Streaming Data Processing - Lab 3 : Streaming Analytics and Dashboards v1.3

## Overview

Data visualization tools can help you make sense of your BigQuery data and help you analyze the data interactively. You can use visualization tools to help you identify trends, respond to them, and make predictions using your data. In this lab, you use Google Data Studio to visualize data in the BigQuery table populated by your Dataflow pipeline in the previous exercise.

At the time of this writing, streaming pipelines are not available in the DataFlow Python SDK. So the streaming labs are written in Java.

## Objectives

In this lab, you will perform the following tasks:

* Connect to a BigQuery data source
* Create reports and charts to visualize BigQuery data

This lab uses Google Data Studio to visualize data in BigQuery using the BigQuery connector. In subsequent tasks, you will create a data source, a report, and charts that visualize data in the sample table.

## Task 1: Preparation

You will be running a sensor simulator from the training VM. There are several files and some setup of the environment required.

### **Open the SSH terminal and connect to the training VM**

1. In the Console, on the **Navigation menu** ( 7a91d354499ac9f1.png), click **Compute Engine** > **VM instances**.
2. Locate the line with the instance called **training-vm**.
3. On the right, under 'Connect' column, click on **SSH** to open a terminal window.
4. In this lab, you will enter CLI commands on the **training-vm**.

### **Verify initialization is complete**

1. The **training-vm** is installing software in the background. Verify that setup is complete by checking that the following directory exists. If it does not exist, wait a few minutes and try again.

ls /training

Wait until setup is complete before proceeding. You can verify the installation of maven with **mvn -version** and the JDK with **java -version**.

### **Copy files**

1. A repository has been downloaded to the VM. Copy the repository to your home directory.

cp -r /training/training-data-analyst/ .

### **Set environment variables**

1. On the **training-vm** SSH terminal enter the following:

source /training/project\_env.sh

This script sets the \_\_$$DEVSHELL\\_PROJECT\\_ID\_\_ and \_\_$$BUCKET\_\_ environment variables.

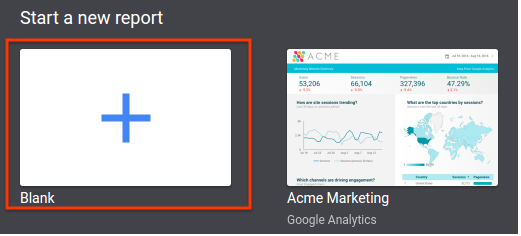
## Task 2: Creating a data source

1. Google Data Studio is a separate service. Open a new browser tab. Navigate to: datastudio.google.com or click on this link: [Google Data Studio](https://datastudio.google.com/)

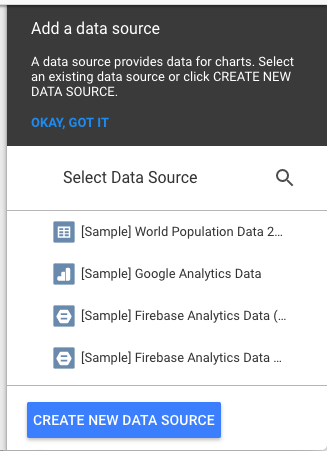
The first step in creating a report in Data Studio is to create a data source for the report. A report may contain one or more data sources. When you create a BigQuery data source, Data Studio uses the BigQuery connector.

You must have the appropriate permissions in order to add a BigQuery data source to a Data Studio report. In addition, the permissions applied to BigQuery datasets will apply to the reports, charts, and dashboards you create in Data Studio. When a Data Studio report is shared, the report components are visible only to users who have appropriate permissions.

1. On the **Reports** page, in the **Start a new report** section, click the **Blank** template. This starts the account setup process.



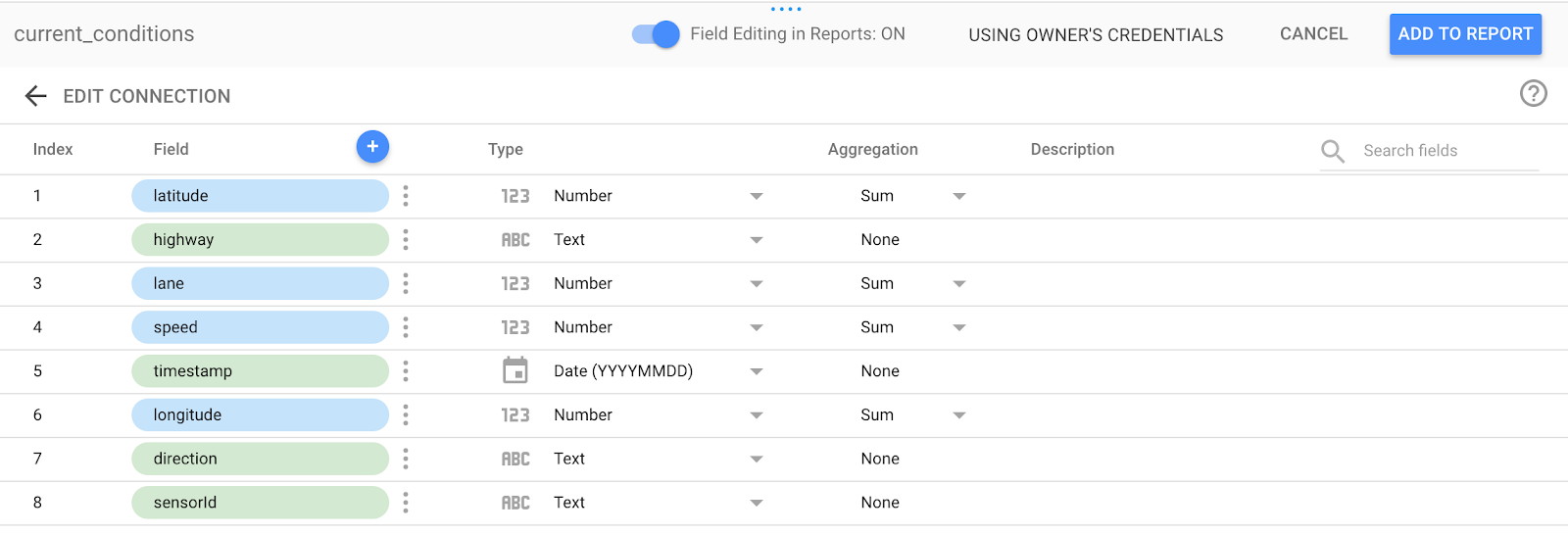
1. On the **Welcome** page, click on **GET STARTED**.
2. On the **Terms** page, click on the **checkbox** to acknowledge the terms. And click **ACCEPT**.
3. On the **Preferences** page, select **No, thanks** for each option to receive email notifications, and click **DONE**.
4. Now that the account is initialized, you need to start the process again.
5. On the **Reports** page, in the **Start a new report** section, click the **Blank** template. This time it will take you to a new page and begin an **Untitled Report**.
6. In the **Add a data source** panel on the right side, click **CREATE NEW DATA SOURCE**.



1. In the **Google Connectors** column on the left, select **BigQuery**.
2. Click on **Authorize**.
3. In the **Sign in** dialog, select your Qwiklabs student account.
4. Click **ALLOW**. to give Google Data Studio permission to view the BigQuery resources in your lab account.
5. Select **My Projects**.
6. In the **Project** column, click on your project name.
7. In the **Dataset** column, click on **demos**.
8. In the **Table** column, click **current\_conditions**.
9. In the upper right corner of the window, click **CONNECT**.

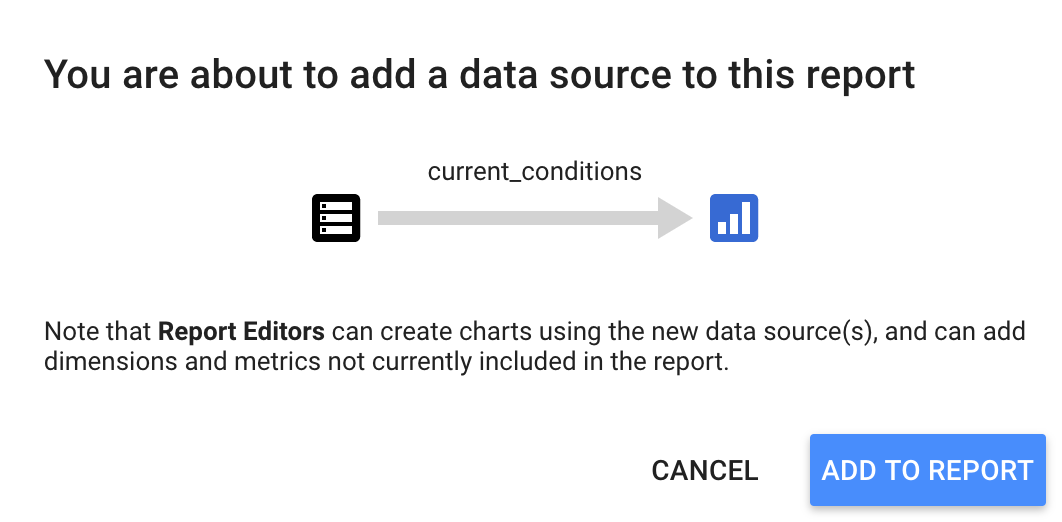
Once Data Studio has connected to the BigQuery data source, the table's fields are displayed. You can use this page to adjust the field properties or to create new calculated fields.

Example:



1. In the upper right corner, click **ADD TO REPORT**.
2. A verification panel opens. Click **ADD TO REPORT**.

Example:



1. This will initiate another sign in process to allow Data Studio to access Google Drive.
2. In the **Sign in** dialog, select your Qwiklabs student account.
3. Click **ALLOW** to give Google Data Studio permission to use the Google Drive resources in your lab account.

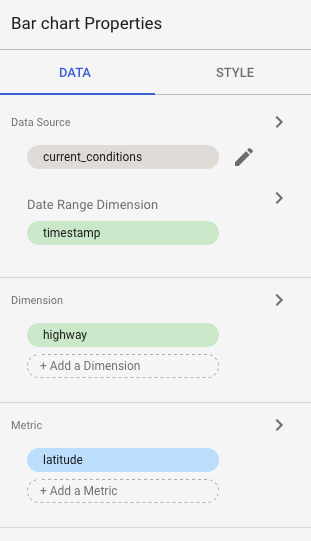
Giving Data Studio permission to Google Cloud account resources is typically a first-time activity and not something you would need to do every time you create a report.

## Task 3: Creating a bar chart using a calculated field

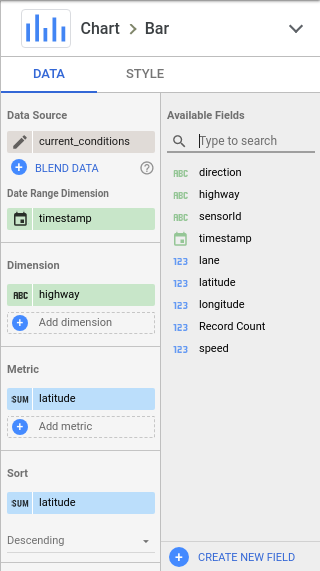
Once you have added the current\_conditions data source to the report, the next step is to create a visualization. Begin by creating a bar chart. The bar chart displays the total number of vehicles captured for each highway. To display this, you create a calculated field as follows.

1. (Optional) At the top of the page, click **Untitled Report** to change the report name. For example, type **<PROJECTID>-report1-yourname**.
2. When the report editor loads, click **Insert > Bar chart**.
3. In the **Bar chart** properties window, on the **Data** tab, notice the value for **Data Source** (**current\_conditions**) and the default values for **Dimension** and **Metric**.
4. If **Dimension** is not set to **highway**, then change **Dimension** to **highway**. In the **Dimension** section, click the existing dimension.

**Example**:



1. In the **Dimension picker**, select **highway**.
2. In the **Metric** section, click **Add metric** and add **latitude**.
3. In the **Metric** section, mouse over **Record Count** and click the **(x)** to remove it.
4. Click on **CREATE NEW FIELD** under the **Available Fields** section.



1. To display a count of the number of vehicles using each highway, create a calculated field. For this lab, you count the entries in the sensorId field. The value is irrespective, we just need the number of occurrences.
2. For **Field Name**, type **vehicles**.
3. Leave the **Field ID** unchanged.
4. For **Formula**, type the following (or use the formula assistant): **COUNT(sensorId)** and press **Enter**.
5. Click **SAVE**.
6. Click **DONE**.

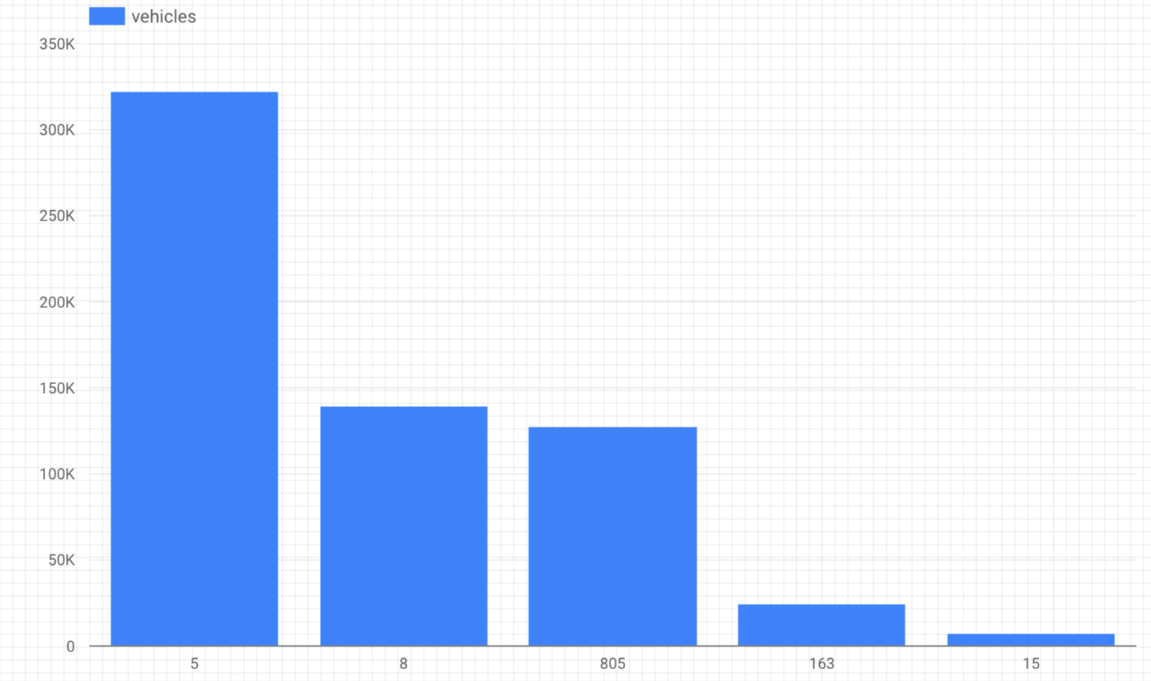
### **Add the metric**

1. In the **Metric picker,** In the **Metric** section, click **Add metric**.
2. Select **vehicles**.

The Bar Chart will show an error. Do you know why?

### **Verify that the value is numeric**

1. Click on the pencil next to **Data Source**, **current\_conditions**.
2. Examine the type associated with **vehicles**. If it is incorrectly set to timestamp, set it to **Numeric** > **Number**. Click **Done**. The error is corrected.
3. In the Metric section, mouse over **latitude** and click the **(x)** to remove it.
4. The Dimension should be set to highway and the Metric should be set to vehicles. Notice the chart is sorted in Descending order by default. The highway with the most vehicles are displayed first.



1. To enhance the chart, change the bar labels. In the **Bar chart** properties window, click the **STYLE** tab.
2. In the **Bar chart** section, check **Show data labels**.
3. The total number of vehicles is displayed above each bar in the chart.

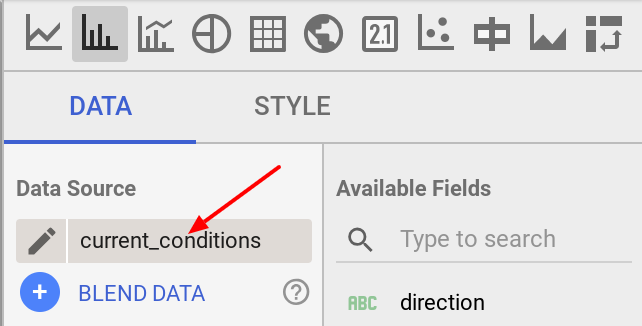
## Task 4: Creating a chart using a custom query

Because Data Studio does not allow aggregations on metrics, some report components are easier to generate using a custom SQL query. The Custom Query option also lets you leverage BigQuery's full query capabilities such as joins, unions, and analytical functions.

Alternatively, you can leverage BigQuery's full query capabilities by creating a view. A view is a virtual table defined by a SQL query. You can query data in a view by adding the dataset containing the view as a data source.

When you specify a SQL query as your BigQuery data source, the results of the query are in table format, which becomes the field definition (schema) for your data source. When you use a custom query as a data source, Data Studio uses your SQL as an inner select statement for each generated query to BigQuery. For more information on custom queries in Data Studio, consult the [online help](https://support.google.com/datastudio/?hl=en#topic=6267740).

1. To add a bar chart to your report that uses a custom query data source:
2. Click **Insert > Bar chart**.
3. In the **Bar chart** properties window, on the **Data** tab, notice the value for Data Source (current\_conditions) and the default values for Dimension and Metric are the same as the previous chart. In the **Data Source** section, click on given data source.



1. Click **Create new data source**.
2. For **Google Connectors**, select **BigQuery**.
3. For **My Projects**, click **Custom Query**.
4. For **Project**, select your project.
5. Type the following in the **Enter custom query** window and replace the <PROJECTID> with your Project ID.

SELECT max(speed) as maxspeed, min(speed) as minspeed,

avg(speed) as avgspeed, highway

FROM `<PROJECTID>.demos.current\_conditions`

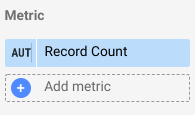
group by highway

This query uses max/min/avg functions to give you the corresponding speed for each highway.

1. At the top of the window, click **Untitled data source**, and change the data source name to **San Diego highway traffic summary**.
2. In the upper right corner of the window, click **Connect**. Once Data Studio has connected to the BigQuery data source, the results of the query are used to determine the table schema.
3. When the schema is displayed, notice the type and aggregation for each field.
4. Click **Add to report**.
5. When prompted, click **Add to report**.

Data Studio may be unable to determine the appropriate Dimension and Metrics for the chart. This requires you to adjust the graph options.

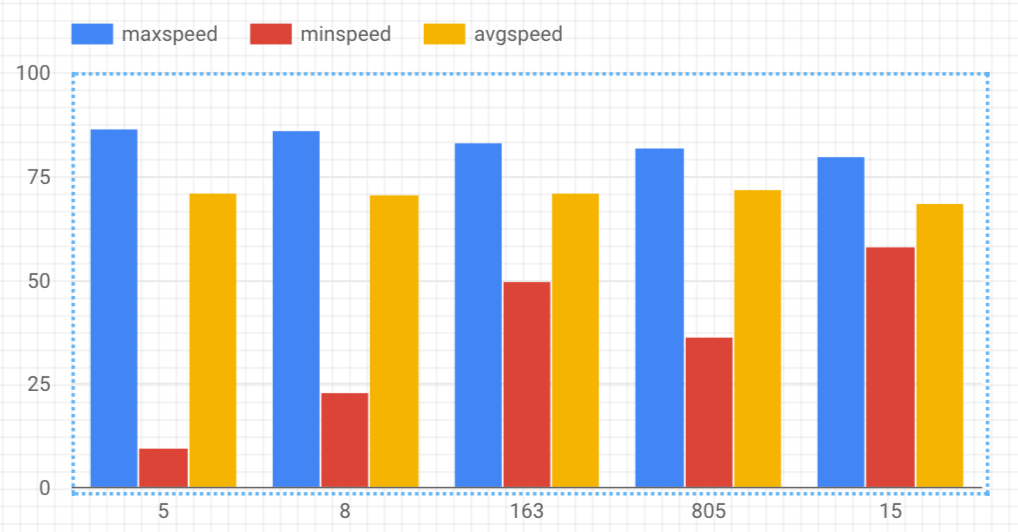
1. In the **Bar chart** properties, on the **Data** tab, in the Metric section, click **Record count**.



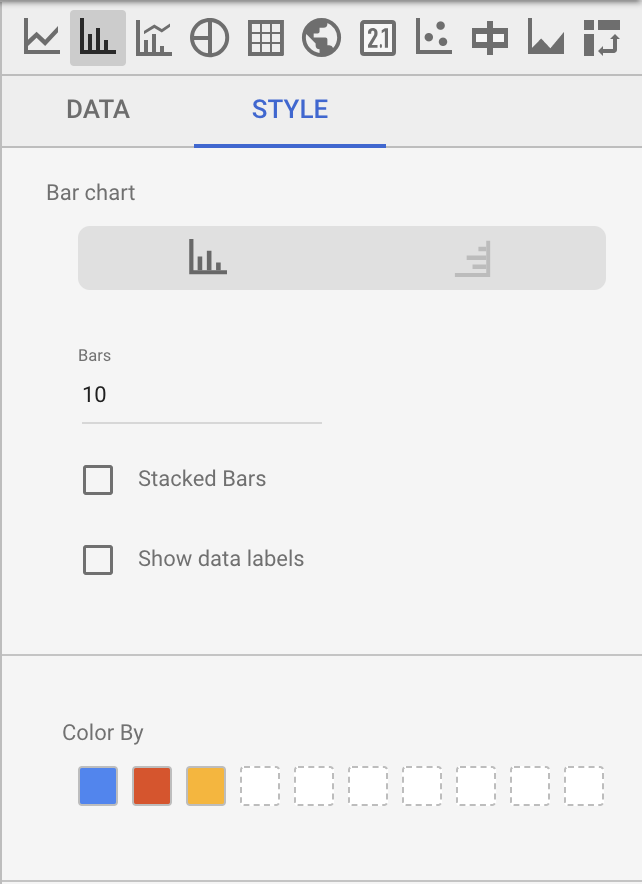
1. In the **Metric picker,** select **maxspeed**.
2. In the **Metric** section, click **Add metric**.
3. In the **Metric picker**, select **minspeed**.
4. In the **Metric** section, click **Add metric**.
5. In the **Metric picker**, select **avgspeed**.
6. Remove the metric other than maxspeed, minspeed and avgspeed, if exist.

Your chart now displays the maximum speed, minimum speed, and average speed for each highway.

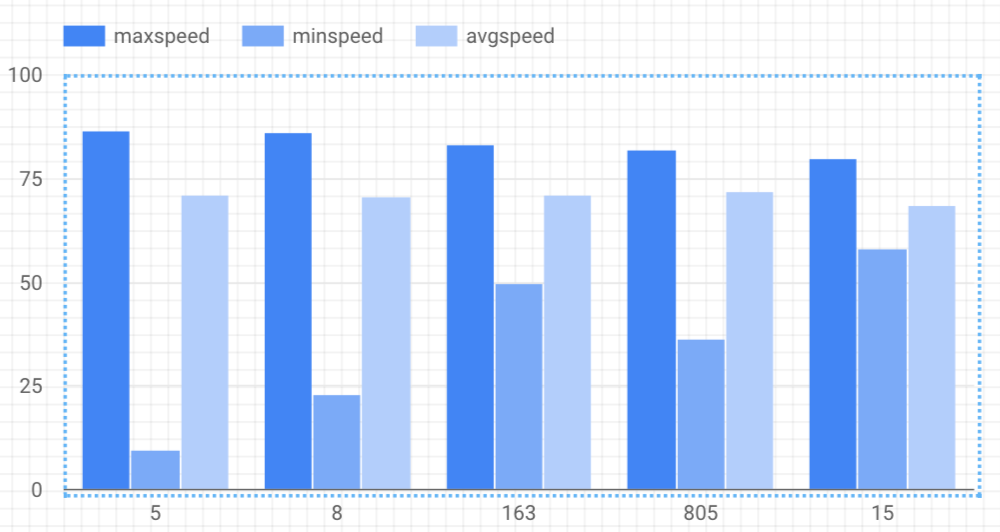
Notice each bar has a default color based on the order the metrics were added to the chart.



1. For readability, change the chart styles. In the **Bar chart** properties, click the **Style** tab.



1. In the **Color By** section, click on the boxes to select different colors.

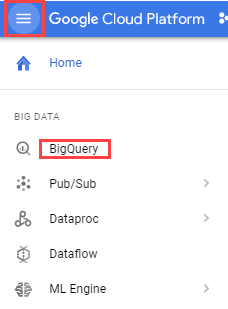


## Task 5: Viewing your query history

You can view queries submitted via the BigQuery Connector by examining your query history in the BigQuery web interface. Using the query history, you can estimate query costs, and you can save queries for use in other scenarios.

### **Open BigQuery Console**

In the Google Cloud Console, select **Navigation menu** > **BigQuery**:



The **Welcome to BigQuery in the Cloud Console** message box opens. This message box provides a link to the quickstart guide and lists UI updates.

Click **Done**.

1. In the left pane the first item in the list will be **Query history**. On your initial visit to the page the query history should appear on the bottom right underneath the Query editor pane. If it is not loaded click the **Query history** link.
2. The list of queries is displayed with the most recent queries first. Click on any Query to view details on the query such as Job ID and Bytes Processed.

