes (GB).
- Every object is encrypted with a unique key.

Simple Storage Service(S3)

- Amazon S3 is **object storage built** to store and retrieve any amount of data from anywhere on the Internet.

- It's a simple storage service that offers an extremely durable, highly available, and infinitely scalable data storage **infrastructure at very low costs**.

- Amazon S3 provides a simple **web service interface** that you can use to store and retrieve any amount of data, at any time, from anywhere on the web.

- Using this web service, you can easily build applications that make use of Internet storage. Since Amazon S3 is **highly scalable** and you only **pay for what you use**, you can start small and grow your application as you wish, with **no compromise on performance or reliability**.

Protecting Data

- Data stored in Amazon S3 is secure by default
- Amazon S3 supports **multiple access control mechanisms**.
- With Amazon S3's data protection features, you can protect your data from both logical and physical **failures**, **guarding against data loss from unintended user actions**, application errors, and infrastructure failures.
- For customers who must **comply with regulatory standards**, Amazon S3's data protection features can be used as part of an overall strategy to achieve compliance.

Data security and reliability features offered by Amazon S3

1. Audit Logs

- Amazon S3 also supports logging of requests made against your Amazon S3 resources. You can configure your Amazon S3 bucket to create access log records for the requests made against it. These server access logs capture all requests made against a bucket or the objects in it and can be used for auditing purposes.
- **CloudWatch** collects metrics, logs, dashboards, alarms, and insights.
- **CloudTrail** records API activity for governance, compliance, and auditing.

2. Versioning

Amazon S3 provides further protection with versioning capability.

- You can use versioning **to preserve, retrieve, and restore every version** of every object stored in your Amazon S3 bucket.
- This allows you to easily recover from both unintended user actions and application failures. By default, requests will retrieve the most recently written version. Older versions of an object can be retrieved by specifying a version in the request.

3. Data Security Details

Amazon S3 provides **four different access control mechanisms:**

- Identity and Access Management (IAM) policies,
- Access Control Lists (ACLs),
- bucket policies,
- query string authentication.

- **IAM** enables organizations with multiple employees to create and manage multiple users under a single AWS account.
- With IAM policies, you can grant IAM users fine-grained control to your Amazon S3 bucket or objects.

- You can use **ACLs** to selectively add (grant) certain permissions on individual objects.

- Amazon S3 **Bucket Policies** can be used to add or deny permissions across some or all of the objects within a single bucket.

- With **Query string authentication**, you have ability to share Amazon S3 objects through URLs that are valid for a predefined expiration time.

Competitor services

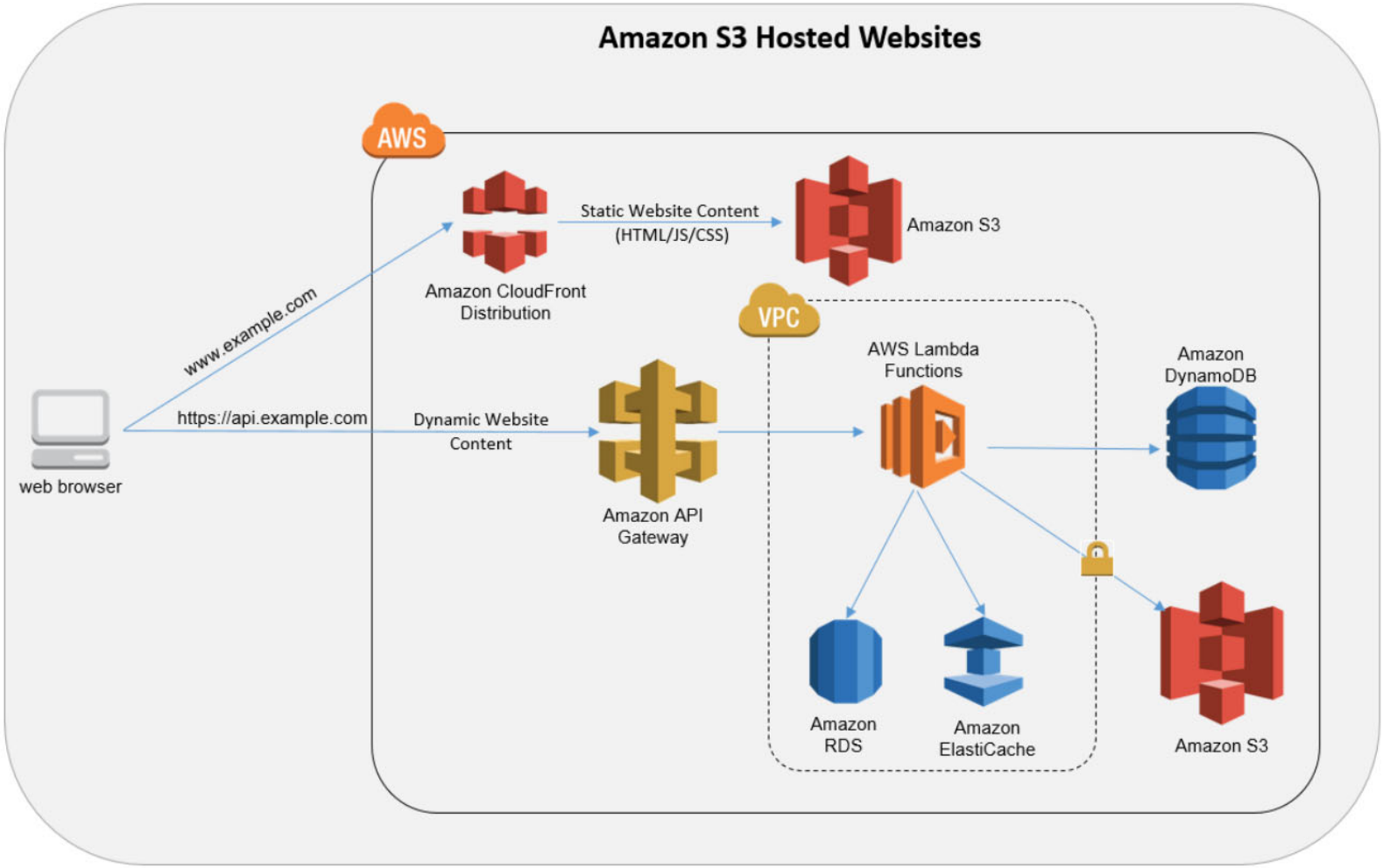
Competitor services to Amazon S3 include other object storage software tool services. Comparable object storage services are offered by other major cloud service providers (CSPs), such as Google, Microsoft, IBM and Alibaba. Main competitor services to Amazon S3 include the following:

- Google Cloud Storage
- Azure Blob storage
- IBM Cloud Object Storage
- DigitalOcean Spaces
- Alibaba Cloud Object Storage Service (OSS)
- Cloudian
- Zadara Storage
- Oracle Cloud Infrastructure Object Storage

Advantages of using Amazon S3

- Amazon S3 is intentionally built with a minimal feature set that focuses on simplicity and robustness. Following are some of the advantages of using Amazon S3:

- **Creating buckets** – Create and name a bucket that stores data. Buckets are the fundamental containers in Amazon S3 for data storage.
- **Storing data** – Store an infinite amount of data in a bucket. Upload as many objects as you like into an Amazon S3 bucket. Each object can contain up to 5 TB of data. Each object is stored and retrieved using a unique developer-assigned key.
- **Downloading data** – Download your data or enable others to do so. Download your data anytime you like, or allow others to do the same.
- **Permissions** – Grant or deny access to others who want to upload or download data into your Amazon S3 bucket. Grant upload and download permissions to three types of users. Authentication mechanisms can help keep data secure from unauthorized access.
- **Standard interfaces** – Use standards-based REST and SOAP interfaces designed to work with any internet-development toolkit.


Amazon S3 Hosted Websites
AWS
Amazon CloudFront Distribution
Static Website Content (HTML/JS/CSS)
Amazon S3
www.example.com
VPC
AWS Lambda Functions
Amazon DynamoDB
web browser
https://api.example.com
Dynamic Website Content
Amazon API Gateway
Amazon RDS
Amazon ElastiCache
Amazon S3

Amazon Glacier

- S3 Glacier is an extremely **low-cost storage service** that provides durable storage with security features for **data archiving and backup.**
- With S3 Glacier, customers can store their data cost effectively for months, years, or even decades.
- S3 Glacier enables customers to **offload the administrative burdens of operating and scaling** storage to AWS, so they don't have to worry about capacity planning, hardware provisioning, data replication, hardware failure detection and recovery, or time-consuming hardware migrations

- Refer Getting started using the Amazon S3 Glacier storage classes | Amazon Web Services

Amazon Glacier

- A developer uses a cold data cloud service such as Amazon Glacier to move **infrequently accessed data to archival storage** to save money on storage costs. He can also move database backups from tape storage media to the cloud for long-term Glacier storage.

Archives and vaults

- A vault is a **container** for storing archives.
- Amazon Glacier stores data in archives and vaults.
- An archive is a block of data that may consist of a single file or aggregated data in the form of **TAR or zip files**.
- Glacier archives range in size from **1 byte to 40 terabytes** (TB); there are no limits to how much data and how many archives an AWS user can store in Glacier. Amazon offers a multipart upload feature for higher throughput and reliability for archives over 100 MB.

S3 Vs Glacier

	S3	**Glacier**
1	Amazon S3 is a durable, secure, simple, and fast storage service	Amazon Glacier is used for archiving solutions
2	Excellent support of Identity and access management makes sure that your objects are used by the intended audience only.	Cheap storage of backup data In Glacier, users create archives and vaults
3	Entire service is accessible through commands and API. It makes integration and orchestration of S3 services really easy and automated.	Can be used as a part of the entire suite of tools from Amazon, without requiring you to leave the familiar stack
4	S3 has quite good reliability and durability to make sure all requests are successfully fulfilled and your objects are always safe.	Accessing data stored in Glacier is slow. That shouldn't be a surprise, but it is undesirable nonetheless.

S3 Vs Glacier

	S3	**Glacier**
5	This service and UI interface is easy to understand.	Retrieving a large amount of data can be expensive; Glacier's intended use is as an archive of rarely-accessed data.
6	It takes almost no learning curve to consume this service.	Some users regard Glacier with fear and uncertainty.
7	The amount of flexibility it provides, its really cost-effective.	Slow retrieval time and high retrieval cost are the greatest risks of using Glacier, and they are also the Glacier interaction that most users have the least experience with.

Also refer
https://tutorialsdojo.com/amazon-s3-vs-glacier/

S3 Lifecycle Management

Using Amazon S3 lifecycle configuration rules,

- One can **significantly reduce your storage costs by automatically transitioning data** from one storage class to another or even automatically deleting data after a period of time.
- **For example**, the **lifecycle rules for backup data** might be:
  - Store backup data initially in Amazon S3 Standard.
  - After 30 days, transition to Amazon Standard-IA.
  - After 90 days, transition to Amazon Glacier.
  - After 3 years, delete.
- Lifecycle configurations are attached to the bucket and can apply to all objects in the bucket or only to objects specified by a prefix.

EFS / EBS/ S3

Feature	EFS	EBS	S3
Type	File storage	Block storage	Object storage
Access	Multiple EC2 instances	Single EC2 instance	Web-based
Use Case	Shared file systems	Databases, OS	Backups, media, big data
Protocol	NFS	Attached as a volume	REST API

Self Assessment

- Why EBS is preferred over instance store volumes for running a database on an EC2 instance and need persistent storage?
- A university needs to archive student records for 10 years. How does Amazon Glacier help in achieving cost-effective long-term storage?
- Why would you choose Amazon EFS over EBS or S3 when multiple EC2 instances need to access the same set of files simultaneously

References

- https://docs.aws.amazon.com/AmazonS3/latest/userguide/Welcome.html
- https://www.educative.io/courses/learn-the-a-to-z-of-amazon-web-services-aws/7nO7GQL4xj8
- https://aws.amazon.com/s3/
- https://searchaws.techtarget.com/definition/Amazon-Simple-Storage-Service-Amazon-S3