FREE eBook

LEARNING

# google-bigquery

Free unaffiliated eBook created from
Stack Overflow contributors.

#google-bigquery

Table of Contents

About 1

Chapter 1: Getting started with google-bigquery 2

Remarks 2

Examples 2

Installation or Setup 2

Analyzing 50 billion Wikipedia pageviews in 5 seconds (BigQuery beginner tutorial) 2

Setup 2

Let's query 2

Cost analysis 3

Loading data into BigQuery 4

Learn more 4

Follow for even more! 4

Chapter 2: Integration of Google's BigQuery with web application 5

Examples 5

Integration of Google's BigQuery API with web application 5

Credits 8

About

You can share this PDF with anyone you feel could benefit from it, downloaded the latest version from: google-bigquery

It is an unofficial and free google-bigquery ebook created for educational purposes. All the content is extracted from Stack Overflow Documentation, which is written by many hardworking individuals at Stack Overflow. It is neither affiliated with Stack Overflow nor official google-bigquery.

Use the content presented in this book at your own risk; it is not guaranteed to be correct nor accurate, please send your feedback and corrections to info@zzzprojects.com

Chapter 1: Getting started with google-bigquery

Remarks

This section provides an overview of what google-bigquery is, and why a developer might want to use it.

It should also mention any large subjects within google-bigquery, and link out to the related topics. Since the Documentation for google-bigquery is new, you may need to create initial versions of those related topics.

Examples

Installation or Setup

Detailed instructions on getting google-bigquery set up or installed.

Analyzing 50 billion Wikipedia pageviews in 5 seconds (BigQuery beginner tutorial)

Hi everyone! This is a demo I love running for people that get started with BigQuery. So let's run some simple queries to get **you** started.

Setup

You will need a Google Cloud project:

1. Go to http://bigquery.cloud.google.com/.
2. If it tells you to create a project, follow the link to create a project, and create a project.
3. Come back to http://bigquery.cloud.google.com/.

Notes:

- You don't need a credit card. Everyone gets a free 1TB for analysis each month.
- BigQuery charges per query. Before running a query you will be able to see how much each query costs.

Let's query

1. Find the pageviews for May 2015 at https://bigquery.cloud.google.com/table/fh-bigquery:wikipedia.pagecounts_201505

Note: Google protects your data with the highest security standards (PCI, ISO, HIPAA, SOC, etc), but it's also easy to share data if you want so - as I did here. https://cloud.google.com/security/

2. This table has 5 columns: *datehour language title requests content_size*. They basically say "this wikipedia page in this language had these many requests at this hour".

3. This table has almost 6 billion rows (379 GB of data).

4. To find out how many pageviews Wikipedia had during May, you can add up all the 6 billion lines of requests:

```
SELECT SUM(requests)
FROM [fh-bigquery:wikipedia.pagecounts_201505]
```

5. Did you notice how fast that was? (1.8s elapsed, 43.1 GB processed for me)

6. Let's do something more complex. Let's run a regular expression over these 6 billion rows. How fast could this be?

```
SELECT SUM(requests) req, title
FROM [fh-bigquery:wikipedia.pagecounts_201505]
WHERE REGEXP_MATCH(title, 'Red.*t')
GROUP BY title
ORDER BY req DESC
LIMIT 100
```

7. How fast was it for you? Did you find Reddit in the results?

Cost analysis

1. This last query processed 269 GB: More than a quarter of the free monthly terabyte. Why?

2. BigQuery looks at the columns you process on your query. 'title' is a big column - it contains text. The 'requests' column is only 43.1 GB.

3. To make your free terabyte last, extract data to smaller tables. For example, I have a table with only the top 65,000 English Wikipedia pages pageviews. The same query processes only 1.18 GB - you can run almost a 1000 of them for free a month.

```
SELECT SUM(requests) req, title
FROM [fh-bigquery:wikipedia.pagecounts_201408_en_top65k]
WHERE REGEXP_MATCH(title, 'Red.*t')
GROUP BY title
ORDER BY req DESC
LIMIT 100
```

4. You can't create tables with the free monthly terabyte - it's only for analysis. Activate your free $300 for new Google Cloud Platform accounts, or ask me here to do an extract for you. I will be happy to do so.

Loading data into BigQuery

To load data into BigQuery, you will need to activate billing for your project - try it with your free $300 for new accounts.

0. Create a dataset in your project to load the data to: https://i.imgur.com/FRClJ3K.jpg.

1. Find the raw logs shared by Wikipedia at https://dumps.wikimedia.org/other/pagecounts-raw/

2. wget one of these files into your computer, like https://dumps.wikimedia.org/other/pagecounts-raw/2015/2015-06/pagecounts-20150616-160000.gz

3. Install the 'bq' tool. https://cloud.google.com/bigquery/bq-command-line-tool

4. Load it into BigQuery:

   ```
    bq load -F" " --quote "" YourProjectName:DatasetName.pagecounts_20150616_16 pagecounts-
   20150616-160000.gz language,title,requests:integer,content_size:integer
   ```

5. Wait a couple minutes. While you wait, let me explain that line: This is not a CSV file, it's a space separated file (-F" ") that doesn't use quotes (--quote ""), we choose a destination table in a dataset in your project (remember to create the dataset first), we chose the file to load, and we define the 4 columns this file has.

6. Note that BigQuery will happily ingest .gz files, up to a certain size. For very large files it's better to un-compress them and put them in Google Cloud Storage first. That's what I did with the reddit comments that /u/Stuck_In_The_Matrix compiled. Those files were large, but BigQuery ingested them in 2 minutes or so.

Learn more

Ready for more advanced examples? See how to query Reddit and how to query the all the NYC taxi trips.

Follow for even more!

- Subscribe to /r/bigquery (and see the sidebar for more links).
- Follow me at https://twitter.com/felipehoffa

Read Getting started with google-bigquery online: https://riptutorial.com/google-bigquery/topic/4799/getting-started-with-google-bigquery

Chapter 2: Integration of Google's BigQuery with web application

Examples

Integration of Google's BigQuery API with web application

In this tutorial , I will explain how to integrate Google's BigQuery API with web application. My web application is going to get the data using BigQuery and plot a graph using d3.js and Javascript.

Each project on Google Developers Console has a clientID and you need to copy the clientID and put it as a config:

```
var gconfig = {
    'client_id': 'ClientID',
    'scope': 'https://www.googleapis.com/auth/bigquery'
};
```

BigQuery API can be accessed in the following way:

```
$.getScript("https://apis.google.com/js/client.js", function(d) {
  function loadGAPI() {
    setTimeout(function() {
      if (!gapi.client) {
        loadGAPI();
      } else {
        loadBigQuery();
      }
    }, 500);
  }

  function loadBigQuery() {
    gapi.client.load('bigquery', 'v2');
    setTimeout(function() {
      if (!gapi.client.bigquery) {
        loadBigQuery();
      } else {
        onClientLoadHandler();
      }
    }, 500);
  }

  loadGAPI();
});
```

Also you'll need to mention the query which you are going to retrieve the data:

```
function runQuery() {
    var request = gapi.client.bigquery.jobs.query({
      'projectId': "bigdatameetup-83116",
      'timeoutMs': '30000',
      'query': 'SELECT DATE(date ) as date,SUM(INTEGER(orders)) as total_orders FROM
[bigdatameetup-83116:Demo_Backup.orders] GROUP BY date ORDER BY date LIMIT 1000; '
    });
    request.execute(function(response) {
      var bqData = [];

      response.result.rows.forEach(function(d) {
        bqData.push({"date": d3.time.format("%Y-%m-%d").parse(d.f[0].v),
          "total_orders": +d.f[1].v});
      });

      drawLineChart(bqData);
    });
  }
```

The rest is the visualization, i.e the creation of Line Chart using d3.js:

```
function drawLineChart(bqData) {
  var WIDTH = config.width, HEIGHT = config.height;
  var Y_AXIS_LABEL = "total_orders";
  var X_DATA_PARSE = d3.time.format("%d-%b-%y").parse;
  var Y_DATA_PARSE = vida.number;
  var X_DATA_TICK = d3.time.format("%b-%y");
  var X_AXIS_COLUMN = "date";
  var Y_AXIS_COLUMN = "total_orders";
  var margin = {top: 20, right: 20, bottom: 30, left: 50},
      width = WIDTH - margin.left - margin.right,
      height = HEIGHT - margin.top - margin.bottom;
  var x = d3.time.scale()
      .range([0, width]);
    var y = d3.scale.linear()
      .range([height, 0]);
  var xAxis = d3.svg.axis()
      .scale(x)
      .orient("bottom")
      .tickFormat(X_DATA_TICK);
  var yAxis = d3.svg.axis()
      .scale(y)
      .orient("left")
      .tickFormat(function(d) {
        return d / 1000000 + "M";
      });
  var line = d3.svg.line()
      .interpolate("basis")
      .x(function(d) { return x(d.x_axis); })
      .y(function(d) { return y(d.y_axis); });

  var svg = d3.select("#canvas-svg").append("svg")
      .attr("width", width + margin.left + margin.right)
      .attr("height", height + margin.top + margin.bottom)
    .append("g")
      .attr("transform", "translate(" + margin.left + "," + margin.top + ")");

  bqData.forEach(function(d) {
    d.x_axis = d[X_AXIS_COLUMN];
    d.y_axis = d[Y_AXIS_COLUMN];
  });

  bqData.sort(function(a, b) {
    return (new Date(a.x_axis)) - (new Date(b.x_axis));
  });

  x.domain(d3.extent(bqData, function(d) { return d.x_axis; }));
  y.domain(d3.extent(bqData, function(d) { return d.y_axis; }));

  svg.append("g")
      .attr("class", "x axis")
      .attr("transform", "translate(0," + height + ")")
      .call(xAxis);

  svg.append("g")
      .attr("class", "y axis")
      .call(yAxis)
    .append("text")
      .attr("transform", "rotate(-90)")
      .attr("y", 6)
      .attr("dy", ".71em")
      .style("text-anchor", "end")
      .text(Y_AXIS_LABEL);

  svg.append("path")
      .datum(bqData)
      .attr("class", "line")
      .attr("d", line);

  }
```

In this example, I have chosen 'Amount' as x-axis and 'Date' as y axis from the public dataset:

> nyc_taxi_public

You can find the full working sample in this link.

`BigQuery Integration with WebApplication`

Read Integration of Google's BigQuery with web application online: https://riptutorial.com/google-bigquery/topic/6139/integration-of-google-s-bigquery-with-web-application

Credits

S. No	Chapters	Contributors
1	Getting started with google-bigquery	Community, Felipe Hoffa
2	Integration of Google's BigQuery with web application	Eduard Malakhov, JL-HN, Sajeetharan