**[BigQuery](https://cloud.google.com/bigquery) :**

<https://cloud.google.com/bigquery/docs/quickstarts/quickstart-cloud-console?_ga=2.38825827.-1345905635.1631520685>

<https://cloud.google.com/bigquery/docs/schemas#python>

<https://cloud.google.com/bigquery/docs/information-schema-tables#bq> #Filter columns using Except statement

<https://cloud.google.com/blog/topics/developers-practitioners/bigquery-explained-blog-series>

<https://towardsdatascience.com/a-simple-way-to-query-table-metadata-in-google-bigquery-92dc7f1edec1>

<https://medium.com/google-cloud/bigquery-explained-querying-your-data-9e017f2714a3>

Reference:

<https://cloud.google.com/bigquery/docs/reference/standard-sql/query-syntax#select_except>

GitHub :

<https://github.com/googleapis/python-bigquery/tree/master>

<https://github.com/googleapis/python-bigquery/tree/master/samples>

<https://hevodata.com/blog/api-to-bigquery/>

<https://github.com/GoogleCloudPlatform/training-data-analyst/tree/master/courses/data-engineering/demos>

**BigData ecosystem:**

**Input Data:**

1. Compute storage
2. Cloud sql
3. Cloud spanner
4. Cloud datastore
5. Cloud bigtable

**Ingestion layer:**

1. Cloud pub/sub
2. Compute data transfer
3. Cloud ioT care
4. Cloud firestore

**Processing layer:**

1. Cloud dataflow
2. Cloud dataproc
3. Cloud dataprep

**Storage layer:**

1. Compute storage
2. Big query
3. Cloud bigtable

**Visualization:**

1. Google sheets
2. Tableau
3. Cloud data studio
4. Cloud datalab
5. Looker

**Features of Big query:**

1. Big query GIS
2. Auto backup
3. Integration with other GCP services
4. Foundation for business intelligence
5. Programmatic interaction
6. High security
7. Repo of logs
8. Federated queries
9. Run ds workloads using storage API
10. Powerful data repositories

Region or Multiregion?

* Choose region if the data would be accessed from only that region or nearby.
  + Region stores/back-up data in regional data centers only, hence resilient to soft failures only.
* Choose multi-region if data would be accessed from multiple far distanced places.
  + Multi-region stores/backup data in geographically separated regions, hence resilient to soft & hard failures.
* Incase of querying data externally, dataset be created in same region or multi-region as of the data location.

Failure Types:

* Soft failures (power failure, network partition)
* Hard failures (damage from floods, earth quakes)

Quotas and limits:

* Can create unlimited datasets per project but approaching 1000 datasets in a project, dashboard performance begin to degrade.
* Unlimited tables in a dataset but approaching 50,000 tables becomes slower.
* Max 2500 authorized view to a dataset’s control list.

Table names are unique for the dataset.

Charge is done in big query is for querying the table and how much data is stored and is partitioned.

Table creation is free.

Limited to 1500 operations/table/day.

Maximum columns in a table, query result or view cab ne 10,000.

Big query engine is for batch processing.

When it is for stream data, cloud dataengine will come forward.

Caching features & Limits:

* To retrieve data from stored cache, the query should be exact replica of the original query.
* Not cached when destination table is specified to store the results in query.
* Not cached if tables/views being used in the query have changed since the last cache.
* Not cached for tables having streaming ingestion and also current\_timestamp, now, current\_user.
* Not cached if the query run against external data sources like bigtable and cloudstorage.
* Less than 10gb

Wildcard tables:

* It enables to query multiple tables at once using concise SQL statements.

**Introduction to SQL for BigQuery and Cloud SQL:**

* <https://google.qwiklabs.com/focuses/2802?parent=catalog>