



# How to use Google Cloud in a datathon

[tiny.cc/koreadatathon](http://tiny.cc/koreadatathon)

Google Cloud



All teams will be saving their work in this project

# Google Cloud project structure

Data-hosting project

**physionet-data, snuop-221609, chrome-coast-249308**

Google Cloud Storage

Data bucket 1

Data bucket 2

Data bucket 3

Google Cloud BigQuery analytical warehouse

Dataset 1

Dataset 2

Dataset 3

Analysis project

**korea-datathon-2019**

SQL through BigQuery

R through RStudio

python in Jupyter notebooks through DataLab



# Intro to BigQuery

Google Cloud

# What is BigQuery?

Google Cloud Platform's enterprise data warehouse for analytics

**Save**  
Fully-managed service with low TCO

**Secure**  
Data encrypted at rest and in motion with fine-grained access control



## Simple

- Analyze data from across Google Cloud Platform using SQL

## Scalable

- Multi-petabyte horizontal scalability, high performance

## Shareable

- Public and commercial datasets to enhance analysis

# New BigQuery UI

# First time in BigQuery

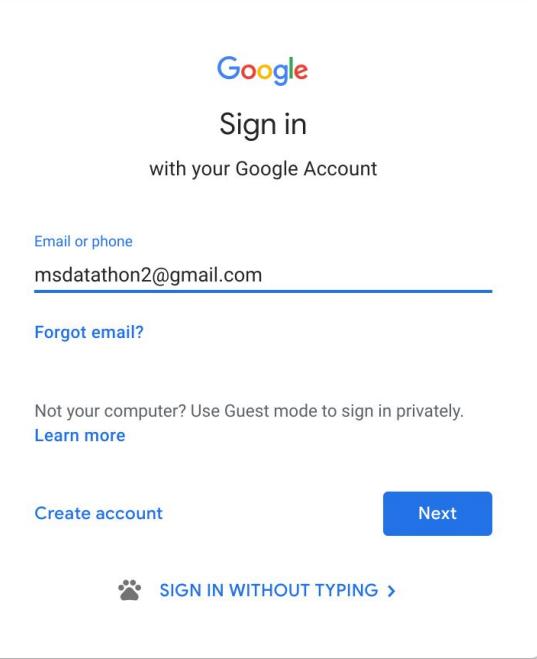
Enter

<http://console.cloud.google.com/bigquery?project=korea-datathon-2019> & log in with your Google ID  
(Gmail or G Suite university domain)



Sign in - Google Accounts

https://accounts.google.com/signin/v2/identifier?service=bigquery&passive=1209600&conti...



The image shows a screenshot of a web browser displaying the Google Sign-in page. The URL in the address bar is https://accounts.google.com/signin/v2/identifier?service=bigquery&passive=1209600&conti... The page features the Google logo and the text "Sign in with your Google Account". A text input field is populated with the email address "msdatathon2@gmail.com". Below the input field are links for "Forgot email?" and "Not your computer? Use Guest mode to sign in privately. Learn more". At the bottom, there are "Create account" and "Next" buttons, and a "SIGN IN WITHOUT TYPING >" button with a paw print icon.

English (United States) ▾

Help Privacy Terms

# First time in BigQuery

Log in with your Google ID  
(Gmail or G Suite university domain)



Sign in - Google Accounts

https://accounts.google.com/signin/v2/sl/pwd?service=bigquery&passive=1209600&continu...

Google Welcome

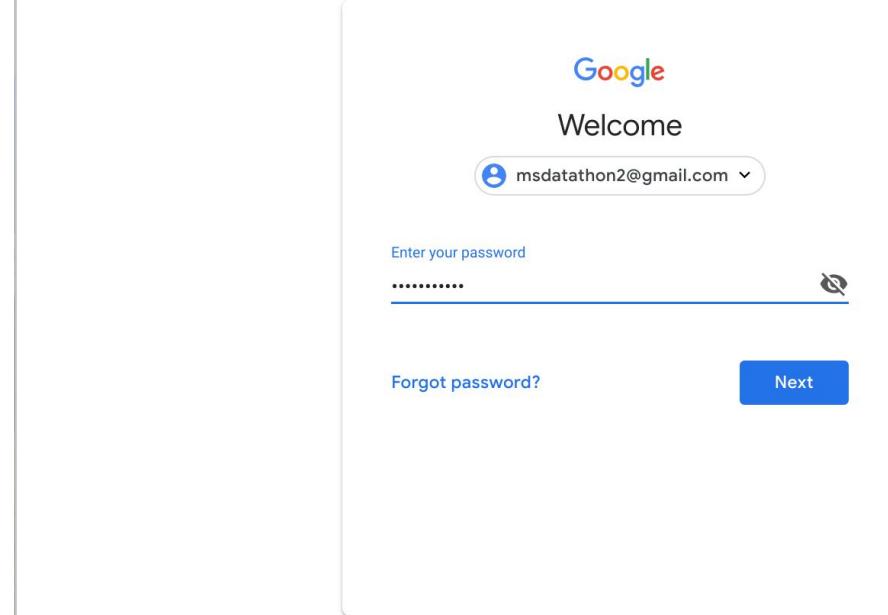
msdatathon2@gmail.com

Enter your password

.....

Forgot password? Next

English (United States) ▾ Help Privacy Terms

A screenshot of a Google sign-in page for BigQuery. The page features the Google logo and the word "Welcome". A user's email address, "msdatathon2@gmail.com", is displayed in a dropdown menu. Below the email is a password input field with placeholder text "Enter your password" and a redacted password. To the right of the input field is a "Next" button. At the bottom of the page are links for "Forgot password?", "Help", "Privacy", and "Terms". The status bar at the top shows the URL "https://accounts.google.com/signin/v2/sl/pwd?service=bigquery&passive=1209600&continu...".

# First time in BigQuery

Accept the conditions



Enable an API - Google API Console +

>Your free trial is waiting: activate now to get \$300 credit to explore Google Cloud products. [Learn more](#)

[DISMISS](#) [ACTIVATE](#)

Google APIs Select a project  [Search](#)

Register your application for BigQuery API in Google API Console

Google API Console allows you to manage your application and monitor API usage.

You have no existing projects. A new project named "My Project" will be created.

Terms of Service  I agree to the [Google Cloud Platform Terms of Service](#), and the terms of service of [any applicable services and APIs](#).

Country of residence

Please email me updates regarding feature announcements, performance suggestions, feedback surveys and special offers.

Yes  No

[Agree and continue](#)

https://console.developers.google.com/flows/enableapi?apiid=bigquery&pli=1

# First time in BigQuery

Accept the conditions



Enable an API - Google API Cons X +

>Your free trial is waiting: activate now to get \$300 credit to explore Google Cloud products. [Learn more](#)

DISMISS ACTIVATE

Google APIs Select a project

Register your application for BigQuery API in Google API Console

Google API Console allows you to manage your application and monitor API usage.

You have no existing projects. A new project named "My Project" will be created.

**Terms of Service**  
 I agree to the [Google Cloud Platform Terms of Service](#), and the terms of service of [any applicable services and APIs](#).

Country of residence

United States

Please email me updates regarding feature announcements, performance suggestions, feedback surveys and special offers.

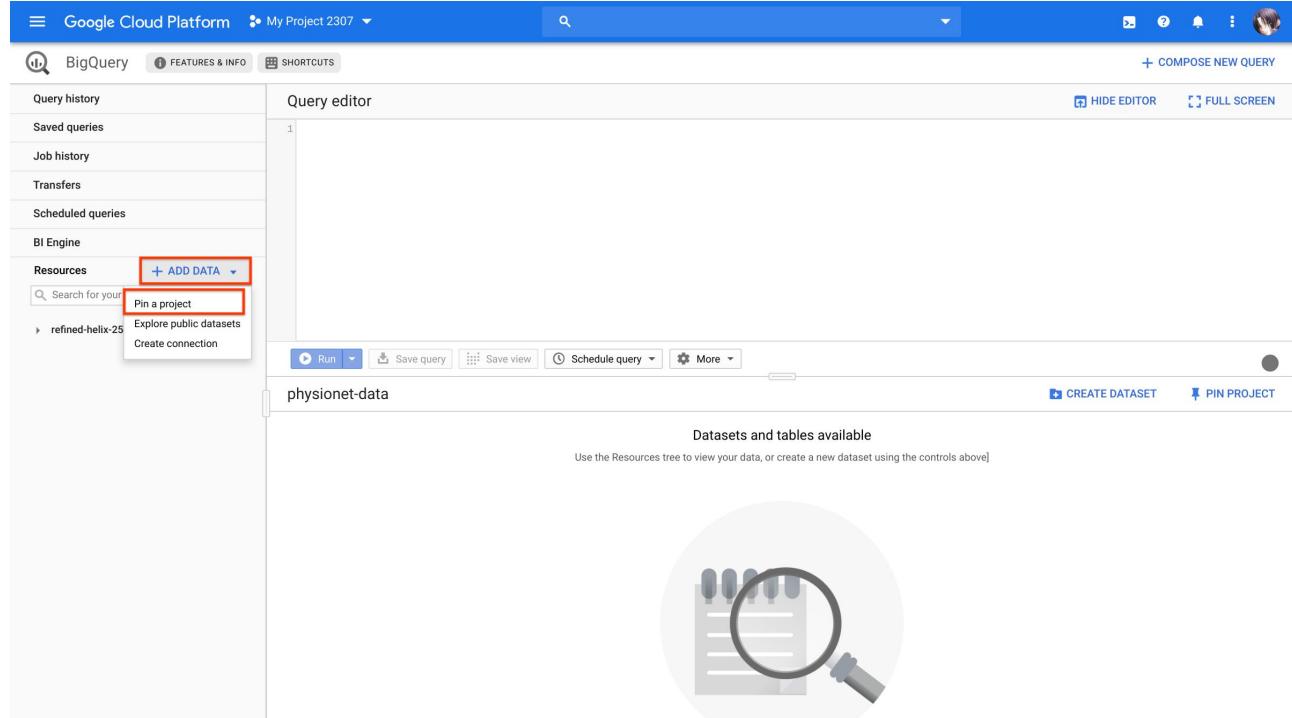
Yes  No

Agree and continue

The screenshot shows the "Enable an API" page in the Google API Console. It displays a registration form for a new project. Two specific sections are highlighted with red boxes: the "Terms of Service" checkbox and the "Please email me updates" section. Both of these highlighted sections contain text and input fields that are typically used for accepting terms and conditions or opting-in to marketing communications. The rest of the page contains standard form fields like "Country of residence" and a "Agree and continue" button.

# Working in BigQuery

Find the dataset you want to query  
We'll use the Admission table from  
**mimiciii-demo** dataset in this  
walkthrough



The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there's a sidebar titled "Resources" with options like "Query history", "Saved queries", "Job history", "Transfers", "Scheduled queries", and "BI Engine". Below this is a search bar and a dropdown menu with options: "+ ADD DATA", "Pin a project" (which is highlighted with a red box), "Explore public datasets", and "Create connection". The main area is titled "Query editor" and shows a single row of code "1". At the bottom of the editor are buttons for "Run", "Save query", "Save view", "Schedule query", and "More". To the right of the editor, the text "physionet-data" is displayed above a section titled "Datasets and tables available". A large circular icon with a magnifying glass over a document is centered at the bottom.

# Working in BigQuery

Fill in the project name (I used **physionet-data** in this case)

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with options like Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, and Resources. Below the sidebar, there is a search bar and a dataset named "refined-helix-253408". In the main area, there is a "Query editor" tab with a single row labeled "1". A modal dialog box titled "Pin a project" is open in the center. It contains instructions: "Select a project from the list below to add it to the Resources tree. It will be pinned for easy access going forward." There are two options: "Enter a project name" (radio button selected) and "Search for a project" (radio button unselected). A dropdown menu under "Search for a project" is labeled "Select a project". At the bottom of the dialog are "CANCEL" and "PIN" buttons. A watermark of a magnifying glass over a grid is visible in the background.

# Working in BigQuery

The project shows up on the left side.

The screenshot shows the Google Cloud Platform BigQuery interface. At the top, there's a navigation bar with the Google Cloud logo, 'Google Cloud Platform', 'My Project 2307', a search bar, and various icons. Below the navigation bar is a sidebar titled 'BigQuery' with links for 'Query history', 'Saved queries', 'Job history', 'Transfers', 'Scheduled queries', 'BI Engine', and 'Resources'. A 'Search for your tables and datasets' input field is also in the sidebar. The main area is titled 'Query editor' and contains a code editor with the number '1'. Below the code editor are buttons for 'Run', 'Save query', 'Save view', 'Schedule query', and 'More'. A dropdown menu shows 'refined-helix-253408' and 'physionet-data', with 'physionet-data' currently selected and highlighted with a red box. To the right of the code editor, there are buttons for 'CREATE DATASET' and 'UNPIN PROJECT'. A large circular icon with a magnifying glass over a document is displayed at the bottom right. The overall interface is clean and modern, designed for efficient data querying and management.

# Working in BigQuery

Drill on the **mimiciii\_demo** dataset to see the tables. Click on the name of each table to see the schema.

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with links for Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, and Resources. Below this is a search bar and a tree view of datasets. Under the 'mimiciii\_demo' dataset, several tables are listed: admissions, callout, caregivers, chartevents, cptevents, d\_cpt, d\_icd\_diagnoses, d\_icd\_procedures, d\_items, d\_labitems, datetimeevents, and diagnoses\_icd. The 'admissions' table is selected, and its schema is displayed in the main pane. The schema table has columns for Field name, Type, Mode, and Description. The 'admissions' table schema is as follows:

Field name	Type	Mode	Description
ROW_ID	INTEGER	NULLABLE	
SUBJECT_ID	INTEGER	NULLABLE	
HADM_ID	INTEGER	NULLABLE	
ADMITTIME	DATETIME	NULLABLE	
DISCHTIME	DATETIME	NULLABLE	
DEATHTIME	DATETIME	NULLABLE	
ADMISSION_TYPE	STRING	NULLABLE	
ADMISSION_LOCATION	STRING	NULLABLE	
DISCHARGE_LOCATION	STRING	NULLABLE	

# Working in BigQuery

Click **Preview** to see a sample of records.

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with links to Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, and Resources. Below this is a search bar for tables and datasets. Under the resources section, the 'mimiciii\_demo' dataset is selected, and its schema is visible. In the main area, an 'Unsaved query' is being edited with the following SQL code:

```
1 SELECT * FROM `physionet-data.mimiciii_demo.admissions` LIMIT 1000
```

Below the code, there are buttons for Run, Save query, Save view, Schedule query, and More. A note indicates that the query will process 23.3 KB when run. The results section shows the 'admissions' table with columns: Row, ROW\_ID, SUBJECT\_ID, HADM\_ID, ADMITTIME, DISCHTIME, DEATHTIME, ADMISSION\_TYPE, ADMISSION\_LOCATION, and DISCHARGE\_LOCATION. The first 8 rows of data are displayed:

Row	ROW_ID	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION
1	12269	10017	199207	2149-05-26T17:19:00	2149-06-03T18:42:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF
2	12280	10029	132349	2139-09-22T10:58:00	2139-10-02T14:29:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF
3	12283	10033	157235	2132-12-05T02:46:00	2132-12-08T15:15:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF
4	12288	10038	111115	2144-02-09T17:53:00	2144-02-21T13:30:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF
5	12293	10043	168674	2185-04-14T00:23:00	2185-04-26T18:20:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF
6	12337	10088	169938	2107-01-04T11:59:00	2107-01-11T15:45:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF
7	12338	10088	168233	2107-01-29T04:00:00	2107-02-10T12:00:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF
8	12345	10094	168074	2180-02-29T18:54:00	2180-03-10T17:35:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF

At the bottom, there are buttons for Rows per page (set to 100), First page, Last page, and Next/Previous page controls.

# Working in BigQuery

Now let's query the table by clicking the **QUERY TABLE** button.

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with links to Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, and Resources. Below this is a search bar for tables and datasets. Under the resources section, there are two datasets: 'eicu\_crd\_demo' and 'mimiciii\_demo'. The 'mimiciii\_demo' dataset is expanded, showing tables like 'admissions', 'callout', 'caregivers', etc. In the main area, an 'Unsaved query' is being edited with the following SQL code:

```
1 SELECT * FROM `physionet-data.mimiciii_demo.admissions` LIMIT 1000
```

Below the code are buttons for Run, Save query, Save view, Schedule query, and More. To the right, it says "This query will process 23.3 KB when run." and has buttons for QUERY TABLE (which is highlighted with a red box), COPY TABLE, DELETE TABLE, and EXPORT. The 'admissions' table is selected. A preview of the table schema is shown with columns: Row, ROW\_ID, SUBJECT\_ID, HADM\_ID, ADMITTIME, DISCHTIME, DEATHTIME, ADMISSION\_TYPE, ADMISSION\_LOCATION, and DISCHARGE\_LOCATION. The preview shows 8 rows of sample data. At the bottom, there are buttons for Rows per page (set to 100), navigation (First page, Previous page, Next page, Last page), and a status message: "1 - 100 of 129".

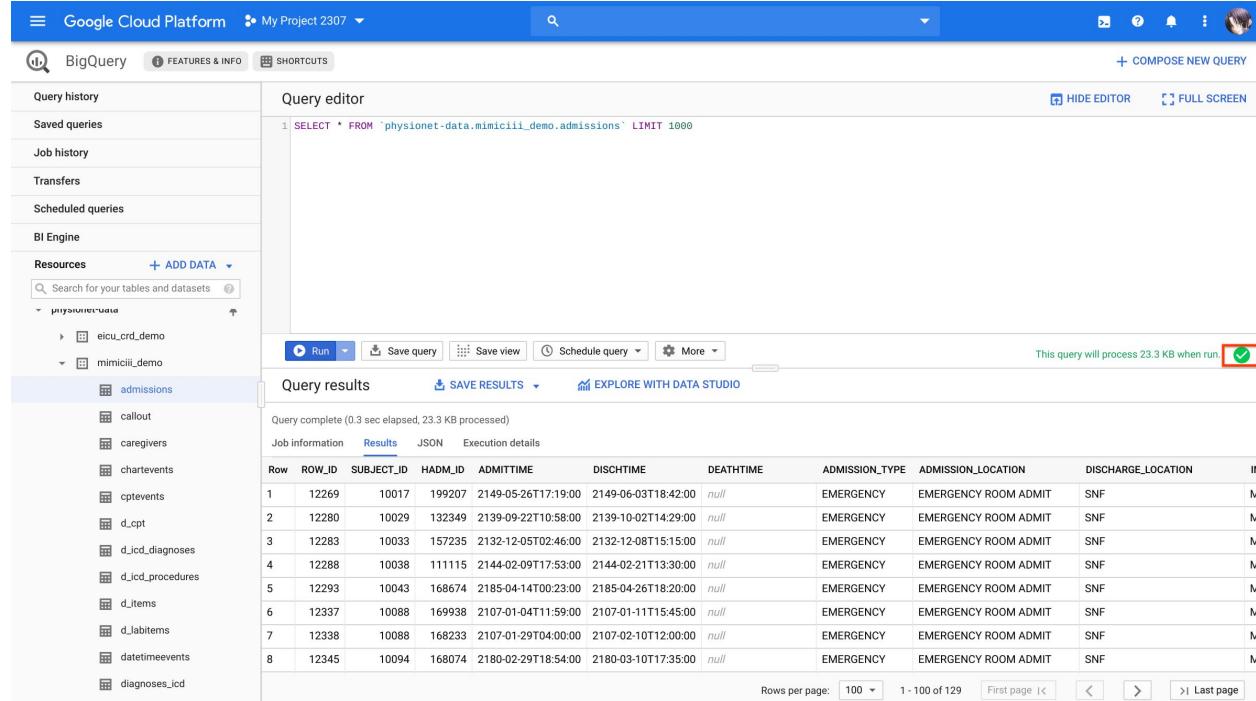


# Working in BigQuery

To query the table, write SQL.

The format for the table name is  
`'project.dataset.table'`

(those are back-ticks)



The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there's a sidebar with links for Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, and Resources. Below this is a search bar for tables and datasets. Under the search bar, the dataset 'mimiciii\_demo' is selected, and its tables are listed: admissions, callout, caregivers, chartevents, cptevents, d\_cpt, d\_icd\_diagnoses, d\_icd\_procedures, d\_items, d\_labitems, datetimeevents, and diagnoses\_icd. The 'admissions' table is currently selected. The main area is the 'Query editor' where the following SQL query is written:

```
1 SELECT * FROM `physionet-data.mimiciii_demo.admissions` LIMIT 1000
```

Below the editor is the 'Query results' section. It shows the first 8 rows of the 'admissions' table:

Row	ROW_ID	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION	INSURANCE
1	12269	10017	199207	2149-05-26T17:19:00	2149-06-03T18:42:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Medicare
2	12280	10029	132349	2139-09-22T10:58:00	2139-10-02T14:29:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Medicare
3	12283	10031	157235	2132-12-05T02:46:00	2132-12-08T15:15:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Medicare
4	12288	10038	111115	2144-02-09T17:53:00	2144-02-21T13:30:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Medicare
5	12293	10043	168674	2185-04-14T00:23:00	2185-04-26T18:20:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Medicare
6	12337	10088	169938	2107-01-04T11:59:00	2107-01-11T15:45:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Medicare
7	12338	10088	168233	2107-01-29T04:00:00	2107-02-10T12:00:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Medicare
8	12345	10094	168074	2180-02-29T18:54:00	2180-03-10T17:35:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Medicare

At the bottom of the results table, there are pagination controls: 'Rows per page: 100', '1 - 100 of 129', 'First page', and 'Last page'.

# Working in BigQuery

Type your SQL into the **Query Editor**.

Check for confirmation that your query will execute.

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with links for 'Query history', 'Saved queries', 'Job history', 'Transfers', 'Scheduled queries', 'BI Engine', and 'Resources'. Below this is a search bar for tables and datasets. Under 'Resources', the 'physionetdata' dataset is selected, showing tables like 'admissions', 'callout', 'caregivers', etc. The main area is the 'Query editor' where the following SQL query is typed:

```
1 SELECT * FROM `physionet-data.mimiciii_demo.admissions` LIMIT 1000
```

Below the editor are buttons for 'Run', 'Save query', 'Save view', 'Schedule query', and 'More'. A note says 'This query will process 23.3 KB when run' with a green checkmark. The 'Query results' section shows the first 8 rows of the query output:



# BigQuery SQL syntax

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

<https://cloud.google.com/bigquery/docs/reference/standard-sql>

It's normal SQL but extremely fast  
with support for user-defined  
functions and embedded hierarchies



## Standard SQL Query Syntax

Query statements scan one or more tables or expressions and return the computed result rows.  
This topic describes the syntax for SQL queries in BigQuery.

### SQL Syntax

```
query_statement:
  [ WITH with_query_name AS ( query_expr ) [, ...] ]
  query_expr

query_expr:
  { select | ( query_expr ) | query_expr set_op query_expr }
  [ ORDER BY expression { ASC | DESC } [, ...] ]
  [ LIMIT count [ OFFSET skip_rows ] ]

select:
  SELECT [ { ALL | DISTINCT } ]
    { expression.* [ EXCEPT ( column_name [, ...] ) ] }
    [ REPLACE ( expression [ AS ] column_name [, ...] ) ]
    | expression [ [ AS ] alias ] [, ...]
  [ FROM from_item [, ...] ]
  [ WHERE bool_expression ]
  [ GROUP BY { expression [, ...] | ROLLUP ( expression [, ...] ) } ]
  [ HAVING bool_expression ]
  [ WINDOW window_name AS ( window_definition ) [, ...] ]

set_op:
  UNION { ALL | DISTINCT } | INTERSECT DISTINCT | EXCEPT DISTINCT

from_item:
  table_name [ [ AS ] alias ] [ FOR SYSTEM TIME AS OF timestamp_expression ] |
  join |
  ( query_expr ) [ [ AS ] alias ] |
  field_path |
  { UNNEST( array_expression ) | UNNEST( array_path ) | array_path }
    [ [ AS ] alias ] [ WITH OFFSET [ [ AS ] alias ] ] |
  with_query_name [ [ AS ] alias ]
}

join:
  from_item [ join_type ] JOIN from_item
  [ ON bool_expression | USING ( join_column [, ...] ) ] |

join_type:
  { INNER | CROSS | FULL [OUTER] | LEFT [OUTER] | RIGHT [OUTER] }
```



Internal: Count: 10, Average: 3.8

Contents	
SQL Syntax	
SELECT list	
SELECT *	
SELECT expression	
SELECT expression.*	
SELECT modifiers	
Aliases	
Analytic functions	
FROM clause	
Syntax	
Subqueries	
Aliases	
JOIN types	
Syntax	
[INNER] JOIN	
CROSS JOIN	
FULL [OUTER] JOIN	
LEFT [OUTER] JOIN	
RIGHT [OUTER] JOIN	
ON clause	
USING clause	
Sequences of JOINs	
WHERE clause	
Syntax	
GROUP BY clause	
Syntax	
HAVING clause	
Syntax	
Mandatory aggregation	
ORDER BY clause	
Syntax	
Set operators	
Syntax	
UNION	
INTERSECT	
EXCEPT	
LIMIT clause and OFFSET clause	
Syntax	
WITH clause	
Aliases	
Explicit alias syntax	
Explicit alias visibility	

# Working in BigQuery

Create a dataset first in order to materialize a view.

Click **CREATE DATASET** button.

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with various options: Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, and Resources. Under Resources, a dataset named "refined-helix-253408" is expanded, showing sub-datasets like "physionet-data" and "mimiciii\_demo". The "mimiciii\_demo" dataset contains tables such as "admissions", "callout", "caregivers", "chartevents", "cptevents", "d\_cpt", "d\_icd\_diagnoses", "d\_icd\_procedures", "d\_items", "d\_labitems", and "d\_labevents". The main area is the "Query editor" where a simple SELECT query is written:

```
1 SELECT * FROM `physionet-data.mimiciii_demo.admissions` LIMIT 1000
```

Below the editor are buttons for Run, Save query, Save view, Schedule query, and More. A message indicates the query will process 23.3 KB. At the bottom right of the editor area, there are two buttons: "CREATE DATASET" (highlighted with a red box) and "PIN PROJECT". A large circular icon with a magnifying glass over a document is positioned at the bottom center.

# Working in BigQuery

Fill in the dataset ID.

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with navigation links: Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, Resources, and + ADD DATA. Below the sidebar, the project structure is displayed under 'refined-helix-253408': 'physionet-data' contains 'eicu\_crd\_demo' and 'mimiciii\_demo', which further contains tables like 'admissions', 'callout', 'caregivers', etc. In the center, the 'Query editor' pane displays a single query: `1 SELECT * FROM `physionet-data.mimiciii_demo.admissions` LIMIT 1000`. Below the query editor are buttons for Run, Save query, Save view, Schedule query, and More. To the right of the query editor, a 'Create dataset' dialog is open. It has fields for 'Dataset ID' (set to 'temp'), 'Data location (Optional)' (set to 'Default'), 'Default table expiration' (radio button selected for 'Never'), and 'Encryption' options (radio button selected for 'Google-managed key'). At the bottom of the dialog are 'Create dataset' and 'Cancel' buttons.



# Working in BigQuery

To materialize a view.

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with navigation links: Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, and Resources. Below this is a search bar for tables and datasets. Under the resources section, the 'physionet-data' dataset is selected, showing its contents: admissions, callout, caregivers, chartevents, cptevents, d\_cpt, d\_icd\_diagnoses, d\_icd\_procedures, d\_items, d\_labitems, datetimeevents, and diagnoses\_icd.

In the main area, the 'Query editor' tab is active. A query is written in SQL:

```
1 SELECT * FROM `physionet-data.mimiciii_demo.admissions` LIMIT 1000
```

Below the editor are buttons for Run, Save query, Save view (which is highlighted with a red box), Schedule query, and More. A note says "This query will process 23.3 KB when run." To the right of the editor is the 'Query results' section. It shows the results of the query, which are the first 8 rows of the 'admissions' table from the 'mimiciii\_demo' dataset. The columns listed are Row, ROW\_ID, SUBJECT\_ID, HADM\_ID, ADMITTIME, DISCHTIME, DEATHTIME, ADMISSION\_TYPE, ADMISSION\_LOCATION, DISCHARGE\_LOCATION, and INS. The results show various patient admissions with details like admission time, discharge time, and location.

Row	ROW_ID	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION	INS
1	12269	10017	199207	2149-05-26T17:19:00	2149-06-03T18:42:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
2	12280	10029	132349	2139-09-22T10:58:00	2139-10-02T14:29:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
3	12283	10031	157235	2132-12-05T02:46:00	2132-12-08T15:15:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
4	12288	10038	111115	2144-02-09T17:53:00	2144-02-21T13:30:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
5	12293	10043	168674	2185-04-14T00:23:00	2185-04-26T18:20:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
6	12337	10088	169938	2107-01-04T11:59:00	2107-01-11T15:45:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
7	12338	10088	168233	2107-01-29T04:00:00	2107-02-10T12:00:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
8	12345	10094	168074	2180-02-29T18:54:00	2180-03-10T17:35:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me

At the bottom of the results table, there are pagination controls: Rows per page: 100, 1-100 of 129, First page, Last page, and Next/Previous buttons.



# Working in BigQuery

Save the view in your project/dataset, and give it a name.

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with navigation links: Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, Resources (+ ADD DATA), and a search bar. The main area shows a query editor with the following code:

```
1 SELECT * FROM `physionet-data.mimiciii_demo.admissions` LIMIT 1000
```

Below the query editor is a "Query results" section showing the output of the query. A "Save view" dialog box is overlaid on the interface. The dialog has the following fields:

- Destination:**
  - Project name: My Project 2307
  - Dataset name: temp
- Table name:** 1000\_admissions

At the bottom right of the dialog are "CANCEL" and "SAVE" buttons. The main interface shows a table with 8 rows of data. The columns include Row, ROW\_ID, SUBJECT\_ID, and various timestamp and location fields. The "Results" tab is selected in the top navigation bar. A message at the bottom of the interface says "This query will process 23.3 KB when run." and includes a green checkmark icon.

Row	ROW_ID	SUBJECT_ID	DISCHARGE_DATE	DISCHARGE_LOCATION	DISCHARGE_TYPE	ADMISSION_DATE	ADMISSION_LOCATION	DISCHARGE_DATE	DISCHARGE_LOCATION	DISCHARGE_TYPE	
1	12269	10017	2013-09-22T10:58:00	2013-10-02T14:29:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	2013-09-22T10:58:00	2013-10-02T14:29:00	SNF	ME
2	12280	10029	132349	2139-09-22T10:58:00	2139-10-02T14:29:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	ME	
3	12283	10033	157235	2132-12-05T02:46:00	2132-12-08T15:15:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	ME	
4	12288	10038	111115	2144-02-09T17:53:00	2144-02-21T13:30:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	ME	
5	12293	10043	168674	2185-04-14T00:23:00	2185-04-26T18:20:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	ME	
6	12337	10088	169938	2107-01-04T11:59:00	2107-01-11T15:45:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	ME	
7	12338	10088	168233	2107-01-29T04:00:00	2107-02-10T12:00:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	ME	
8	12345	10094	168074	2180-02-29T18:54:00	2180-03-10T17:35:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	ME	

At the bottom of the interface, there are buttons for "Rows per page: 100", "First page < > Last page", and navigation icons for the results table.



# Working in BigQuery

You can use the view as a normal table now.

The screenshot shows the Google Cloud Platform BigQuery interface. On the left, there is a sidebar with navigation links: Query history, Saved queries, Job history, Transfers, Scheduled queries, BI Engine, Resources, and a search bar. Below this is a tree view of datasets and tables:

- refined-helix-253408 (selected)
- temp (empty)
- 1000\_admissions (highlighted with a red box)
- physionet-data
  - eicu\_crd\_demo (empty)
  - mimiciii\_demo
    - admissions (highlighted with a red box)
    - callout
    - caregivers
    - chartevents
    - cptevents
    - d\_cpt
    - d\_icd\_diagnoses
    - d\_icd\_procedures
    - items

The main area displays a query editor with the following code:

```
refined-helix-253408:temp.1000_admissions
1 SELECT * FROM `physionet-data.mimiciii_demo.admissions` LIMIT 1000
```

Below the editor are buttons for Run, Save query, Save view, Schedule query, and More. A note says "This query will process 23.3 KB when run." The "Query results" section shows the execution details and a table of results. The table has columns: Row, ROW\_ID, SUBJECT\_ID, HADM\_ID, ADMITTIME, DISCHTIME, DEATHTIME, ADMISSION\_TYPE, ADMISSION\_LOCATION, DISCHARGE\_LOCATION, and INS. The results are as follows:

Row	ROW_ID	SUBJECT_ID	HADM_ID	ADMITTIME	DISCHTIME	DEATHTIME	ADMISSION_TYPE	ADMISSION_LOCATION	DISCHARGE_LOCATION	INS
1	12269	10017	199207	2149-05-26T17:19:00	2149-06-03T18:42:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
2	12280	10029	132349	2139-09-22T10:58:00	2139-10-02T14:29:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
3	12283	10031	157235	2132-12-05T02:46:00	2132-12-08T15:15:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
4	12288	10038	111115	2144-02-09T17:53:00	2144-02-21T13:30:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
5	12293	10043	168674	2185-04-14T00:23:00	2185-04-26T18:20:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
6	12337	10088	169938	2107-01-04T11:59:00	2107-01-11T15:45:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
7	12338	10088	168233	2107-01-29T04:00:00	2107-02-10T12:00:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me
8	12345	10094	168074	2180-02-29T18:54:00	2180-03-10T17:35:00	null	EMERGENCY	EMERGENCY ROOM ADMIT	SNF	Me

At the bottom, there are buttons for Rows per page, First page, Last page, and Next/Previous page.

# Accessing RStudio

## RStudio is hosted on Google Compute Engine

<http://34.85.6.31:8787>

Talk to me for login credentials

The RStudio tutorial is accessed from the right pane



R version 3.3.3 (2017-03-06) -- "Another Canoe"  
Copyright (C) 2017 The R Foundation for Statistical Computing  
Platform: x86\_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

>

Environment is empty

Name	Size	Modified
bigquery_tutorial.Rmd	9.3 KB	Sep 8, 2018, 12:13 AM

# Working in Colab

Free Jupyter notebook environment in the cloud, stored on Google Drive

[https://colab.research.google.com/github/GoogleCloudPlatform/healthcare/blob/master/datathon/mimic\\_eicu/tutorials/bigquery\\_tutorial.ipynb](https://colab.research.google.com/github/GoogleCloudPlatform/healthcare/blob/master/datathon/mimic_eicu/tutorials/bigquery_tutorial.ipynb)

Run the code



Make a copy of the notebook

Share it with others in your team

The screenshot shows a Google Colab notebook interface. At the top, there's a toolbar with a 'File' button highlighted in green, a 'SHARE' button in the top right corner, and other options like 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. Below the toolbar, the title 'Datathon Tutorial' is displayed. A sub-section titled 'Prerequisites' is shown, explaining the need for a Google account and linking to account creation. A 'Setup' section follows, providing instructions on creating a copy of the notebook and sharing it. A code cell at the bottom contains Python code for importing libraries:

```
# Import libraries
import numpy as np
import os
import pandas as pd
import matplotlib.pyplot as plt
```

# Accessing BigQuery with other tools

If the analysis tool you want to use does not have a native BigQuery connector, you may be able to use a JDBC/ODBC connector to connect to BigQuery.

Matlab example:

<https://github.com/GoogleCloudPlatform/healthcare/blob/master/datathon/anzics18/tutorial.matlab>

# Find the tutorials

Hosted on Github, no need for an account

[https://github.com/GoogleCloudPlatform/healthcare/tree/master/datathon/mimic\\_eicus](https://github.com/GoogleCloudPlatform/healthcare/tree/master/datathon/mimic_eicus)

- BigQuery ML
- Colab
- ...

GoogleCloudPlatform / healthcare

Code Issues Pull requests Projects Security Insights

Branch: master · healthcare / datathon / mimic\_eicus

Cloud Healthcare Team and reidhayes Fix bigrquery errors in certain regions.

tutorials Fix bigrquery errors in certain regions. 3 months ago

FAQ.md Remove references to analyst(1-5) example accounts 8 months ago

README.md Remove references to analyst(1-5) example accounts 8 months ago

## Datathon Tutorials

Welcome to datathon on GCP!

We have prepared tutorials to get you started on [BigQuery](#), the tool to filter, join, aggregate and extract data from the raw datasets for analysis. In each of the tutorials, some comprehensive examples are included to show how to view the datasets, run transformations and analyze them.

- For Python users, please start from the [Python colab](#) (a copy is available in the `tutorials` folder as well), which is a Jupyter notebook hosted in Google Drive, and can be shared with other people for collaboration. It has the most comprehensive examples, including how to train machine learning models on the MIMIC demo dataset with [Tensorflow](#).
- For people who have experience with R, start with the [R tutorial](#), which provides an interactive interface to go through the tutorial in RStudio.

