

Simple Storage Service (S3)

S3 overview

Basically, a place where you can upload files, S3 is object-based storage, allows you to upload files. You can't install or run anything on S3. Not suitable to install an OS or run a database on.

Let's start with S3. S3 is an object storage service for the cloud that is highly available. Objects are stored in buckets. So instead of files and directories, you have objects and buckets. S3 gives you essentially unlimited storage because it can hold millions of objects per bucket. Objects can be public or private, and you can upload objects via the console, the CLI, or through code. S3 Objects can range in size from a min of 0 bytes to a max of 5 TB in size.

- So public access is granted to buckets and objects in those buckets through access control lists, bucket policies, access point policies, or all 3. You can enable versioning to create multiple versions of your objects, to protect against accidental deletion, or to roll back to a previous version. You can use S3 access logs to log all the access to S3 buckets and objects. So you can see the root cause of issues or track down any suspicious activity. And S3 is a regional service, but the bucket name is globally unique. So this means, after a bucket is created, the name of that bucket cannot be used by anyone else in any AWS Region, unless that bucket is deleted.
- While Amazon S3 buckets must, as we mentioned earlier, exist within a single region, S3 is nevertheless considered a global service (open the S3 Console page and look at the region indicator).
- While Amazon S3 buckets must, as we mentioned earlier, exist within a single region, S3 is nevertheless considered a global service (open the S3 Console page and look at the region indicator).

Let's say you're hosting large media files in S3 buckets. If users would have to retrieve their files directly from the bucket each time they were requested, delivery—specially to end users living continents away from the bucket location—would be relatively slow. But if you could store cached copies of the most popular files on servers located geographically close to your users, then they wouldn't have to wait for the original file to be retrieved but could be enjoying the cached copy in a fraction of the time.

- Retention settings range from 1 day to 10 years. You can also archive your logs by exporting them to an S3 bucket.
- You can use S3 to store any kind of file. Although AWS calls it "storage for the internet," you can implement access controls and encryption to restrict access to your files to specific individuals and IP addresses. S3 opens up a variety of uses, both inside and outside of AWS. Many AWS services use S3 to store logs or retrieve data for processing, such as with analytics. You can even use S3 to host static websites!
- Amazon S3 Standard - is designed for 99.99% availability. How can AWS offer 11 9's of durability and 99.99% availability for S3 Standard.
 - Default storage class
 - Perfect for frequently access data
 - Use cases include websites, content distribution, mobile and gaming apps, and big data analytics.
- S3 Standard-IA
 - Rapid access – used for data that is access less frequently but requires rapid access when needed
 - Low per-GB storage price and a per-GB retrieval fee

- 30 days minimum duration charge per object
 - Great for long-term storage, backups, and as a data store for disaster recover files
 - **99.99% availability**
- S3 One Zone – IA
 - 20% less than standard-IA
 - This isn't a good idea as this is stored within a single AZ
 - Great for long-lived, infrequent access non-critical data
- Data is only stored in 1 availability zone. Which cost 20 % less S3 standard. Data can be lost since it is stored in 1 AZ. This is recommended that can be easily recreated.
- **S3 Glacier**
 - Offers long-term archiving of infrequently accessed data, such as backups that must be retained for many years.
 - Pay for each time you access the data
 - 3 Options listed below
 - 1-5 mins
 - 3-5 hours
 - 5-12 hours
 - This is recommended for long-term backups and cheaper storage options
- **S3 Glacier Deep Archive**
 - **Cheapest of all options due to longer access times**
 - 2 options for retrieval times
 - 12 hours or 48 hours
 - Recommended for long-term data archival accessed once or twice a year
- S3 Outpost
 - Provided storage on-premises
 - A single storage class
 - Recommended for data that needs to be kept local
- S3 Intelligent-Tiering
 - Automatically moves your data to the most cost-effective tier based on how frequently you access each object

NOTE

> Across the S3 storage classes all of them have the 99.99% durability (11 9's)

	S3 Standard	S3 Intelligent-Tiering*	S3 Standard-IA	S3 One Zone-IA+	S3 Glacier	S3 Glacier Deep Archive
Designed for Durability	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)	99.999999999% (11 9's)
Designed for Availability	99.99%	99.9%	99.9%	99.5%	99.99%	99.99%
Availability SLA	99.9%	99%	99%	99%	99.9%	99.9%
Availability Zone(s)	≥3	≥3	≥3	1	≥3	≥3
Min Capacity Charge per Object	N/A	N/A	128KB	128KB	40KB	40KB
Minimum Storage Duration Charge	N/A	30 days	30 days	30 days	90 days	180 days
Retrieval Fee	N/A	N/A	Per GB retrieved	Per GB retrieved	Per GB retrieved	Per GB retrieved
Storage Type	Object	Object	Object	Object	Object	Object
Lifecycle Transitions	Yes	Yes	Yes	Yes	Yes	Yes

S3 Storage Classes - Storage Costs	Storage Pricing
S3 Standard - General-purpose storage for any type of data, typically used for frequently accessed data	
First 50 TB / Month	\$0.023 per GB
Next 450 TB / Month	\$0.022 per GB
Over 500 TB / Month	\$0.021 per GB
S3 Intelligent-Tiering* - Automatic cost savings for data with unknown or changing access patterns	
Frequent Access Tier, First 50 TB / Month	\$0.023 per GB
Frequent Access Tier, Next 450 TB / Month	\$0.022 per GB
Frequent Access Tier, Over 500 TB / Month	\$0.021 per GB
Frequent Access Tier, All Storage / Month	\$0.0125 per GB
Monitoring and Automation, All Storage / Month	\$0.0025 per 1,000 objects
S3 Standard-Infrequent Access * - For long-lived but infrequently accessed data that needs millisecond access	
All Storage / Month	\$0.0125 per GB
S3 One Zone-Infrequent Access * - For re-createable infrequently accessed data that needs millisecond access	
All Storage / Month	\$0.01 per GB
S3 Glacier ** - For long-term backups and archives with retrieval options from 1 minute to 12 hours	
All Storage / Month	\$0.004 per GB
S3 Glacier Deep Archive ** - For long-term data archiving that is accessed once or twice in a year and can be restored within 12 hours	
All Storage / Month	\$0.00099 per GB

^ ABOVE

Is the cost from highest to lowest

Study the retrieval options, storage, etc

Storage Class	Availability and Durability	AZ(s)	Use Case
S3 Standard	99.99% Availability 11 9's Durability	>=3	Suitable for most workloads (e.g., websites, content distribution, mobile and gaming applications, and big data analytics)
S3 Standard-Infrequent Access	99.9% Availability 11 9's Durability	>=3	Long-term, infrequently accessed critical data (e.g., backups, data store for disaster recovery files, etc.)
S3 One Zone-Infrequent Access	99.5% Availability 11 9's Durability	1	Long-term, infrequently accessed, non-critical data
S3 Glacier	99.99% Availability 11 9's Durability	>=3	Long-term data archiving that occasionally needs to be accessed within a few hours or minutes
S3 Glacier Deep Archive	99.99% Availability 11 9's Durability	>=3	Rarely accessed data archiving with a default retrieval time of 12 hours (e.g., financial records for regulatory purposes)
S3 Intelligent-Tiering	99.9% Availability 11 9's Durability	>=3	Unknown or unpredictable access patterns

^ABOVE***Exam Tips, just remember the use cases for each S3 class***

S3 in the real world

1. Static websites

- a. Use static websites to s3 and user CloudFront for global distribution
- 2. Data Archive
 - . Amazon Glacier as a storage option
- 3. Analytics Systems
 - . Store data in S3 for use with services like redshift and Athena
- 4. Mobile applications
 - . Users can upload files to an S3 bucket

Uploading Files

When you upload a file to an S3 bucket, you will receive an HTTP 200 code if the upload was successful

- Strong Read-After-Write Consistency (As soon as you written something to S3 it should be immediately available)
 - After a successful write
 - of a new object (PUT) or an overwrite of an existing object, and subsequent read request immediately receives the latest version of the object
 - Strong Consistency
 - For list operations, so after a write, you can immediately perform a listing of objects in a bucket with all changes reflected.



Object ACL – Go down on an individual object public using object ACLs

Bucket Policy – If you want to make it public, you want to use this. Apply on an entire bucket level

Hosting a static website using S3



- You can use S3 to host static websites, such as .html sites
- Static Content
 - You can use S3 to host static content only (not dynamic)

❖ Versioning Objects in S3

- Advantages

The infographic lists five advantages of S3 Versioning, each preceded by a checkmark icon in a circle:

- ✓ **All Versions**
All versions of an object are stored in S3. This includes all writes and even if you delete an object.
- ✓ **Backup**
Can be a great backup tool.
- ✓ **Cannot Be Disabled**
Once enabled, versioning cannot be disabled — only suspended.
- ✓ **Lifecycle Rules**
Can be integrated with lifecycle rules.
- ✓ **Supports MFA**
Can support multi-factor authentication.

>Lifecycle management with S3

Lifecycle management automates moving your objects between the different storage tiers, thereby maximizing cost effectiveness

- Keep for 30 days > S3 Standard
- After 30 days > S3 IA
- After 90 days > Glacier

>S3 Object Lock and Glacier Vault lock

- **S3 Object lock** (Comes in 2 modes Governance and Compliance)
 - Use this to store objects using a **write once, read many (WORM)** model. It can help prevent objects from being deleted or modified for a fixed amount of time or indefinitely.
 - Can be applied on individual objects or applied across the bucket as a whole.
 - **Governance Mode**
 - Users can't overwrite or delete an object version or alter its lock settings unless they have special permissions.
 - Made to protect you
 - **Compliance Mode**
 - Can't be overwritten or deleted by any user, even the root user can't during the duration of the retention period.
- Retention Periods
 - Protects an object version for a fixed amount of time. Object version can be overwritten or deleted by any user
- Legal holds
 - Prevents an object version from being overwritten or deleted
 - Legal holds can be freely placed and removed by any user who has
 - **S3:PutObjectLegalHold** permission
- **S3 Glacier Vault Lock**
 - Easily deploy and enforce compliance controls for individual S3 glacier vaults with a vault lock policy.
 - Once locked, the policy can no longer be changed
 - **WORM models with Glacier, think of scenario questions for this.**

>Encrypting S3 Objects

- Types of encryption
 - Encryption in transit
 - **SSL/TLS**
 - **HTTPS**
 - Encryption at Rest: **Server-Side encryption**
 - **SSE-S3**: S3-managed keys, using AES 256-bit encryption
 - **SSE-KMS**: AWS Key Management Service-managed keys

- **SSE-C**: Customer-provided keys

You can enforce encryption through your console and have it through bucket policies. You can create a bucket policy that denies any S3 PUT request that doesn't include the x-amz-server-side-encryption parameter in the request header.

df

- Encryption at rest: Client-Side encryption
 - You encrypt the files yourself before you upload them to S3

>Optimizing S3 Performance

- S3 Prefixes is just folders inside our buckets. The more prefixes you have, the better performance you get.
- S3 has extremely low latency. You can get the first byte out of S3 within 100-200 milliseconds
 - To get better performance
 - Spread the reads different directories or prefixes
 - If you are using 2 prefixes, you can achieve 11,000 request per second
 - The more folders and subfolders the better performance you will get
- Limitations within KMS (Key management service)
 - When you upload/download a file will count towards the KMS quota
 - Region-Specific, however, it's either 5,500, 10,000 or 30k request per second
 - Currently you cannot request a quota increase for KMS

>Backing up Data with S3 replication

- S3 Replication
 - You can replicate objects from one bucket to another
 - Versioning must be enabled on both the source and destination buckets
 - Delete markers are not replicated by default

Tips for Exam

- S3 is object-based storage
 - Upload and store files
- Not OS or DB storage
 - Not suitable to install an OS or run a database on
- Files up to 5 TB
 - Files can be from 0 bytes to 5 TB
- Unlimited Storage
 - The total volume of data and the number of objects you can store is unlimited.
- Files are stored in buckets and S3 is a universal space, has to be unique that anyone else in the world
- Successful CLI or API uploads will generate an HTTP 200 status code
- Key
 - The object name
- Value

- The data itself, which is made up of a sequence of bytes
- Version ID
 - Allows you to store multiple versions of the same object
- Metadata
 - Data about the data you are storing
- Use multipart uploads to increase performance when uploading files to S3
- S3 largest object that can be stored should be used for any files over 100MB and must be used for any file over 5 GB
- Use S3 byte-range fetches to increase performance when downloading files to S3
- Once versioning on your bucket is suspended, and someone deletes it. There is no way of restoring it. As it is not possible since versioning was suspended. The object is gone.
- S3 buckets are not public, they are private by default.
- Lifecycle policies can't work backwards. You can use a lifecycle policy to migrate objects from the more frequently accessed storage classes to the longer-term options, but not the other way around.
- Availability is the ability to access your data, while durability is the ability of AWS to ensure your data is properly stored in S3.