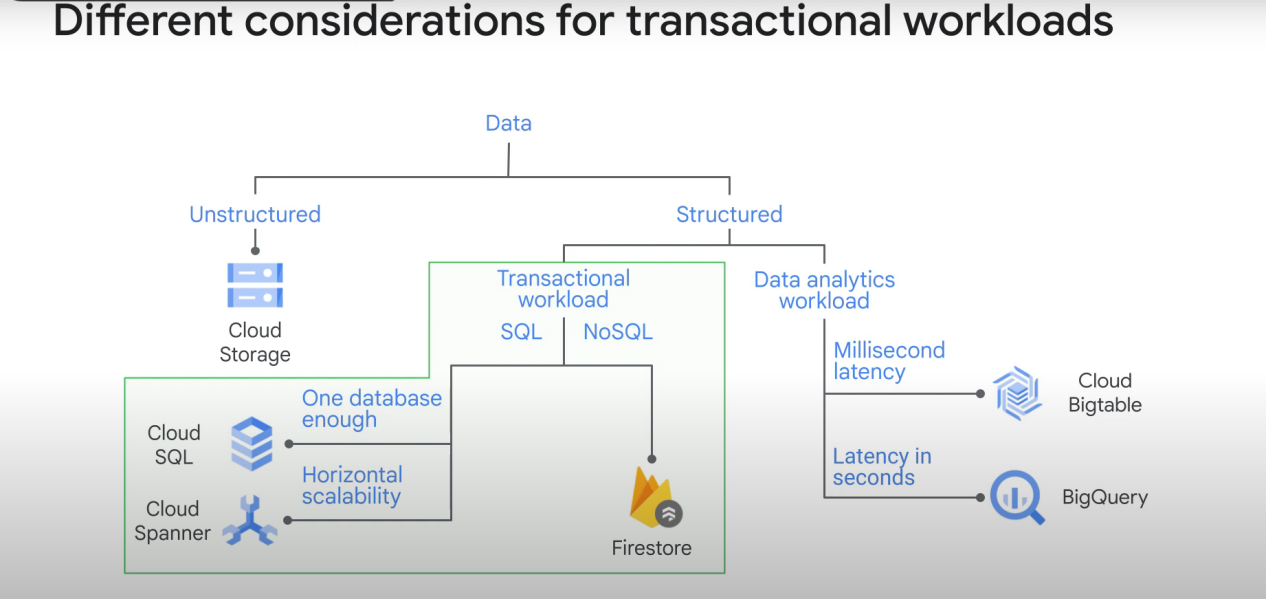
You can use the following as a data lake:

- Cloud Storage (general), good for unstructured data

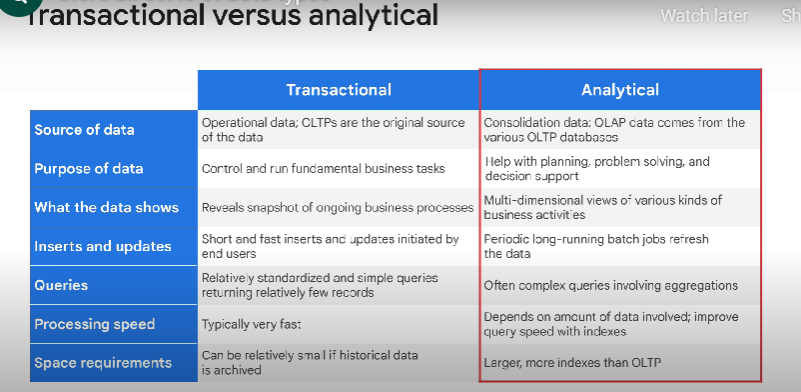
- Could SQL (Relational DB)

- Cloud Spanner if a lot of data (Relational DB)

- Firestore/Cloud bigtable (NoSQL)



Cloud Bigtable is mainly for high throuput streaming (large quantity of data)



Transactional: 80% write operation 20% read. Keeps a current snapshot of the system state (bank account system, e-commerce website, etc...) (OLTP)

Analytical: 20% Write, 80% read. (analyze large amount of data, for instance to get statistics about monthly users activity in a website) (Called OLAP)

Transferring to data warehouse:

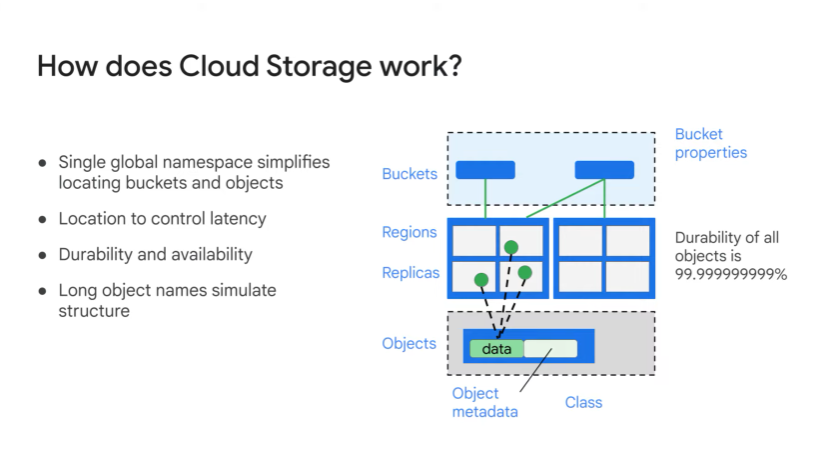
EL: Extract and Load, no need for transform. For instance, bigquery can directly read avro files. No need to transform the data, you just need to load it in bigquery.

ETL: extract, load, transform == we need to transform/Clean the data before loading it to the data

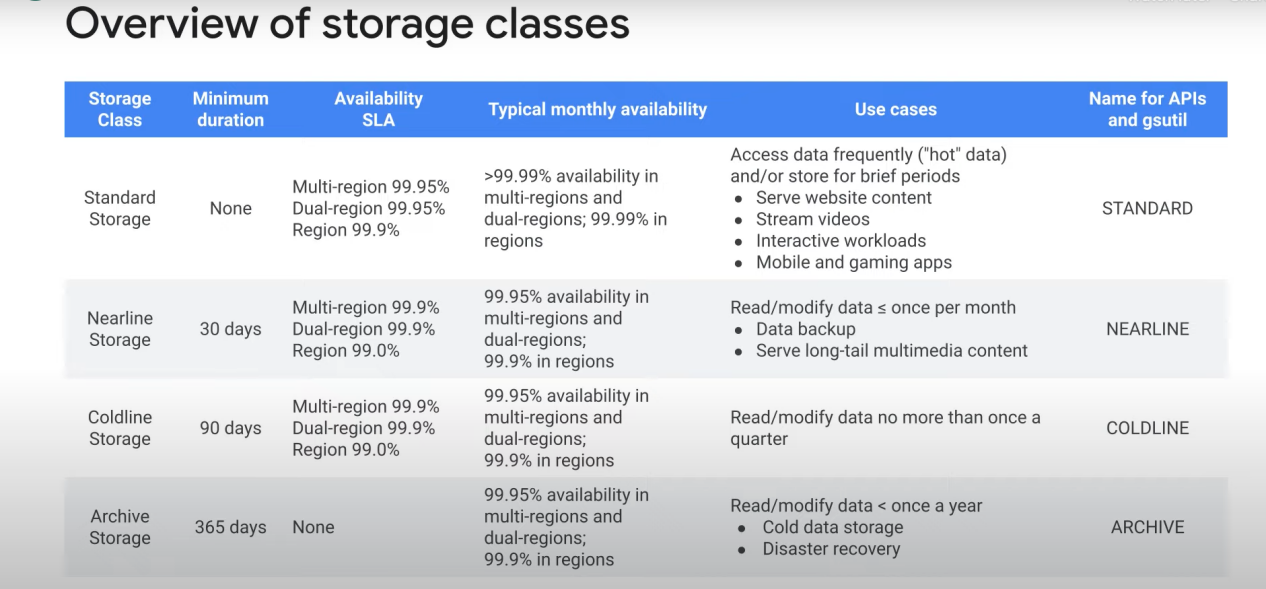
sink.

ELT: same, but we transform the data after loading it in the cloud/data sink. We may use this if our transformation is not going to reduce considerably the size of the data for instance.

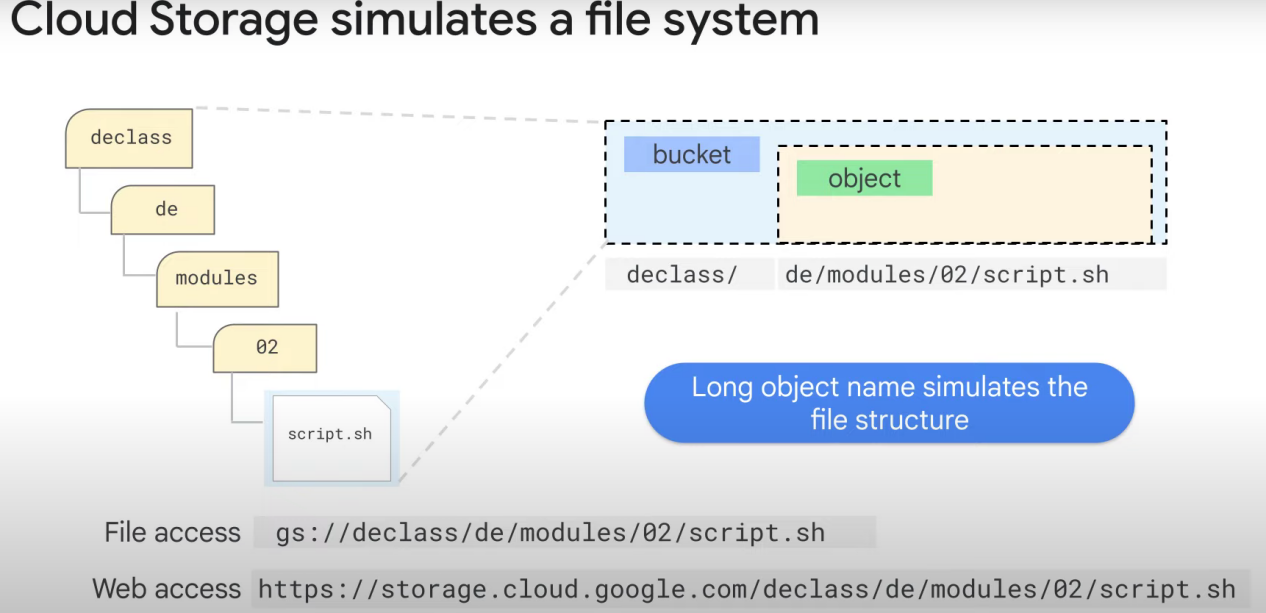
**Cloud storage:**



Types of classes in cloud storage



1 object = 1 class I think. And you can automatically move objects from one class to another if it hasn’t been used for more than n days.



Note: Serverless = fully managed + other advantages like dynamic cost management