What is an external table?

Any data or file system which is not a part of or reside in BigQuery is called external data. This document explains how to query data stored outside of BigQuery.

Data sources associated with External table

Following data sources are associated with external table.

* *Bigtable*
* *Cloud storage*
* *Drive*

To create an external file:

External file can be a permanent table or a temporary table

Create a permanent table

A permanent external table is contained inside a dataset, and can manage it in the same way as we do query with a standard BigQuery table.

Data sources associated with External table

We can create a file from the following external data sources.

* *Bigtable*
* *Cloud storage*
  + CSV file
  + ORC file
  + JSON
* *Drive*
  + CSV
  + JSON
  + Google sheets

Create a table file for an external data source

Create a project

Under the project name Create dataset

* Dataset ID
* Data location
* Create **dataset**

Create a table under dataset

* Source
  + Create table from > Google cloud storage



**External table limitations**

* BigQuery does not guarantee data consistency for external data tables.
* Query performance is not high as querying data in native BigQuery table.
* The performance of the query depends on external storage type.
  + Querying data stored in Cloud storage is faster than querying data stored in Google Drive. This means the query performance differs for querying data from Cloud storage as compared to Google drive storage data.
* External tables are for read only purpose. We cannot modify external tables using DML statements.
* We cannot reference an external table in a wildcard table query.
  + Wildcard table
    - A wildcard table represents a union of all the tables that match the wildcard expression.
    - Example :
    - select FROM  
        `bigquery-public-data.noaa\_gsod.gsod\*`
    - For each row in the wildcard table contains a specific column \_TABLE\_SUFFIX, which contains the value matched by the wildcard character.
    - Limitations
* External tables do not support clustering. They support partitioning in limited ways.
* BigQuery writes all query results to a table. Other than cloud storage, when do we query external tables, the results are not catched.
* Querying data in Cloud is only available for specific regions and zones,
* There is a limitation on number of queries to run. We are limited to 4 concurrent queries against a Cloud Bigtable external data source.