

Google Cloud Platform

Interfacing with External Tools

BigQuery for Data Analysts

V1.2

Approximate timing: 20 minutes

Agenda

- 1 Interfacing with Spreadsheets
- 2 ODBC & JDBC Drivers
- 3 Encrypted BigQuery Client
- 4 BigQuery and R
- 5 Lab

BigQuery Connector for Microsoft Excel (1 of 2)

- Supports Excel 2007 and up
- Supports Windows and Mac
- Access through authorization key
 - Time sensitive
 - Min 1 hour – Max 30 days
 - Key can be revoked

Notes:

Although spreadsheet are not designed to handle big data, many business run on them and use them daily. Spreadsheets are understandable by both technical and non-technical staff. Spreadsheets allow for use of simple charts and graphs to be easily built.

You could also connect to BigQuery from Excel via an ODBC connector.

BigQuery Connector for Microsoft Excel (2 of 2)

- Go to <https://bigquery-connector.appspot.com>
- Select Google account to use
- Record unique key and download IQY file
- Follow site instructions to execute query from Excel

Using Google Sheets with BigQuery (1 of 2)

- Extend Google Sheets using App Script
- Rich interface
- JavaScript-based language
 - Create buttons, pulldowns, and so on
 - Create dynamic query parameters
 - Create visualizations

Using Google Sheets with BigQuery (2 of 2)

- OWOX [BigQuery Reports Add-On](#)
 - Save queries with preset variables
 - Create visualizations
 - Share results
- Alternative to writing scripts
- [Free version available](#)

Agenda

- 1 Interfacing with Spreadsheets
- 2 ODBC & JDBC Drivers
- 3 Encrypted BigQuery Client
- 4 BigQuery and R
- 5 Lab

Simba ODBC/JDBC Drivers ^{Beta}

- [Simba ODBC/JDBC Drivers](#)
- 32-bit and 64-bit
- Available for Mac, Linux, Windows
- Supports ANSI SQL-92: SELECT, JOIN, WHERE, HAVING, GROUP BY, ORDER BY, TOP and most SQL-92 scalar and aggregate functions
- Supports BigQuery's SQL subset: SELECT, HAVING, WHERE, GROUP BY, ORDER BY, LIMIT, CASE and all functions
- Supports all BigQuery data types (STRING, INTEGER, FLOAT, BOOLEAN, TIMESTAMP)

Notes:

Google has partnered with Simba Technologies to provide updated ODBC and JDBC drivers that leverage the power of BigQuery's Standard SQL (support is also provided for legacy SQL). For more information on the Simba ODBC/JDBC drivers for BigQuery, see:

<https://cloud.google.com/bigquery/partners/simba-beta-drivers>.

As of this writing, the drivers had the following limitations:

- The drivers do not yet support parameterized queries.
- The drivers do not yet support DML statements (such as INSERT INTO).
- The drivers will yield errors when executing queries that have complex type representations such as nested and/or repeated values.
- There is limited support for ODBC escape sequences.

Other JDBC Drivers

- [Starschema JDBC driver for BigQuery](#)
 - Supports server and OAuth2 authentication
 - Supports handling metadata
 - Query transformation capabilities
 - Released to open source - No longer under active development
- [CData JDBC driver for BigQuery](#)
 - Abstracts BigQuery data source into tables, views, stored procedures use to access data

Notes:

Although the Starschema driver is available, not active work has been done since June 2013. This means that any enhancements to BigQuery may not be reflected in the driver.

Agenda

- 1 Interfacing with Spreadsheets
- 2 ODBC & JDBC Drivers
- 3 Encrypted BigQuery Client
- 4 BigQuery and R
- 5 Lab

Encrypted Client (1 of 2)

- An experimental extension to the BigQuery client
- Offers client-side encryption for a subset of query types
- Implemented in Python
- Encrypts data before loading and transforms query to work on top of encrypted data
- Only available as a replacement for bq CLI

Encrypted Client (2 of 2)

- Supports multiple encryption modes
 - *Pseudonym* - encrypts the data the same way, given a particular key
 - *Probabilistic* - encrypts the same text differently every time
 - *Homomorphic* - Encrypts numeric fields with special mathematical properties allowing mathematical operations on encrypted data to yield encrypted results
 - *Searchwords* - Encrypts data so you can find a particular word within a longer string – same word is encrypted the same way every time
 - *Probabilistic_searchwords* - combines the two types of encryption so that a word is encrypted a different way every time
 - *None* - No encryption

Client Interaction

- Normal client interaction
 - Data on client in normal (unencrypted) state
 - Data moves between client and BigQuery over SSH
 - Data is encrypted in flight and at rest once in BigQuery
- Encrypted client interaction
 - Interface encrypts input data on client
 - Encrypted data moves between client and BigQuery over SSH
 - Data encrypted in flight and at rest once in BigQuery
 - Query results to client are encrypted
 - Client interface decrypts the results

Agenda

- 1 Interfacing with Spreadsheets
- 2 ODBC & JDBC Drivers
- 3 Encrypted BigQuery Client
- 4 BigQuery and R
- 5 Lab

BigQuery and R (1 of 2)

- Environment for statistical computing
- Contains large, integrated collection of data analysis tools
- Graphical facilities
- Simple and effective programming language

Notes:

Reference – www.r-project.org/about.html

BigQuery and R (2 of 2)

- BigQuery added as extension package
- BigQuery allows R to process very large datasets
- Hundreds of modeling packages available
- R provides very sophisticated analysis
- Easy setup and use

Notes:

Available extension packages and documentation at

http://cran.us.r-project.org/web/packages/available_packages_by_name.html.

One limitation of R requires that all the data it operates on reside in memory. If you analyze a billion row dataset, it is unlikely that you will have enough memory. The BigQuery extension package allows BigQuery do a phase of that work in the service sending the results to R for further modeling or other sophisticated analysis.

R extends BigQuery with hundreds of additional functions that BigQuery does not have.

Easy setup: Add the BigQuery package, grant a one-time authorization to BigQuery, then execute queries.

Agenda

- 1 Interfacing with Spreadsheets
- 2 ODBC & JDBC Drivers
- 3 Encrypted BigQuery Client
- 4 BigQuery and R
- 5 Lab

Lab

Use the BigQuery Reports add-on for Google Sheets to query data

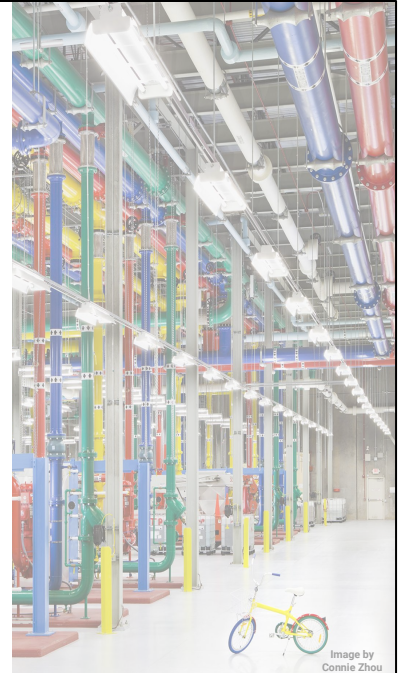


Image by
Connie Zhou

Resources

- Third party tools
<https://cloud.google.com/bigquery/third-party-tools>
- Client-side Encryption for BigQuery
<https://code.google.com/p/encrypted-bigquery-client/>
- R
<https://www.r-project.org/about.html>

