Assignment and Exam Content

Cloud BigQuery

Cloud Gurus Seattle, USA
Training material

Always Delete your Cloud Resources to Avoid $$ Charges.

Cloud BigQuery Lab

Cloud BigQuery Lab Contains following topics

A Create dataset

B Create Tables , Add Data , Retrieve data.

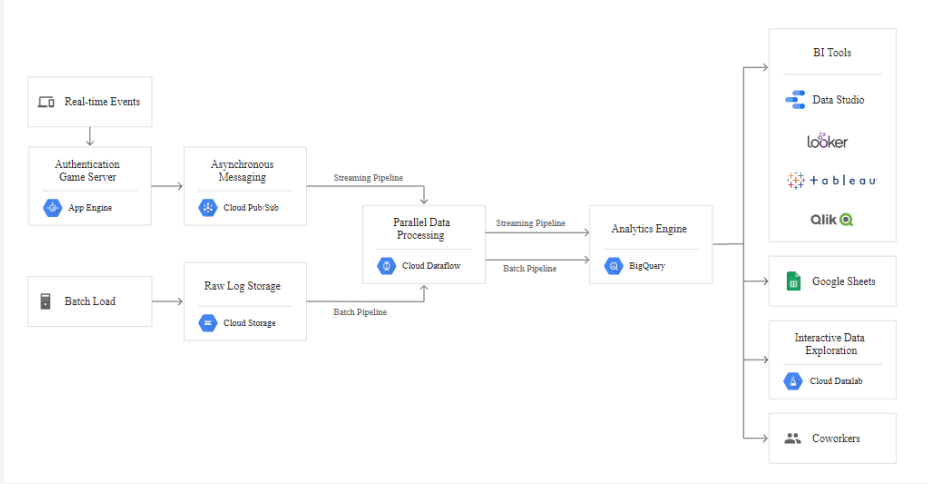
C Exam TIPS

Cloud Gurus Seattle, USA
Training material

Cloud BigQuery is relational data warehouse systems out of Google Cloud Platform

BigQuery is a serverless, highly-scalable, and cost-effective cloud data warehouse with an in-memory BI Engine and machine learning built in.

You don't have to provision any cloud infrastructure .its fully managed service by GCP.
https://cloud.google.com/bigquery/#focus-on-the-analytics-not-your-infrastructure



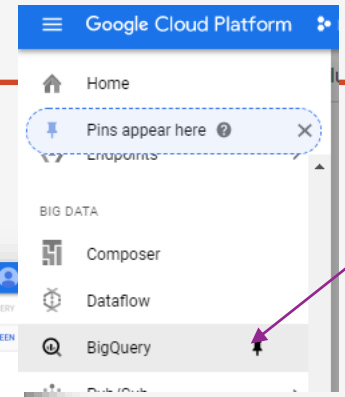
Always Delete your Cloud Resources to Avoid $$ Charges.

A

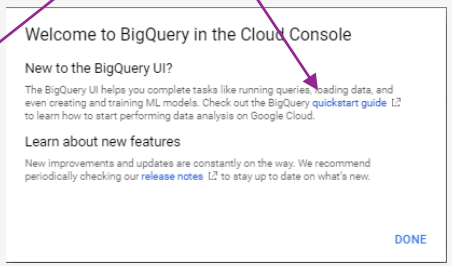
Create BigQuery Dataset

1 Go To -> BIGDATA -> BigQuery

There is not dataset - When you 1st see this screen.



Cloud BigQuery

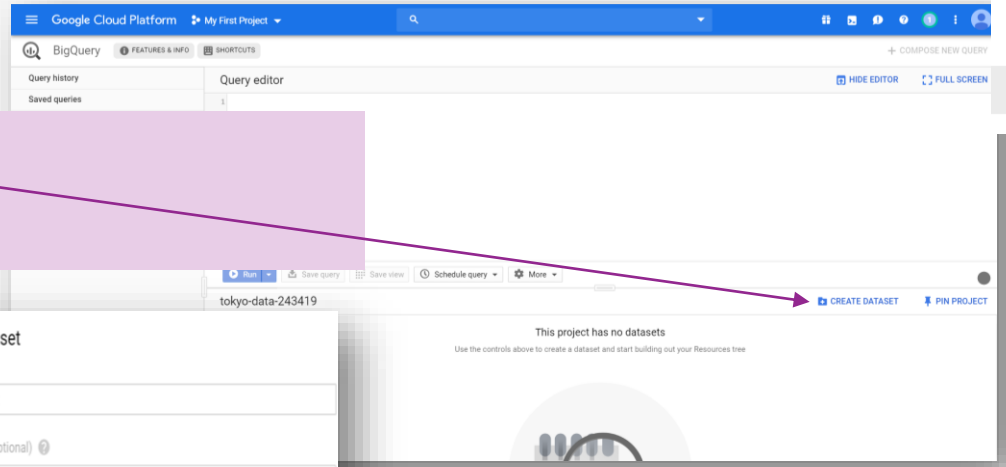


2

Create Dataset

Dataset is kind of database in traditional world

You can add tables into dataset



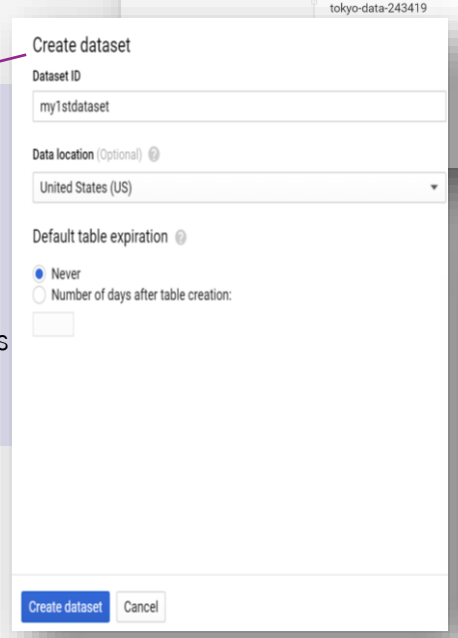
Dataset Creation

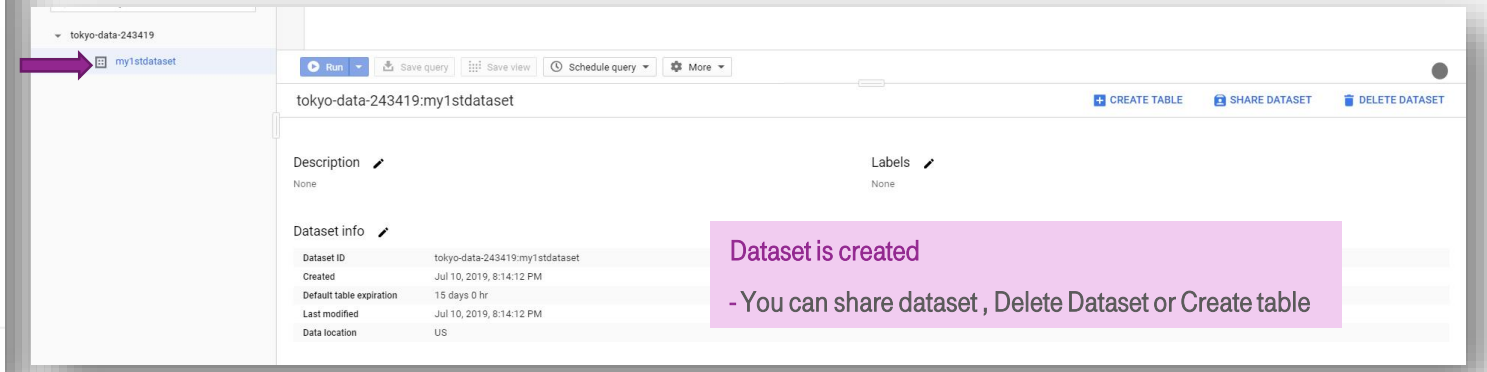
Choose name of dataset

Data Location – your nearest location

Default table expirations

You can provide deletions for any tables as default and BigQuery will manage the deletions of table.





Dataset is created

- You can share dataset , Delete Dataset or Create table

A

Create BigQuery Table

3

Create Table

BigQuery table is like any other RDBMS table.

It can be created is variety of ways in BigQuery

Hit Create Table on Dataset to open Create table UI.

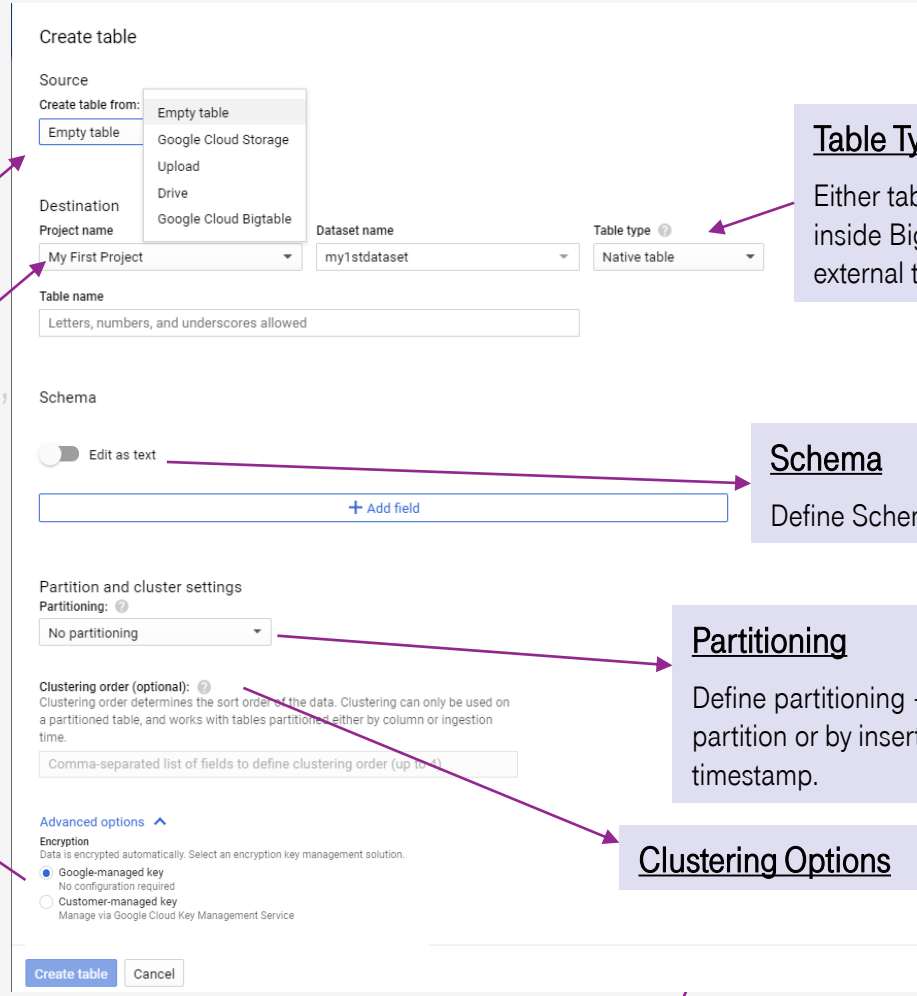


Table Type

Either table stored inside BigQuery or external to BigQuery

Source

You can create either empty table or create it using existing file on cloud storage, on your computer , Google Drive, or even table from Bigtable.

Schema

Define Schema

Destination

Choose – where you want to create table – Project Name -> Dataset name

Partitioning

Define partitioning – no partition or by insert timestamp.

Encryptions options.

Clustering Options

Provide your Choice and create at least one table

Choose – where you want to create table – Project Name -> Dataset name

A

Try something with bg utility : DataSet -> Sample Commands

bq is used as command line utility for BigQuery operations

Important for exams. – Purpose of bq.

Load Files

$ bq –location=US load –source_format=CSV mydataset.mytable ./myfile.csv qtr:STRING,sales:FLOAT,year:STRING
$ bq –location=asia-northeast1 load –source_format=CSV mydataset.mytable ./myfile.csv qtr:STRING,sales:FLOAT,year:STRING
$ bq mk –table [PROJECT_ID]:[DATASET].[TABLE] [SCHEMA]
$ bq mk –table mydataset.mytable qtr:STRING,sales:FLOAT,year:STRING
$ bq mk –table mydataset.mytable ./myschema.json

Schema

bq show –schema –format=prettyjson [PROJECT_ID]:[DATASET].[TABLE] > [PATH_TO_FILE]

Schema Form Json
bq –location=US load –source_format=CSV mydataset.mytable ./myfile.csv ./myschema.json

Autodetect schema
bq –location=US load –autodetect –source_format=CSV mydataset.mytable ./myfile.csv

Update Schema
Step 1. Get Schema
bq show –schema –format=prettyjson mydataset.mytable > /tmp/myschema.json
Step 2 : Modify the schema

Step 3. Apply Schema
bq update mydataset.mytable /tmp/myschema.json
bq update –description "Description of mytable" mydataset.mytable
bq update –description "Description of mytable" myotherproject:mydataset.mytable

Change table Expiration time
bq update –expiration 432000 mydataset.mytable

Browsing Table
bq head –max_rows=10 mydataset.mytable
bq head myotherproject:mydataset.mytable
bq head –start_row 100 –selected_fields "field1,field2" mydataset.mytable

Exporting Data
bq –location=US extract –compression GZIP 'mydataset.mytable' gs://example-bucket/myfile.csv
bq –location=US extract –destination_format NEWLINE_DELIMITED_JSON 'mydataset.mytable' gs://example-bucket/myfile.json
bq –location=US extract –destination_format AVRO –compression SNAPPY 'mydataset.mytable' gs://example-bucket/myfile.avro

Copy Table
bq –location=US cp mydataset.mytable mydataset2.mytable2
bq –location=asia-northeast1 cp mydataset.mytable mydataset2.mytable2

Delete
bq rm -t myotherproject:mydataset.mytable
bq rm -t mydataset.mytable

A

Try something with bg utility : Tables -> Sample Commands

4

Create database

bq –location=US mk -d –default_table_expiration 3600 –description "This is my dataset." mydataset

bq is used as command line utility for BigQuery operations

Important for exams. – Purpose of bq.

Listing Dataset

Listing

bq ls –format=prettyjson –project_id
bq ls –format=prettyjson

bq show –format=prettyjson [PROJECT_ID]:[DATASET]
bq query –nouse_legacy_sql \
'SELECT * EXCEPT(schema_owner) FROM INFORMATION_SCHEMA.SCHEMATA'

Update Dataset

bq update –description "Description of mydataset" mydataset
bq update –default_table_expiration 7200 mydataset

Remove

bq rm -r -f -d [PROJECT_ID]:[DATASET]

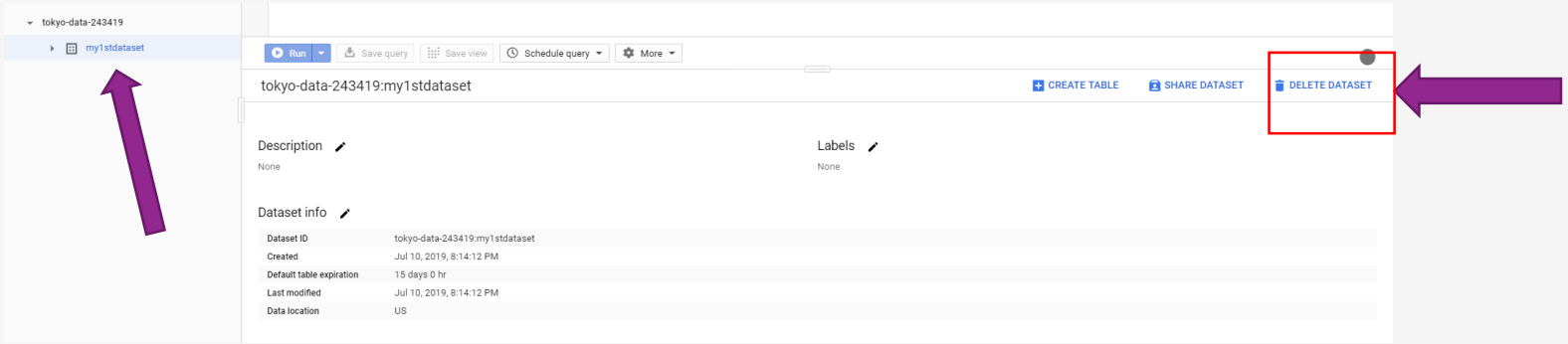
Access Control Access.

bq show –format=prettyjson [PROJECT_ID]:[DATASET] > [PATH_TO_FILE]
bq update –source [PATH_TO_FILE] [PROJECT_ID]:[DATASET]

2

Cloud BigQuery : Delete Dataset

1



Always delete data set

Even though the GCP Charge for small data is very small but it can get considerable if you store it for longer period without use - Delete Dataset....

Always Delete your Cloud Resources to Avoid $$ Charges.

Cloud BigQuery : Try Yourself

1. Create different dataset and create tables using different source files

   Dry run Queries to get cost of the Query -> bq parameter. -> https://cloud.google.com/bigquery/docs/best-practices-costs

   Try Some QuickStarts https://cloud.google.com/bigquery/docs/quickstarts

2. **Exam Tips**

Important concepts are

1. **Which Command utility used -> bq.**
2. **Realtime low latency data access vs batch jobs -> BigQuery cant be used for low latency use -: Low latency used –always prefer BigTable.**
3. **Use cases e.g. Machine Learning data store with SQL**
4. **Log Analysis for long term – use BigQuery sink**
5. **Billing Analysis for large companies..**
6. **Cost Optimizations – Using estimation tool ->** https://cloud.google.com/bigquery/docs/best-practices-costs

Cloud Gurus Seattle, USA
Training material

End of Cloud BigQuery Assignment