AWS

Simple Storage Service (Amazon S3)

what is it?

An S3 bucket is where objects are stored, similar to files and folders on your local machine. There is unlimited storage available, across 100 buckets, and files can be from 0 bytes to 5TB.

Each object consists of:

- **Key** - the name of the object
- **Value** - the data in the file itself made of bytes
- **VersionID**
- **Metadata**

when to use it?

- **Analytics / Data Lake** - Uncouple storage and compute to scale either up or down as needed using Amazon Athena as the query service over the top and AWS Glue as a data catalogue.
- **Archive** - When data goes from 'hot', frequently accessed, to 'cold', infrequently accessed, it can be moved to Amazon Glacier for a more cost-effective option.
- **Data Staging** - Temporary data storage before being loading into AWS Redshift.
- **Static website** - Host a website using S3 for storage and Route 53 as the DNS.

AWS

Simple Storage Service (Amazon S3)

storage options

S3

- The most expensive as it promises 11 9's of durability.
- Good for cloud apps, big data, websites, content distribution.

S3:IA

- Costs 50% less than standard S3 as availability is reduced
- Recommended for non-critical data that **CANNOT** be easily reproduced and needs to be retrieved quickly
- Good for disaster recovery, backups

S3:IA - One Zone

- Costs less than S3:IA as durability is reduced
- Recommended for non-critical data that **CAN** be easily reproduced and needs to be retrieved quickly
- Good for secondary backups as objects are only stored in one zone

Glacier

- Much cheaper as there is a 3 - 5 hour retrieval time
- Good for long term storage, archives and 'cold' data

Deep Glacier

- The cheapest option as there is a 12 hour retrieval time
- Used for documents that need to be kept for compliance reasons for 7+ years

AWS

Simple Storage Service (Amazon S3)

security

- S3 is secure by default and each new bucket and the objects in it are private.
- To keep objects even more secure you can use bucket policies, similar to IAM policies, to fine tune access.
- Presigned URLs are another option to provide security when temporary access to an object is required.
- A URL is generated via the AWS CLI and SDK which can then be used to provide temporary access to write or download object data.

encryption

Client side

This is when the client encrypts the objects and uploads to Amazon S3.

Server side

This is when the data is encrypted when written and decrypts when it is being used.

- **SSE-AES** - S3 handles the key using the AES-256 algorithm
- **SSE-KMS** - Envelope encryption via AWS KMS, you manage keys
- **SSE-C** - Customer provided key, you manage the keys

AWS

Simple Storage Service (Amazon S3)

versioning

When versioning is turned on deleted files have a delete tag added which hides the file. To restore the file, delete the tag.

Key things to know:

- Each version takes up storage space. So a 1GB file edited three times with versioning on takes up 3GB of space
- Once turned on, versioning can only be suspended, not removed
- Versions that are deleted on the other hand are actually deleted
- Enabling multi factor authentication gives extra protection from accidental deletion

replication

Replication enables automatic, asynchronous copying of objects across Amazon S3 buckets. Buckets that are configured for object replication can be owned by the same AWS account or by different accounts. Objects can be replicated across regions or within the same region.

Key things to know:

- **Cross-Region replication** (**CRR**) is used to copy objects across Amazon S3 buckets in different AWS Regions
- **Same-Region replication** (**SRR**) is used to copy objects across Amazon S3 buckets in the same AWS Region
- Existing files won't be copied until there's been a new version, which will also replicate all previous versions and permissions