1. What is Google Cloud Storage and what are its storage classes?

Answer:

GCS is a scalable, durable, and secure object storage service provided by Google Cloud for storing unstructured data like files, images, backups, and logs.

Object Storage Service, is a service provided by Google Cloud Services. That allows to you stores and retrieve any amount of data at any time. It is designed for developers and enterprises to store unstructured data.

Storage classes:

- **Standard** – For frequently accessed data.
- **Nearline** – For infrequently accessed data (once a month).
- **Coldline** – For rarely accessed data (once a quarter).
- **Archive** – For long-term, rarely accessed data.

A. Can you change the storage class of an object?

-- Yes, using lifecycle rules or manually through the gsutil rewrite command or GCP console.

B. What is the minimum storage duration for Nearline/Coldline/Archive?

-- Nearline: 30 days, Coldline: 90 days, Archive: 365 days.

2. How do you upload and download data to/from GCS?

Answer:

You can upload/download using:

- GCP Console (UI)
- gsutil CLI
- APIs

A. What is gsutil cp command?

-- It copies files to/from GCS buckets, e.g., gsutil cp file.txt gs://my-bucket/

B. How do you make uploads faster for large files?

--Use parallel composite uploads or resumable uploads.

3. How can you automate data transfer to GCS?

Answer:

You can use:

- Cloud Storage Transfer Service
- Scheduled gsutil scripts
- Cloud Functions or Workflows
- Cloud Composer (Airflow)

A. Can you move data from AWS S3 to GCS?

--Yes, using Storage Transfer Service or gsutil.

B. How would you schedule daily ingestion from GCS to BigQuery?

-- Use Cloud Composer, Scheduled Queries, or a Cloud Function trigger**.**

4. What's the difference between a bucket and an object in GCS?

Answer:

Bucket: A container that holds objects and metadata.

Object: The actual data file along with its metadata.

A. Can you have nested folders in GCS?

--GCS is a flat namespace, but simulates folder structure using slashes (/) in object names.

B. do you list all objects under a folder?

--gsutil ls gs://bucket/folder-name/**

5. A large file upload fails halfway. How would you handle it?

Answer:

- Use resumable uploads
- Check logs and resume from the failure point

- If a large file upload fails halfway, I would use **resumable uploads** in GCS. This allows the upload to resume from the point of failure without starting over, saving time and bandwidth. You initiate the upload, get a session URI, and resume using that URI. For command-line, gsutil handles this automatically with large files.

6. How would you manage hundreds of TBs of data in GCS while keeping costs low?
Expected Topics:

- Use proper storage class
- Lifecycle rules
- Compression before upload

Operation	gsutil Command	gcloud Command
Create Bucket	gsutil mb gs://my-bucket	gcloud storage buckets create my-bucket
List Buckets	gsutil ls	gcloud storage buckets list
List Bucket Contents	gsutil ls gs://my-bucket/	gcloud storage objects list --bucket=my-bucket
Upload File	gsutil cp file.txt gs://my-bucket/	gcloud storage cp file.txt gs://my-bucket/
Download File	gsutil cp gs://my-bucket/file.txt ./	gcloud storage cp gs://my-bucket/file.txt ./
Copy Between Buckets	gsutil cp gs://bucket-a/file gs://bucket-b/	gcloud storage cp gs://bucket-a/file gs://bucket-b/
Delete File	gsutil rm gs://my-bucket/file.txt	gcloud storage rm gs://my-bucket/file.txt
Delete Bucket	gsutil rb gs://my-bucket/	gcloud storage buckets delete my-bucket
View File Metadata	gsutil ls -L gs://my-bucket/file.txt	gcloud storage objects describe gs://my-bucket/file.txt