Solutions (https://cloud.google.com/solutions/) Solutions

Analyzing Financial Time Series Using BigQuery and Datalab

This solution illustrates the power and utility of BigQuery (https://cloud.google.com/bigquery/) and Datalab (https://cloud.google.com/datalab/) as tools for quantitative analysis. The solution provides an introduction (this document) and gets you set up to run a notebook-based Datalab tutorial.

If you're a quantitative analyst, you use a variety of tools and techniques to mine big data, such as market transaction histories, for information that can give you insight into market trends. Because quotes and trades happen at predictable intervals, such data represents a financial time series that you can analyze by using established techniques, including frequency analysis and moving averages.

But dealing with massive datasets can be challenging. Traditional tools might not scale as the dataset grows. Storage requirements can grow as fast as the dataset, so downloading data to your computer's hard drive is no longer a workable approach. And it can take a long time to retrieve the right data subsets from a traditional database query.

BigQuery solves these issues by enabling you to run SQL queries and to get results quickly through the processing power of Google's infrastructure. You can use BigQuery on the web, and you can use it on the command line and through APIs. When combined with other components of Google Cloud or third-party tools, BigQuery enables you to build the data-analysis applications you need now yet still be confident that you can scale them in the future.

In this solution, you use a powerful pattern for data analysis: BigQuery takes care of the heavy lifting in SQL, and Datalab does detailed data manipulation and visualization in Python.

Security is always important when working with financial data. Google Cloud helps to keep your data safe, secure, and private (https://support.google.com/googleforwork/answer/6057301?rd=1) in

several ways, and all data is encrypted during transmission and at rest. Google Cloud is also ISO 27001, ISO 27017, ISO 27018, SOC3, FINRA, and PCI compliant. (https://cloud.google.com/security/compliance/)

Objectives

- Load a dataset into BigQuery.
- Use BigQuery and Datalab to query financial time-series data.
- Visualize your query results in Datalab.

Costs

This tutorial uses the following billable components of Google Cloud:

- **Datalab**: The resources needed to run Datalab on Google Cloud are billable. These resources include one Compute Engine virtual machine, two persistent disks, and space for Cloud Storage backups. For details, refer to the Datalab Pricing page (https://cloud.google.com/datalab/docs/resources/pricing).
- **BigQuery**: This tutorial stores close to 100 MB of data in BigQuery and processes under 300 MB to execute the queries once. This amount of data is covered under the free limits provided by BigQuery each month. For complete details about BigQuery costs, see the BigQuery Pricing page (https://cloud.google.com/bigquery/pricing).

To generate a cost estimate based on your projected usage, use the pricing calculator (https://cloud.google.com/products/calculator). New Google Cloud users might be eligible for a free trial (https://cloud.google.com/free-trial).

Before you begin

Before you start the tutorial, you need to set up Datalab.

Use Cloud Shell

LAUNCH FROM CLOUD SHELL (HTTPS://CLOUD.GOOGLE.COM/DATALAB/DOCS/HOW-TO/DATALAB-USING-

Use the Cloud SDK

If you have the SDK installed:

LAUNCH FROM THE CLOUD SDK (HTTPS://CLOUD.GOOGLE.COM/DATALAB/DOCS/QUICKSTART)

If you *don't* have the SDK installed but want to use it to set up Datalab: Install and initialize the Cloud SDK (https://cloud.google.com/sdk/docs/).
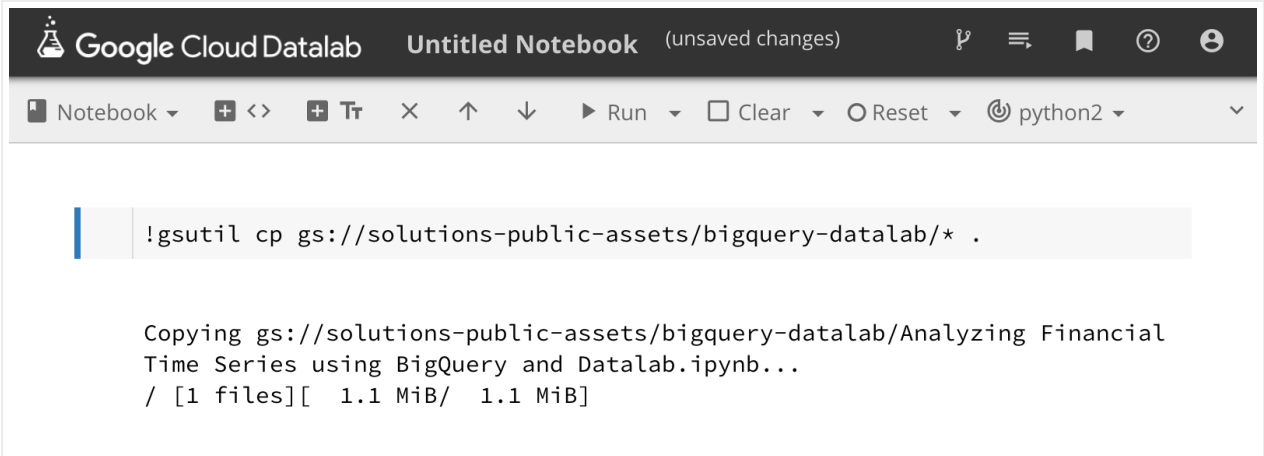
Completing the tutorial in the notebook

1. On the Datalab home page, add a new notebook by clicking ➕ **Notebook** on the top left.

   A new tab that contains an empty notebook with a code cell opens in your browser.

2. Copy the following code into that cell and click **Run** to execute it.

```
!gsutil cp gs://solutions-public-assets/bigquery-datalab/* .
```

3. Return to the original tab to see additional files. Click **Analyzing Financial Time Series using BigQuery and Datalab.ipynb** to begin working interactively through the tutorial.
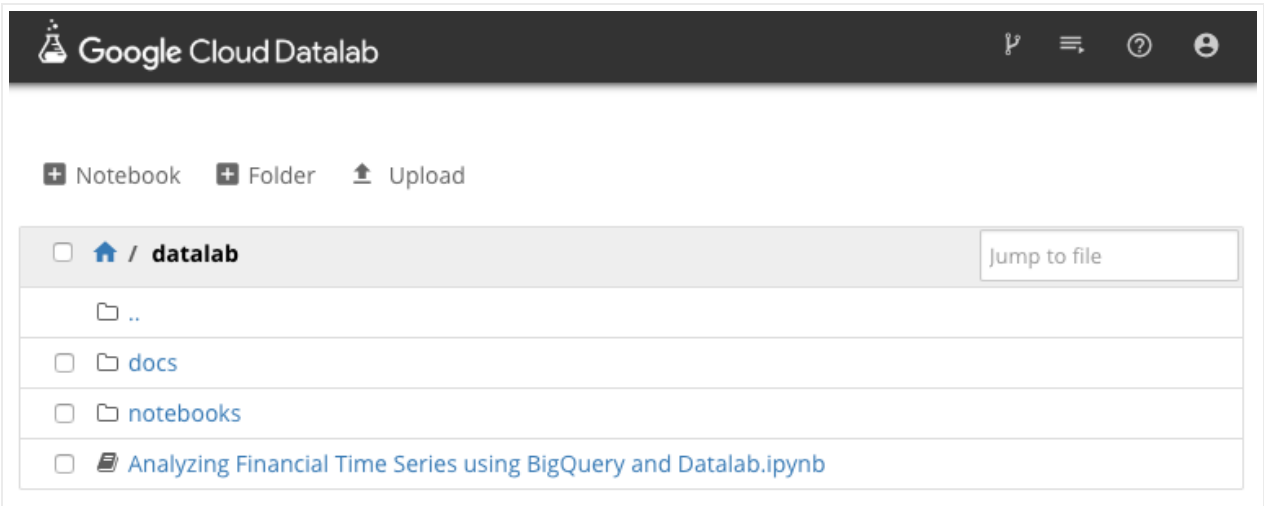
4. If you are unfamiliar with Datalab notebooks, go through the **Introduction to Notebooks.ipynb** document that is in the **docs / intro** subfolder.
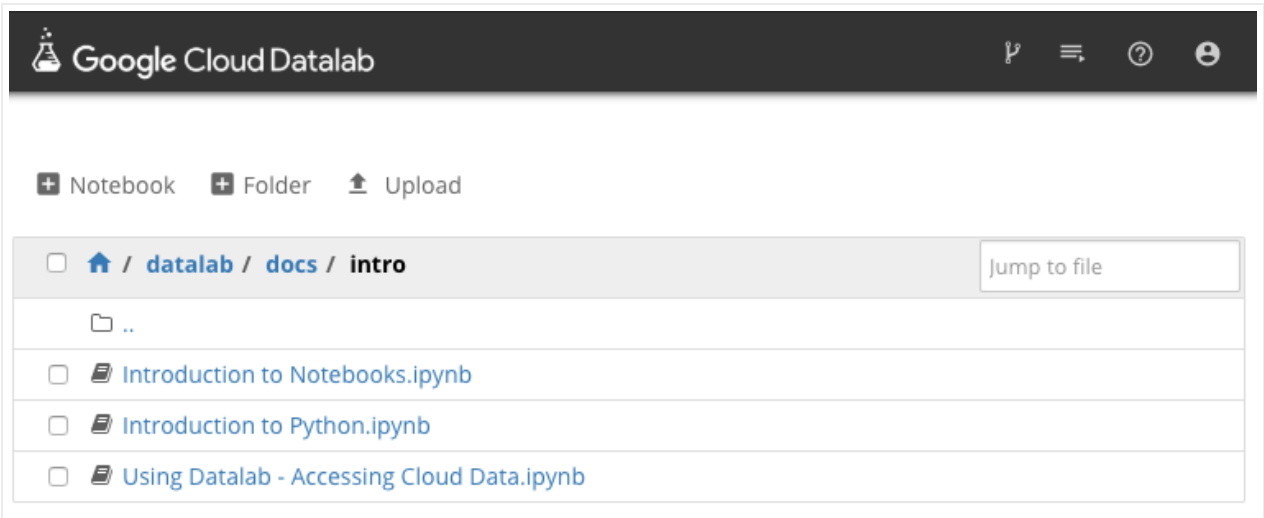
5. Follow the rest of the tutorial in the notebook.