Cloud Storage Documentation: <https://cloud.google.com/storage/docs/>

Cloud Storage

-stage input into a relational database which is Cloud SQL, into BigQuery which is a data warehouse, or into Dataproc which is a Hadoop cluster.  So cloud storage is blob storage.

 So you are basically storing raw data in any format, directly onto cloud storage. In order to get there though, normally what you might want to do is to get this data from somewhere else.

 It may be in your data center, it could be on instruments out in the field, it could be logs that are being created. Tends to be that you're basically ingesting extracting this data. You're doing some processing, and that processing could happen on a compute engine. So you could have a compute engine VM and you're basically doing a whole bunch of processing. And if you need to do processing, you need very fast seeks, reads and writes of the data. And a good way to do that is to basically store that data on disk. The problem is that any disk, although you have persistent disks,

 let's ignore that for now. Typical disks are associated with the compute engine that they're attached to. When the compute engine goes away, the disk also goes away.

To use a command line tool called gsutil, you need to install G cloud SDK.

Interacting with cloud storage:

**gsutil cp sales\*.csv gs://acme-sales/data/**

Commands:

copy **cp**

remove **rm**

move **mv**

list **ls**

make a bucket **mb**

remove a bucket **rb**

mirror of something local in the cloud **rsync**

rsync only uploads files that have changed

access control list (permissions) acl **acl**

Can set up transfer service (one time or re-occurring)

GCP gives you durability, reliability, and global reach

--Versioning, redundancy, edge-caching

Control access to Cloud Storage at that object level (bucket level, every bucket belongs to a project, which billing account will be responsible for paying for it, add and remove files from bucket)

Access control can be given to authenticated users and all users (don’t have to be logged in, just give http url).

Control latency and availability with zones and region

Choose the closest zone/region so as to reduce latency.

Distribute your apps and data across zones to reduce service disruptions (might run application in multiple applications).

Distribute your apps and data across regions for global availability.

Compute nodes on GCP are:

--allocated on demand, and you pay for the time that they are up.

--cheaper if you allow them to be shutdown at any time (preemtible VM).

GCS is a good option for storing data that:

(NO. not good for real-time from sensors and other devices, not high throughput)

--may be required to be read at some later time (persistent storage).

--may be imported into a cluster for analysis.

Compute Engine: <https://cloud.google.com/compute/>

Storage: <https://cloud.google.com/storage/>

Pricing calculator

<https://cloud.google.com/pricing/>

Cloud launcher

<https://cloud.google.com/launcher/>

Pricing Philosophy

<https://cloud.google.com/pricing/philosophy/>