**BigQuery data ingestion**

Different platforms to access data from BigQuery.

* File upload
* Google cloud storage
  + We can create a storage bucket, upload the file there and connect to BigQuery
* Google sheets
  + If we have data in Google drive in form of google sheets, we can use that as an external table.
* Data flow(ETL operation)

**How does BigQuery store data?**

****

BigQuery stores data in column format called Capacitor(file format)

1. **Capacitor**
   * + Stores type of data, how the data is being used, reshuffle rows and columns.
     + Example : Run length encoding

|  |
| --- |
| Q1:4 |
| Q12:2 |

|  |
| --- |
| 0:2 |
| 2:1 |
| 1:3 |

|  |  |
| --- | --- |
| Q1 | 0 |
| Q1 | 0 |
| Q1 | 2 |
| Q1 | 1 |
| Q2 | 1 |
| Q2 | 1 |

* Capacitor**:** reorders the rows to obtain a compact encoding.

Where the data get stored

1. **Colossus:** The **distributed file system** provided by Google.

* **Colossus** has a ton of disks that live on a ton of servers but there may be a disadvantage (if the number increases some of disks can be failed on each day).
* **What Colossus does?**
  + - Uses **erasure encoding** which breaks data into fragments and saves pieces across a set of different disks.
    - By using multi region location like US or EU, BigQuery stores a copy of data. Which makes data to be recoverable in the event of disaster.
  + Colossus protects data by ensuring 100% encrypted.

A best practice when optimizing costs is to keep your data in BigQuery. Rather than exporting your older data to another storage option (such as Cloud Storage), take advantage of BigQuery’s long-term storage pricing.

Note: Add external data to BigQuery through cloud storage (use Bucketing)

**Load data …from cloud storage….**

Create a project

Create a table (drive, cloud storage..)

Go to Cloud storage…….Create a new Bucket(name your bucket…..choose other options)

Upload file in **Bucket**

**Writing or Save query results**

https://cloud.google.com/bigquery/docs/writing-results