BigQuery in Action

Learn How to Use Google Cloud BigQuery for Data Analytics and Data Warehousing

Dan Sullivan

Lead Solutions Architect, Hydrolix
Author of study guides for Google Cloud certifications

Pearson

Audience Poll Question

What is your experience level with databases and SQL? (Single response)

- Expert level understanding of databases and SQL
- Experienced with databases and SQL
- Limited experience with SQL

Lesson 1: Overview of BigQuery

- Understand the advantages of serverless data warehousing
- Become familiar with the BigQuery Console
- Know key BigQuery concepts, including Projects, Datasets, and Tables
- Know how to view a schema structure and understand the structures and data types used in a table
- Know the role of Jobs in BigQuery

BigQuery

- Serverless, petabyte scale data warehouse
- Uses SQL but is not a relational database
- Analytical database
- Other features
  - Machine learning
  - BI Engine
  - GIS
  - BigQuery Omni

Organizational Structure

- Projects
- Datasets
- Tables and Views
- Rows and Columns

Query Interface

RUN | SAVE | SCHEDULE | MORE | This query will process 559.8 KiB when run.

```
1  select
2    *
3  from
4    `bigquery-public-data.geo_international_ports.world_port_index`
5  limit 20
```

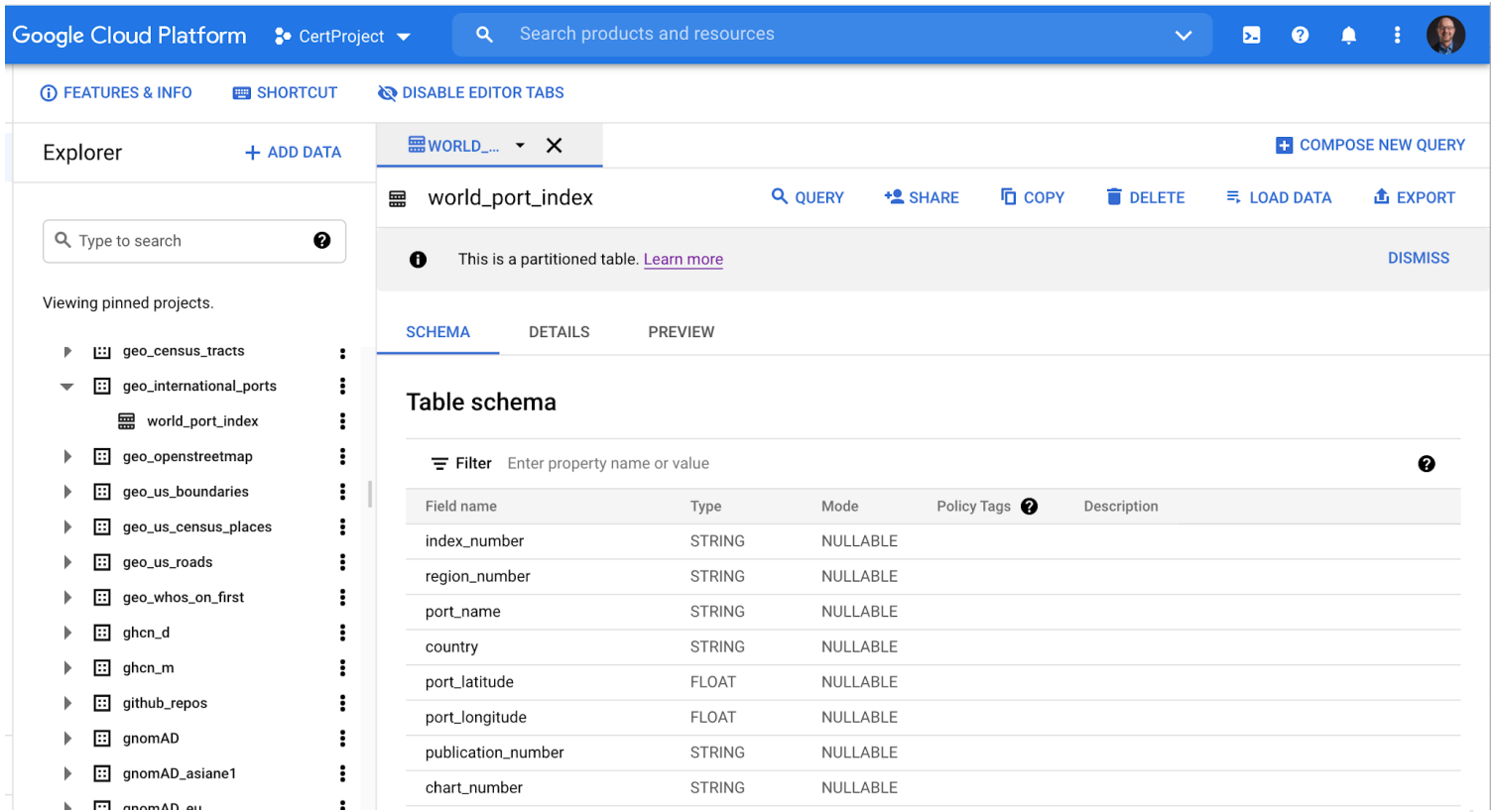
Query results | SAVE RESULTS | EXPLORE DATA

Query complete (0.3 sec elapsed, 559.8 KB processed)

Job information | Results | JSON | Execution details

Row	index_number	region_number	port_name	country	port_latitude	port_longitude	publication_number	chart_number	harbor_size
1	50922	50920	MERAK MAS TERMINAL	ID	-5.916667	105.983333	150	*null*	M
2	47855	47840	DORALEH	DJ	11.6	43.083333	172	*null*	S
3	8822	8650	SOUTHWEST PASS	US	28.45	-90.7	CP05	*null*	S
4	2200	1930	PORT DE BECANCOUR	CA	46.4	-72.383333	SCOR	*null*	V
5	62635	62530	PORT BERINGOVSKY	RU	63.066667	179.366667	155	*null*	*null*
6	61245	61100	ISHIKARI BAY NEW PORT	JP	43.216667	141.283333	159	*null*	M

Schemas



Pearson

Federated Data

- Federated query
  - Bigtable
  - Cloud SQL
  - Cloud Storage
    - Parquet
    - ORC
  - Google Drive
    - CSV
    - JSON
    - Avro
    - Sheets

Lesson 2: Creating Tables, Loading Data, and Linking Datasets

- Create tables in BigQuery
- Load data into BigQuery tables
- Preview data in a BigQuery table
- Link data from data Analytics Hub

Lesson 2: Exercise

- Login to your Google Cloud account
- Open the BigQuery console
- Verify you are using the project you want to use
- Use the Add Data option to link the NYC Citi Bike Trips dataset
- Preview data in each of the tables in the dataset

Lesson 3: Querying Data in BigQuery

- Use basic select statements
- Join tables in BigQuery
- Use aggregate functions such as MIN, MAX, and SUM
- Group rows
- Filter results using WHERE and HAVING

Lesson 3: Querying Data in BigQuery

- Write a query to return the station ID and name of stations that have more than 70 bikes available.

- Modify the above query to join the stations table to the trips table. Add trip duration and bike ID columns to the query.

- Finally, modify the query to return the minimum and maximum trip durations.

Lesson 4: Using Math, Statistics, and Data Conversion Functions in BigQuery

- Use basic math functions
- Use statistical functions
- Use CASTing to convert between data types

Lesson 4: Exercise

- Write a query to divide trip duration by 60 and round to 2 decimal places
- Write a query to find the standard deviation of trip duration
- Write a query to convert trip duration to a string and append “ min” as a suffix.

Lesson 5: Working with Strings and Regular Expressions

- Apply trimming and padding to strings
- Manipulate strings to create substrings
- Split strings on delimiters
- Matching strings using regular expressions

Lesson 5: Exercise

- Left pad the station ID to 10 characters using a space as a padding character.
- Write a query to return the first 5 characters of the station name
- Write a query to split station name using a space as the delimiter

Lesson 6: Working with Dates and Times

- Use a variety of date functions
- Use time and timestamp functions

Lesson 6: Exercise

- Write a query to extract the day of week and day of year from the start time column in the citibikes_trip table.

- Write a query that returns the start time of a trip and the datetime of one week after the start of the trip.

- Write a query that returns the start time of a trip and the start time truncated to the day interval.

Lesson 7: Working with Arrays and Structs

- Query arrays
- Unnest arrays
- Query structs
- Query array of structs

Lesson 7: Exercises

- Link the Google Analytics Sample dataset from the Analytics Hub

- Write a query to return the visitor ID, the start time, and the total number of page views.

- Write a query to return the visitor ID, the start time, and the product name of items in the hits structure.

Conclusions and Next Steps