Project

MAANG Stock Prices Analysis

Milestone 6

Cloud Problem Definition

Indra Kumar Chandaka

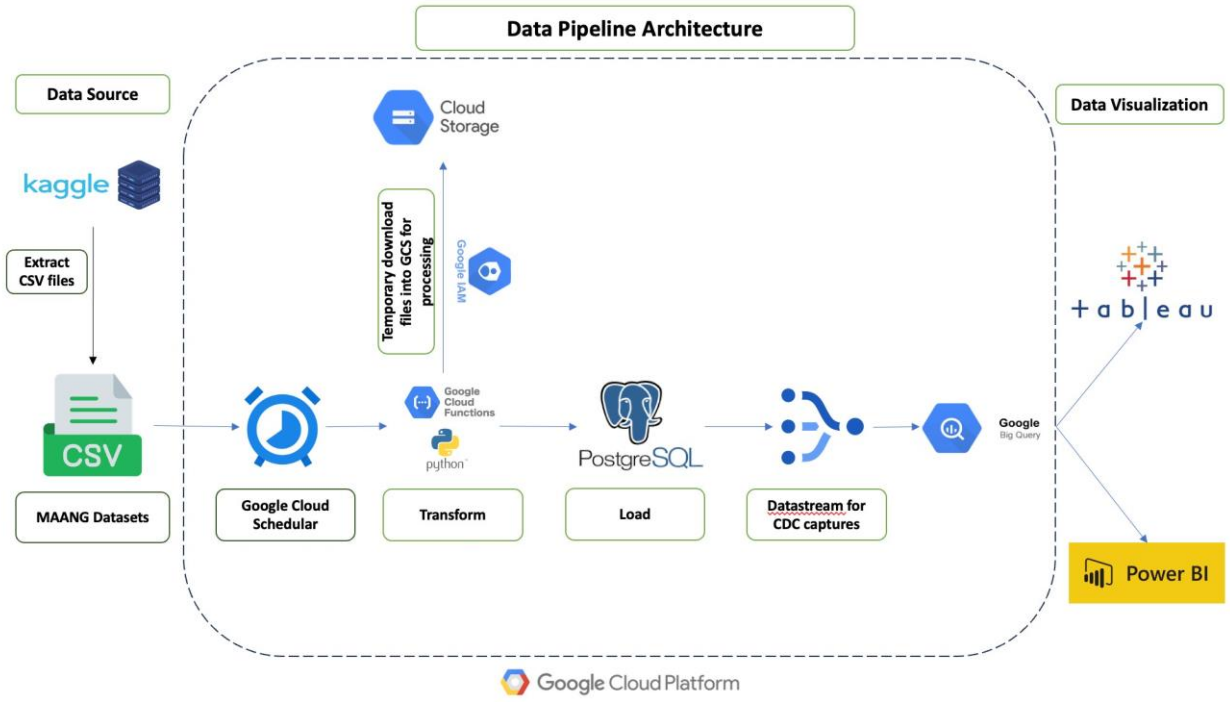
ichandak@gitam.in

Submission Date: Dec 1st, 2023

Objec&ve:

This project seeks to develop a data driven solu4on to address the complexi4es of analyzing the historical stock prices of MAANG companies and iden4fy the market condi4ons and external factors that could help make informed investment decisions. With this, the project aims to contribute to a broader understanding of MAANG stock prices, providing investors with niche perspec4ves of market dynamics and risks involved. The project aims to predict the effec4veness of trading strategies by back tracing with historical data insights and provide ac4onable insights to make data driven decisions about the trends, opportuni4es, and vola4lity.
Furthermore, by developing an efficient cloud-based data engineering solu4on which can handle large volumes of data for real 4me analysis.

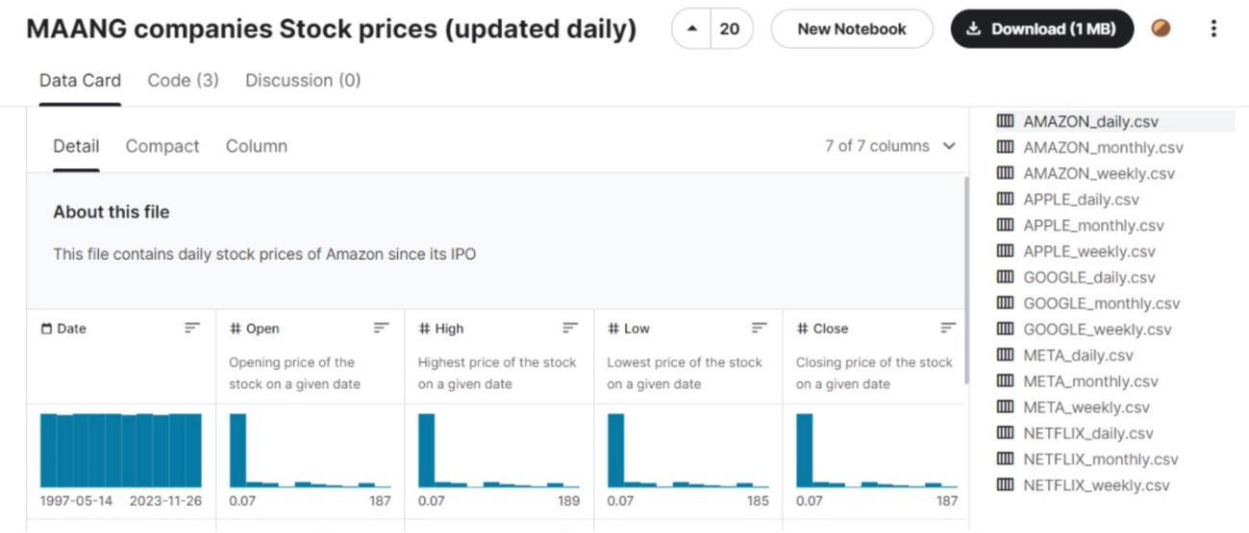
Data Pipeline Architecture:



The MAANG stocks dataset from Kaggle is updated on a daily, weekly, and monthly basis. Here our aim is to retrieve the data daily and transform it to the database (Postgres) using the Kaggle API. ANer this, we read the data in cloud func4ons using python and this is triggered by Goggle cloud scheduler based on the requirements. As the cloud func4on is serverless, we need a temporary storage which can be used to hold the extracted files in google cloud storage. Then we transform the data (data types, renaming columns, adding columns) and then files are ingested to three tables based on daily, weekly, and monthly into the Postgres DB. Then we use DataStream to capture for the changes (Auto incremental load) and move it into BigQuery.

Data source:

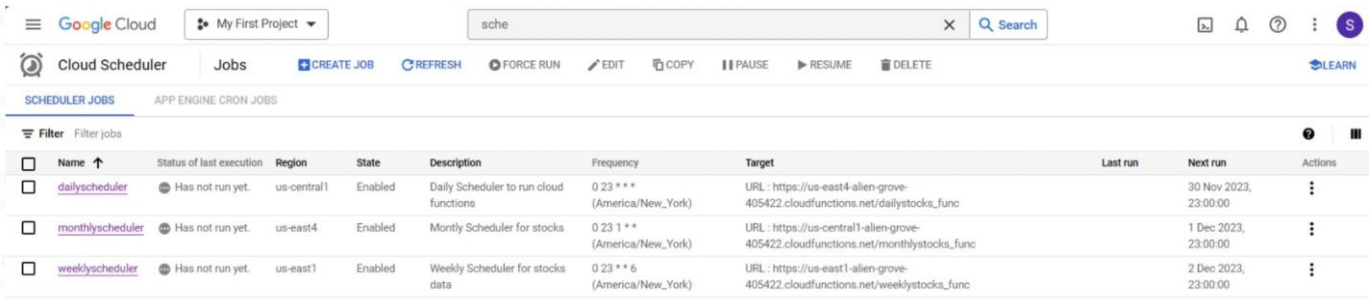
h"ps://www.kaggle.com/datasets/nikhil1e9/ne7lix-stock-price



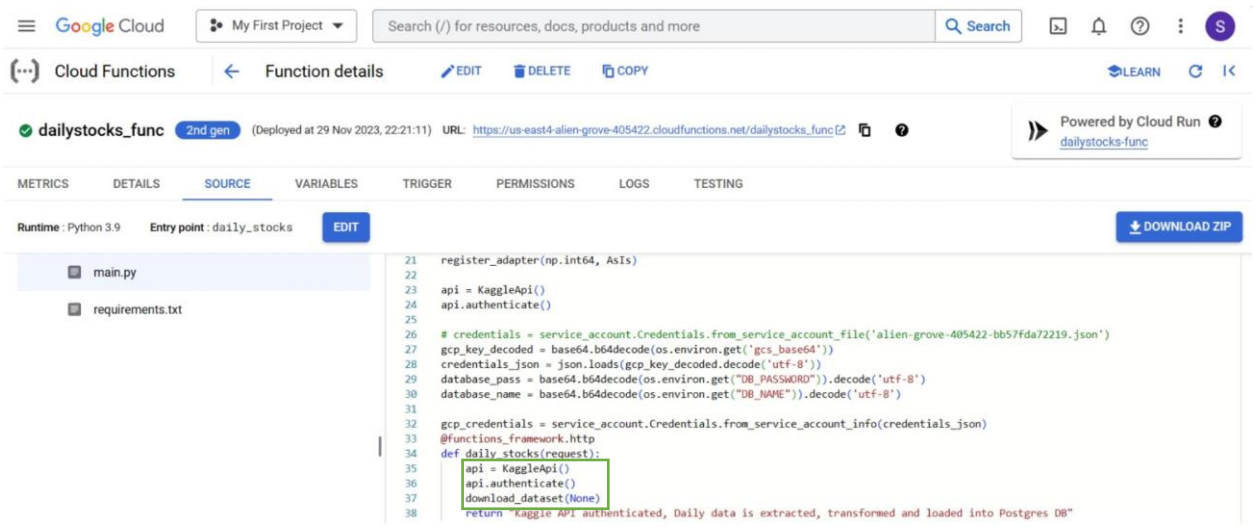
Data Inges&on:

We have adopted py scripts for daily, weekly, and monthly to transform and insert the 15 files into cloud postgres. Here, we used the Kaggle API instead of downloading the 15 files into the local or cloud bucket, as the data is very dynamic and changes daily.

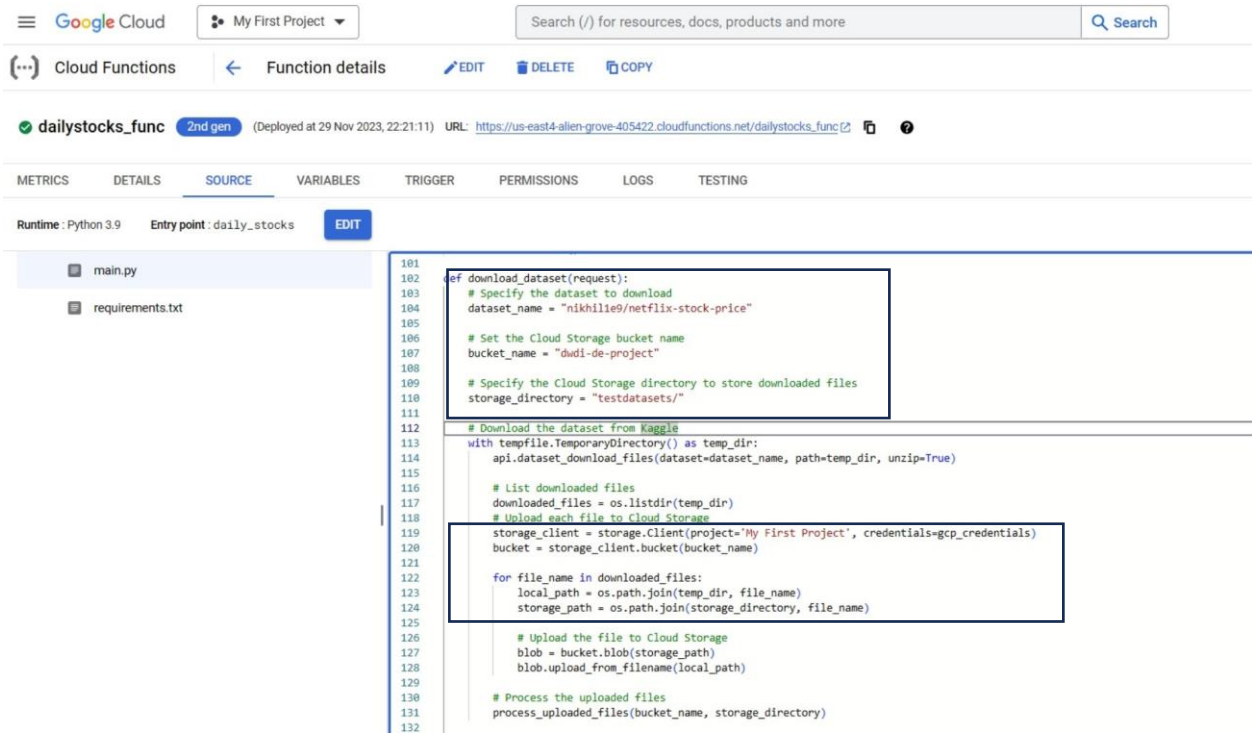
We are cloud scheduler to trigger the cloud func4ons.



Authen4ca4on of Kaggle API by providing creden4als:

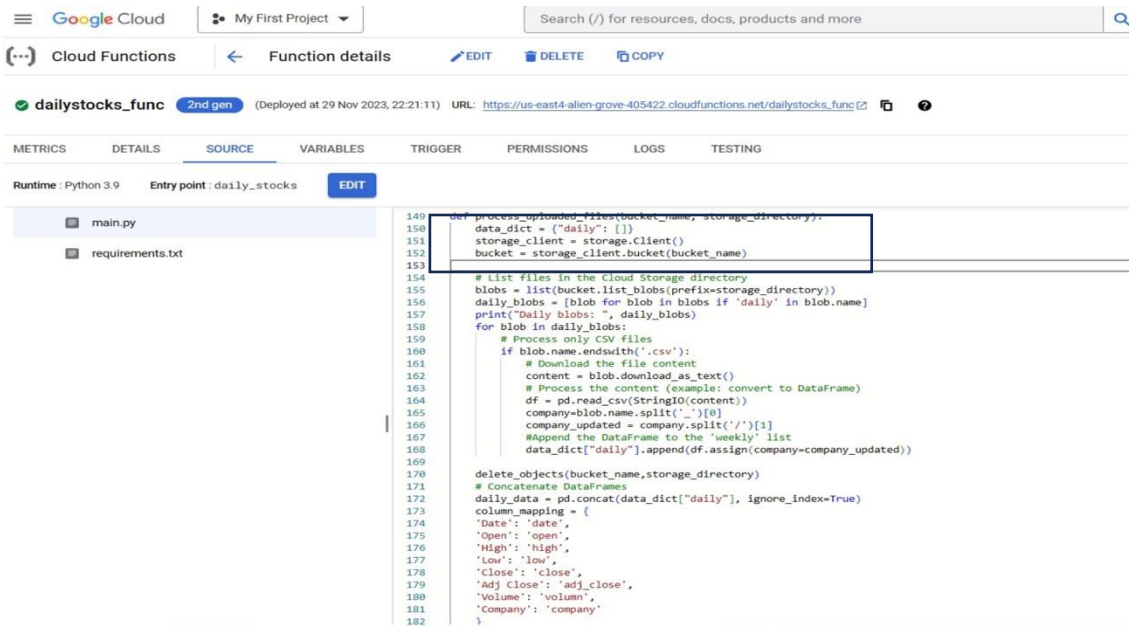


Providing the dataset link to download from Kaggle. Furthermore, the details of temporary storage bucket are also provided. It automa4cally authen4cates using GCP default creden4als to get the bucket access:

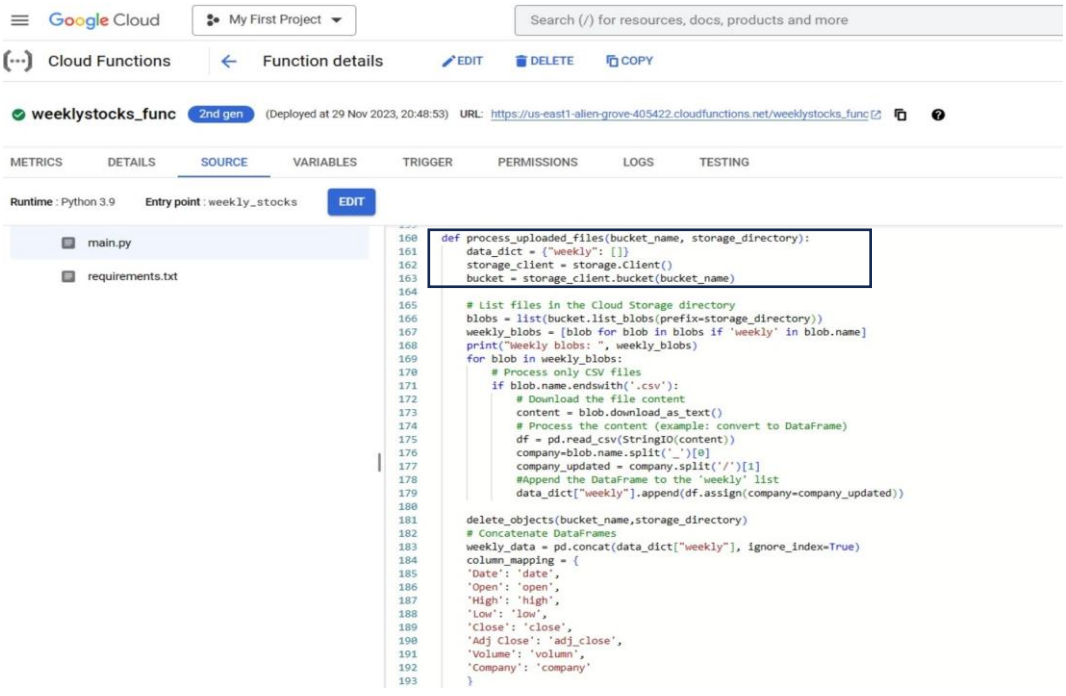


Now we call the process to upload files into cloud storage by providing the temporary path. Below is the screenshot for Daily, weekly, and monthly. Here the data is read and transformed. Now here we deleted the data stored in the bucket i.e., blobs.

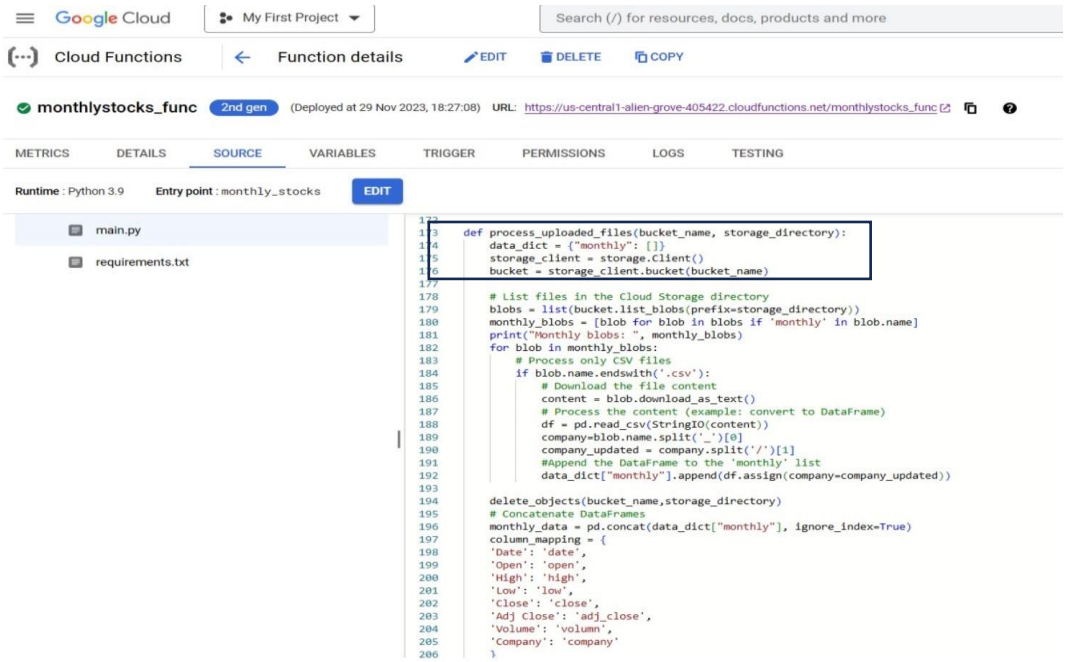
Daily -



Weekly –

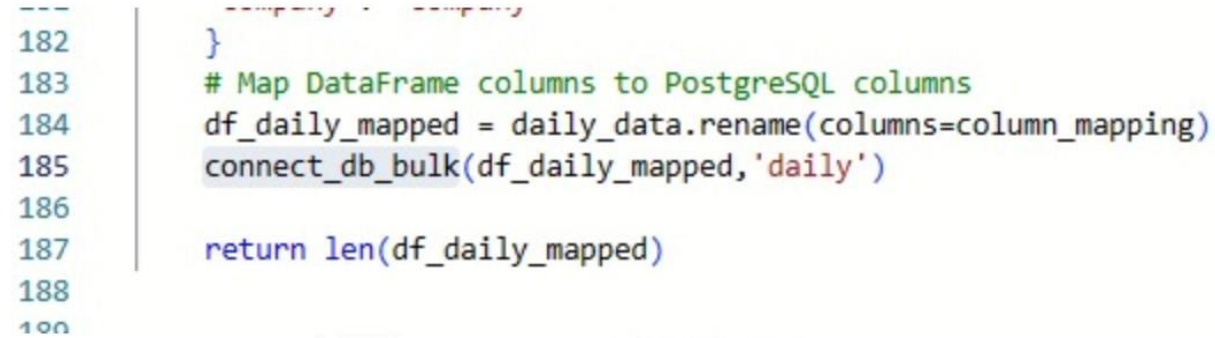


Monthly -

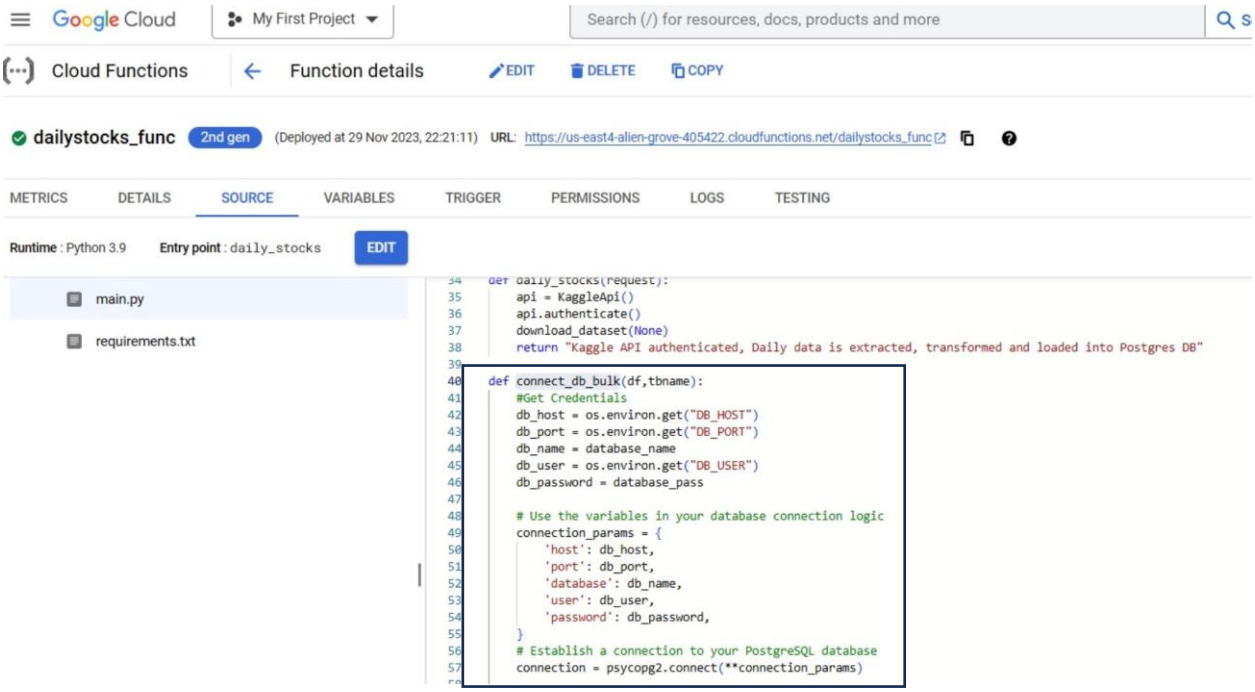


Now, aNer this the data gets downloaded into the cloud storage bucket for temporary purpose. We read the extracted files using the py script and read the files based on the files name and then read it as dataframes to transform such as renaming columns, adding column (Company name) and modifying the datatypes. Then we check now if the data is already present in PostGres Cloud using the py script. In case it is not present, we insert the data required to avoid data redundancy.

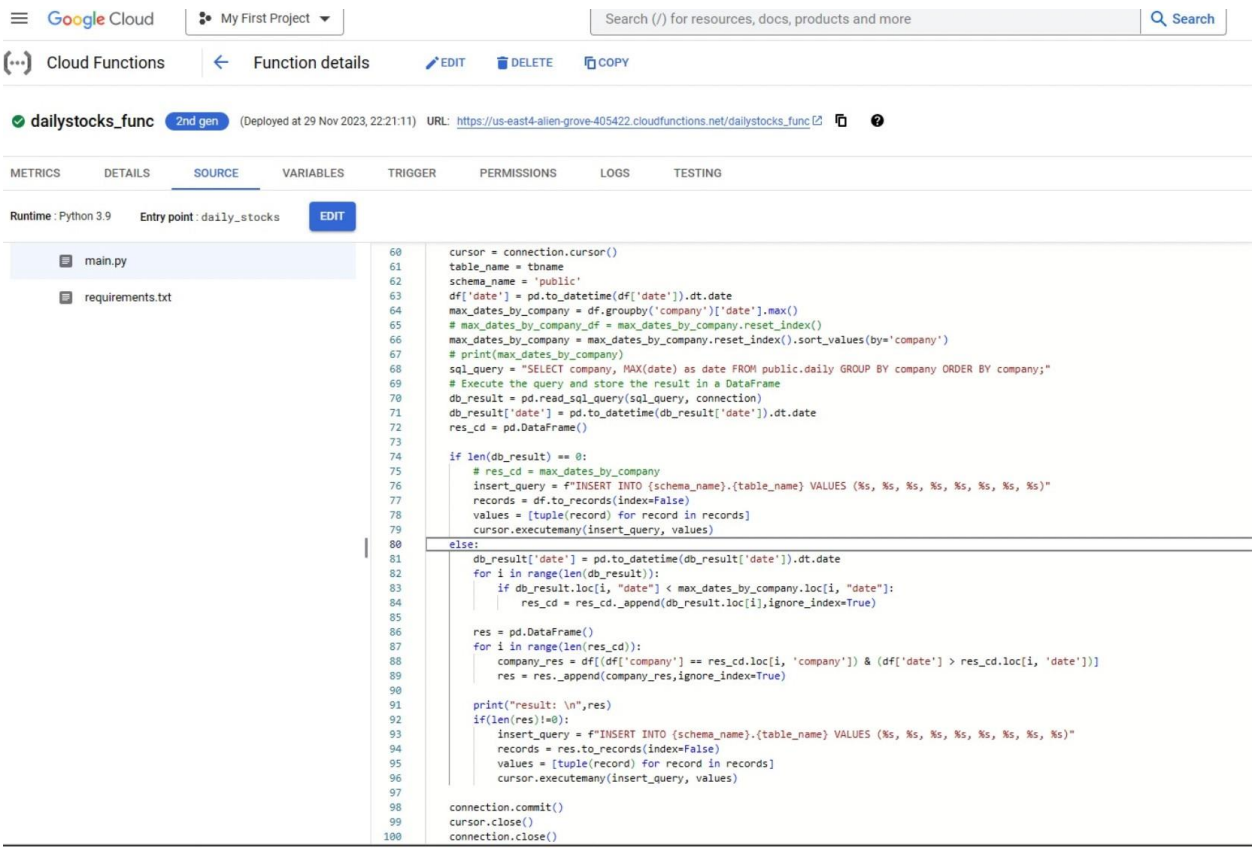
ANer reading the data, we call the connectDB func4on to insert it into Cloud Postgres. Now the transformed data is fed into cloud postgres.



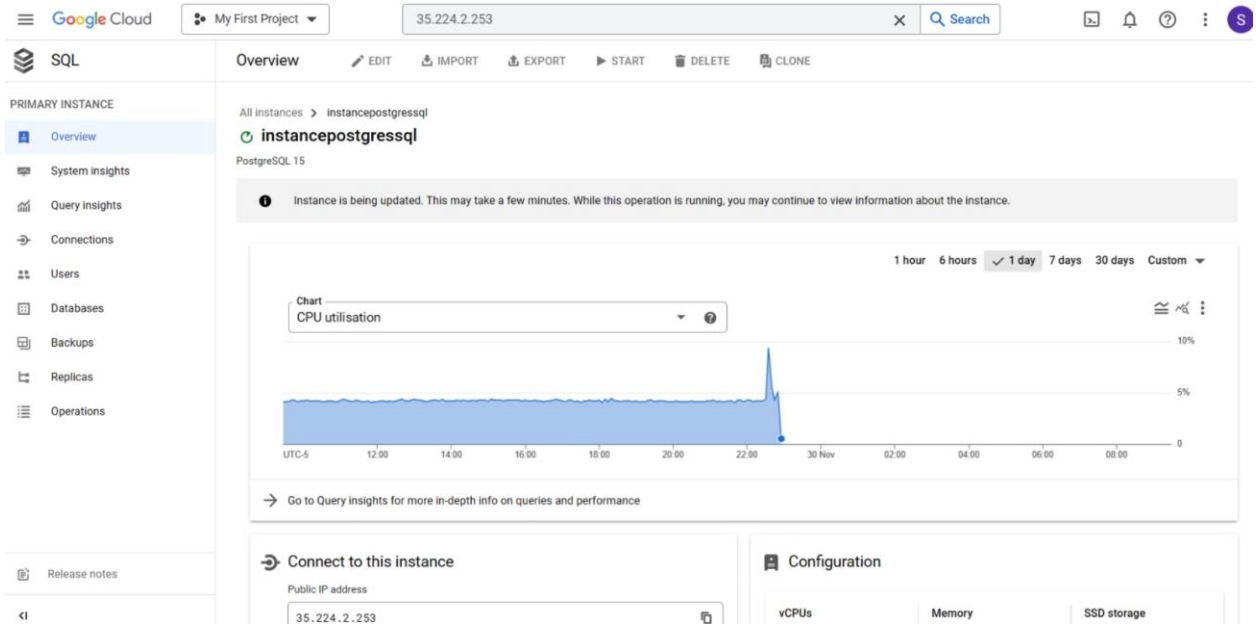
Now we provide the DB connec4ons to connect to postgres DB and the password and DB name has been encoded in Base 64 format to ensure security.



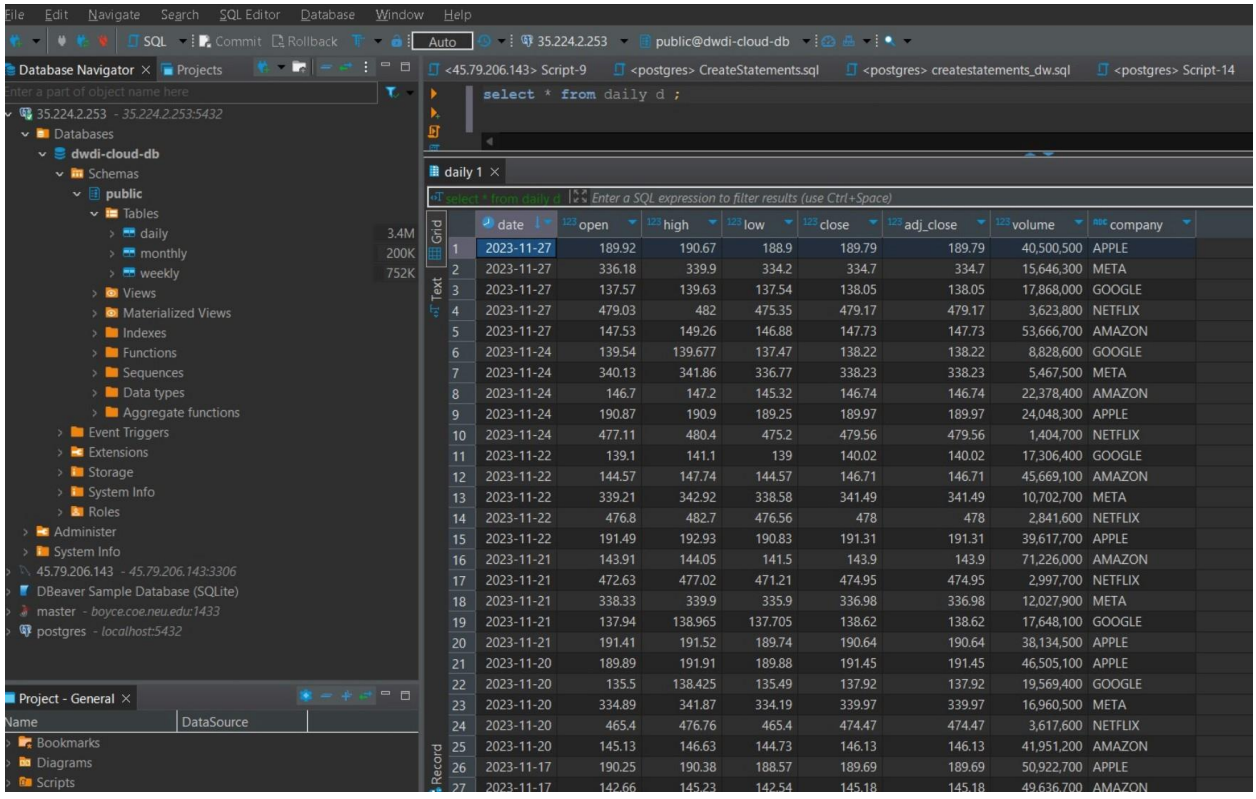
We are now checking of the data is present or not, if not we insert the data in order to avoid data redundancy.



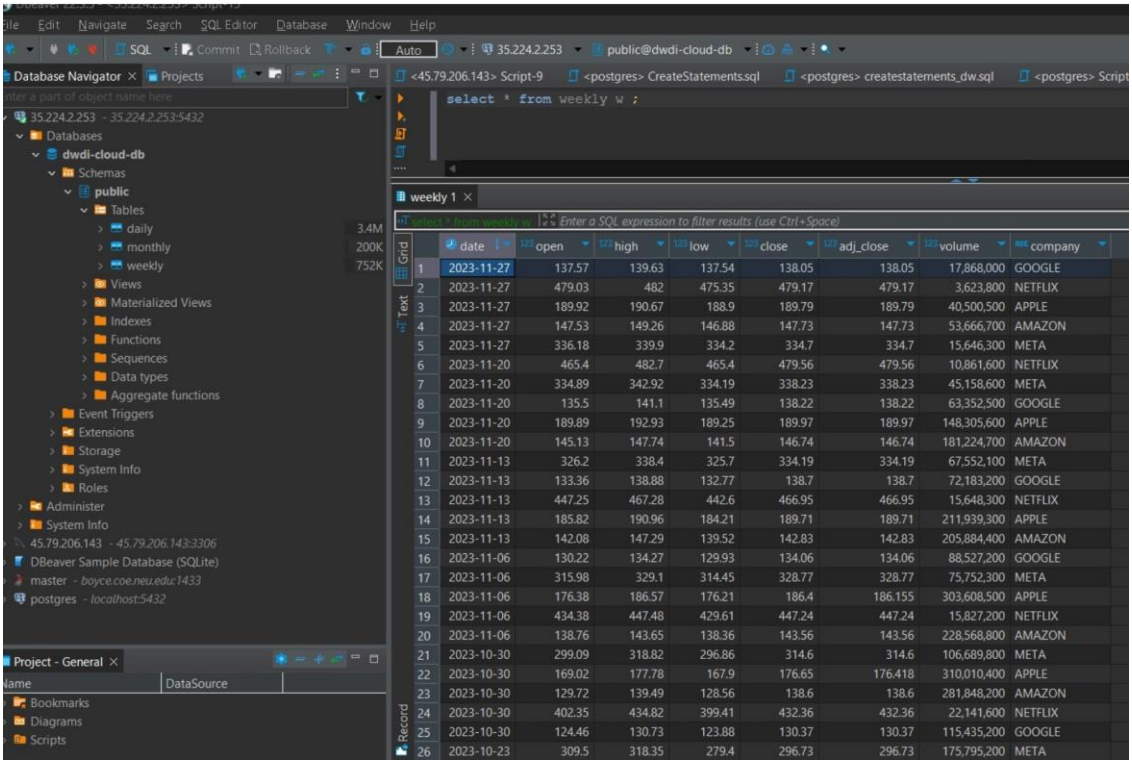
Loading into Postgres:



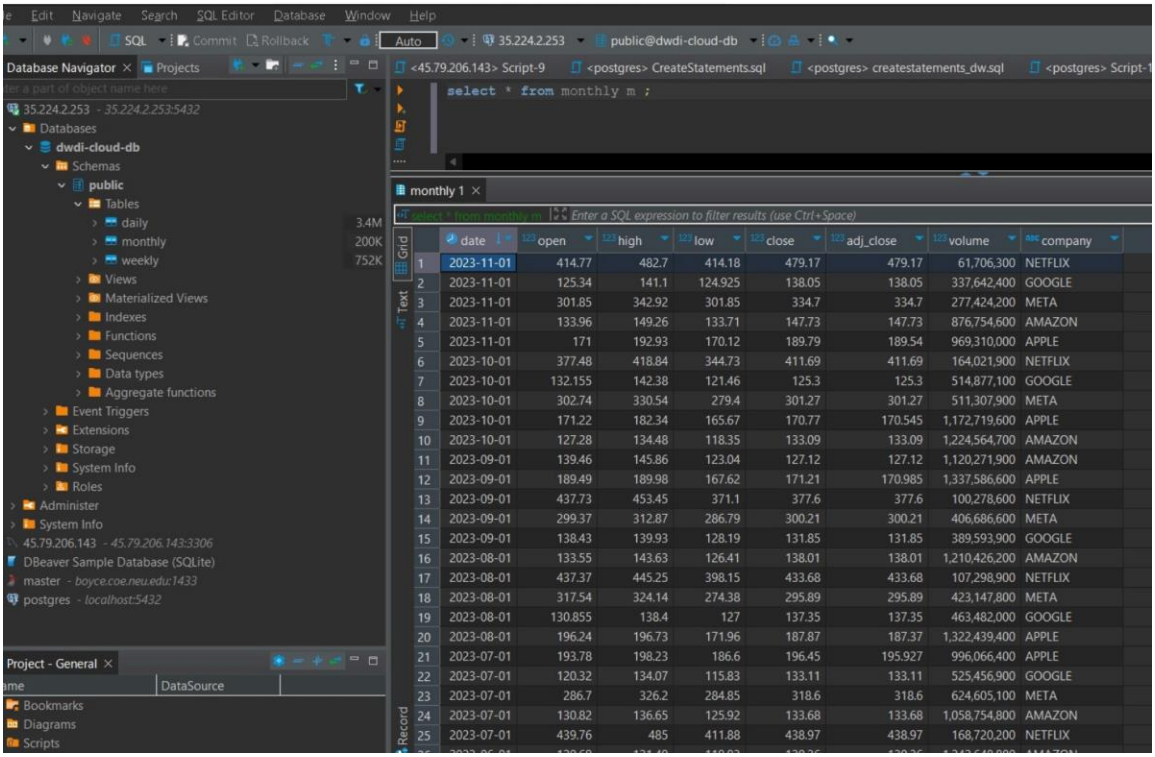
Daily data:



Weekly data:

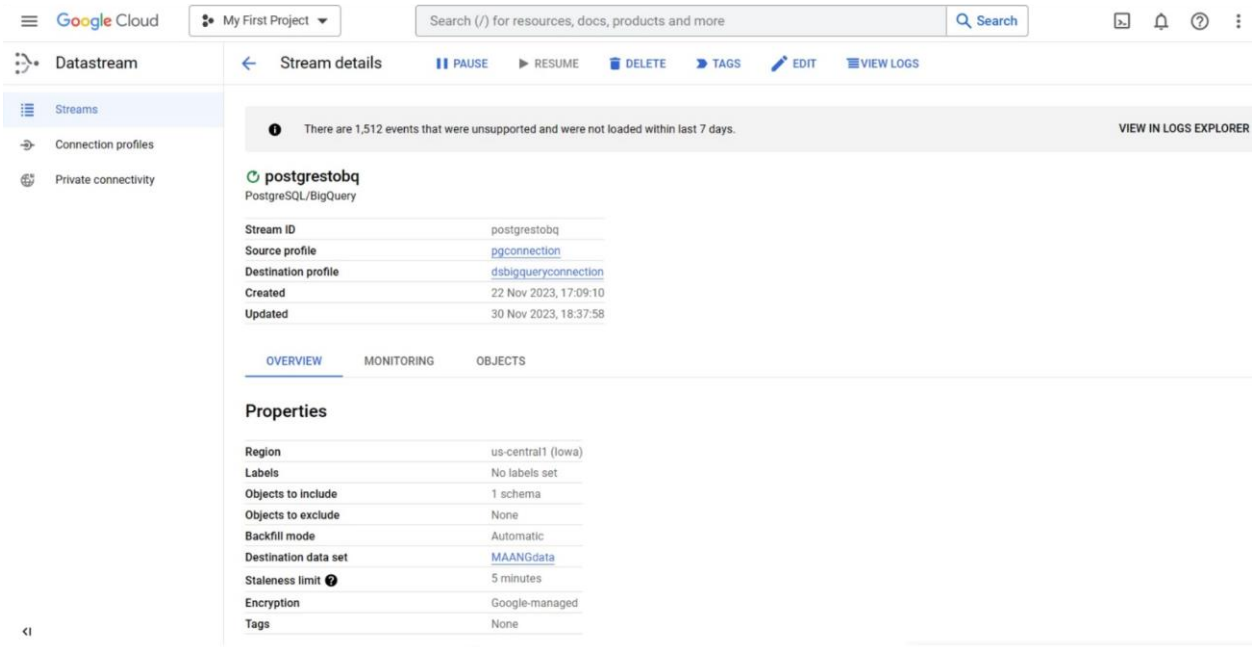


Monthly data:

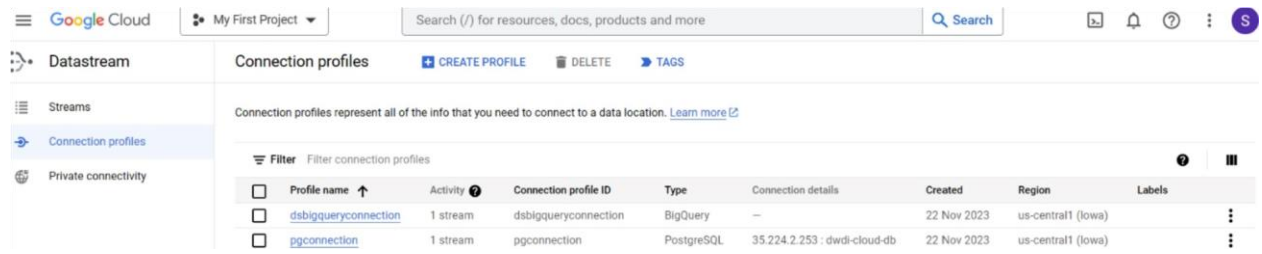


Data Inges&on into BigQuery:

Now the updated data in postgres will be inserted into BigQuery using DataStream.

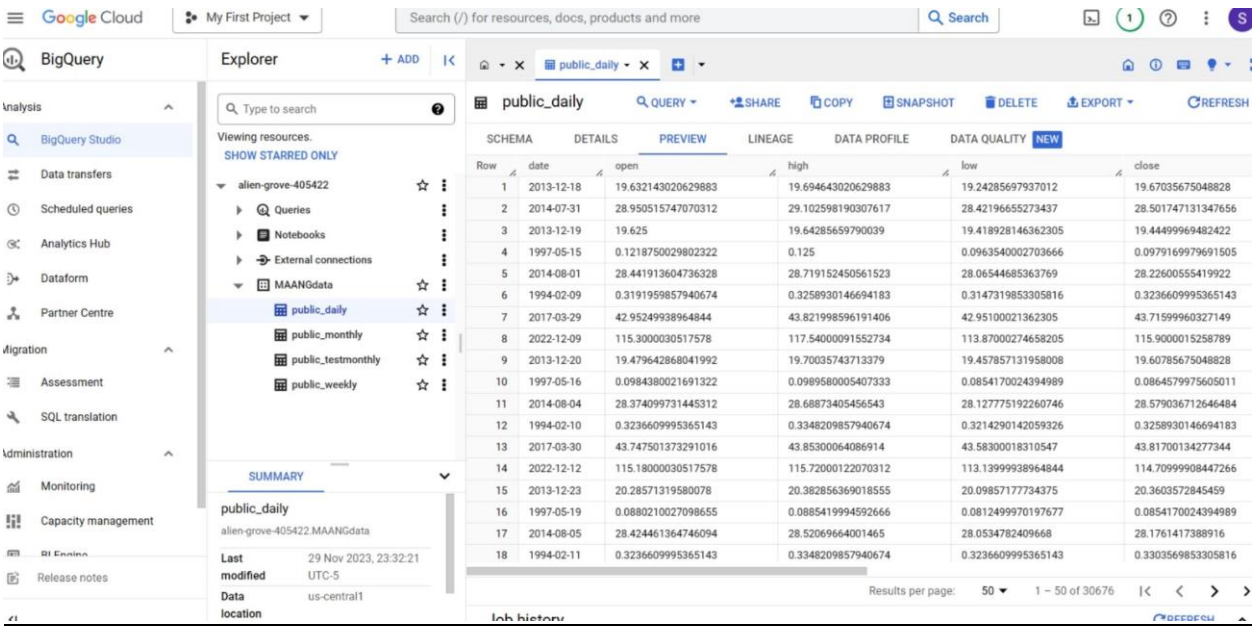


Below is the connec4on profile in DataStream:

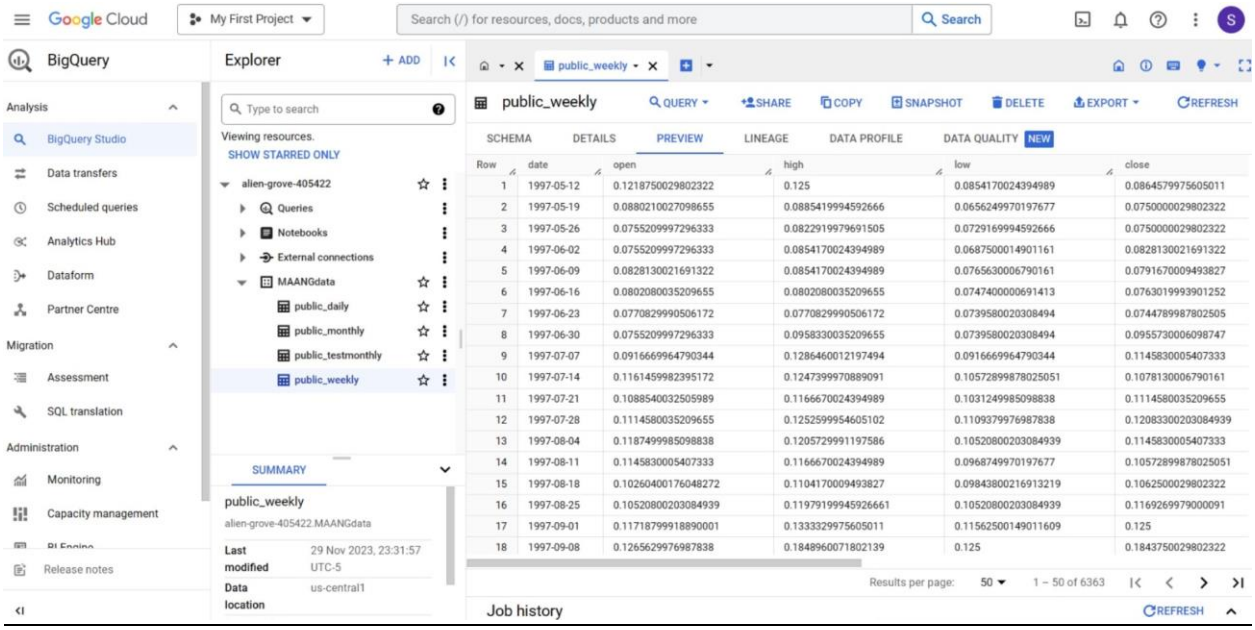


In order to capture the changes in the postgres database, we implemented DataStream to capture the incremental load in data in BigQuery.

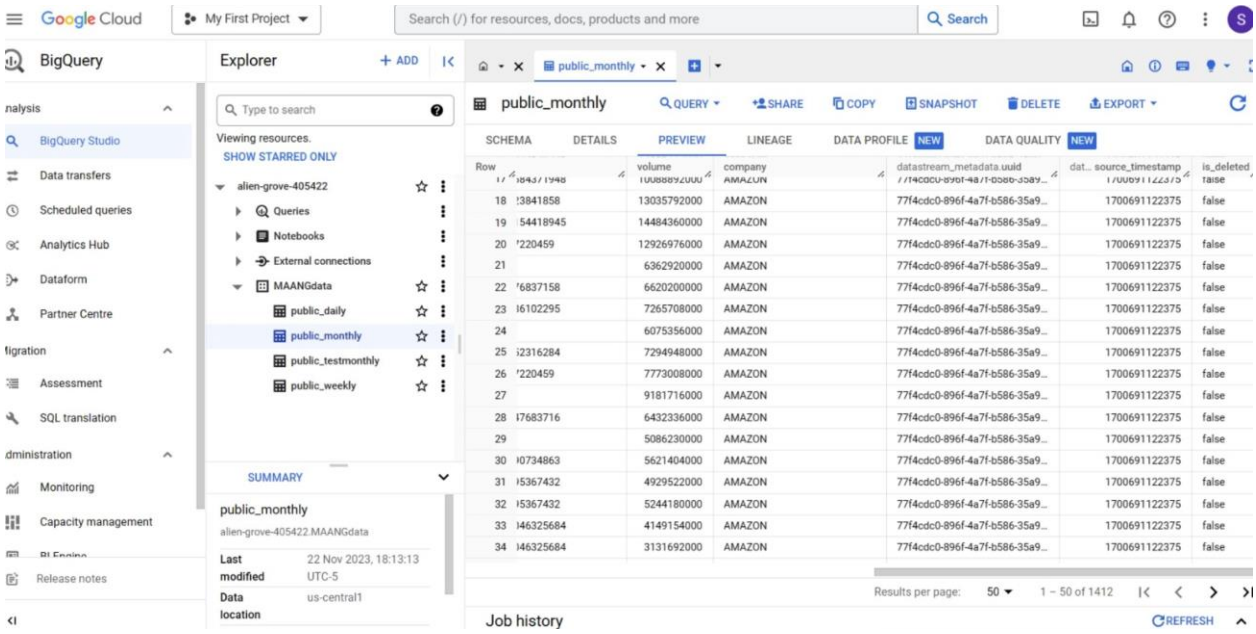
Daily:


Google Cloud
My First Project
Search (/) for resources, docs, products and more
Search
BigQuery
Explorer
+ ADD
Type to search
Viewing resources.
SHOW STARRED ONLY
alien-grove-405422
Queries
Notebooks
External connections
MAANGdata
public_daily
public_monthly
public_testmonthly
public_weekly
BigQuery Studio
Data transfers
Scheduled queries
Analytics Hub
Dataform
Partner Centre
Assessment
SQL translation
Monitoring
Capacity management
Release notes
SUMMARY
public_daily
alien-grove-405422.MAANGdata
Last modified
29 Nov 2023, 23:32:21
UTC-5
Data location
us-central1
public_daily
QUERY
SHARE
COPY
SNAPSHOT
DELETE
EXPORT
REFRESH
SCHEMA
DETAILS
PREVIEW
LINEAGE
DATA PROFILE
DATA QUALITY
NEW
Row
date
open
high
low
close
1 2013-12-18 19.632143020629883 19.694643020629883 19.24285697937012 19.67035675048828
2 2014-07-31 28.950515747070312 29.102598190307617 28.42196655273437 28.501747131347656
3 2013-12-19 19.625 19.64285659790039 19.418928146362305 19.44499969482422
4 1997-05-15 0.1218750029802322 0.125 0.0963540002703666 0.0979169979691505
5 2014-08-01 28.441913604736328 28.719152450561523 28.06544685363769 28.22600555419922
6 1994-02-09 0.3191959857940674 0.3258930146694183 0.3147319853305816 0.3236609995365143
7 2017-03-29 42.95249938964844 43.821998596191406 42.95100021362305 43.71599960327149
8 2022-12-09 115.3000030517578 117.54000091552734 113.87000274658205 115.9000015258789
9 2013-12-20 19.479642868041992 19.70035743713379 19.457857131958008 19.60785675048828
10 1997-05-16 0.0984380021691322 0.0989580005407333 0.0854170024394989 0.0864579975605011
11 2014-08-04 28.374099731445312 28.68873405456543 28.127775192260746 28.579036712646484
12 1994-02-10 0.3236609995365143 0.3348209857940674 0.3214290142059326 0.3258930146694183
13 2017-03-30 43.747501373291016 43.85300064086914 43.58300018310547 43.81700134277344
14 2022-12-12 115.18000030517578 115.72000122070312 113.13999938964844 114.70999908447266
15 2013-12-23 20.28571319580078 20.382856369018555 20.09857177734375 20.3603572845459
16 1997-05-19 0.0880210027098655 0.0885419994592666 0.0812499970197677 0.0854170024394989
17 2014-08-05 28.424461364746094 28.52069664001465 28.0534782409668 28.1761417388916
18 1994-02-11 0.3236609995365143 0.3348209857940674 0.3236609995365143 0.3303569853305816
Results per page: 50
1 – 50 of 30676
Job history
REFRESH

Weekly:

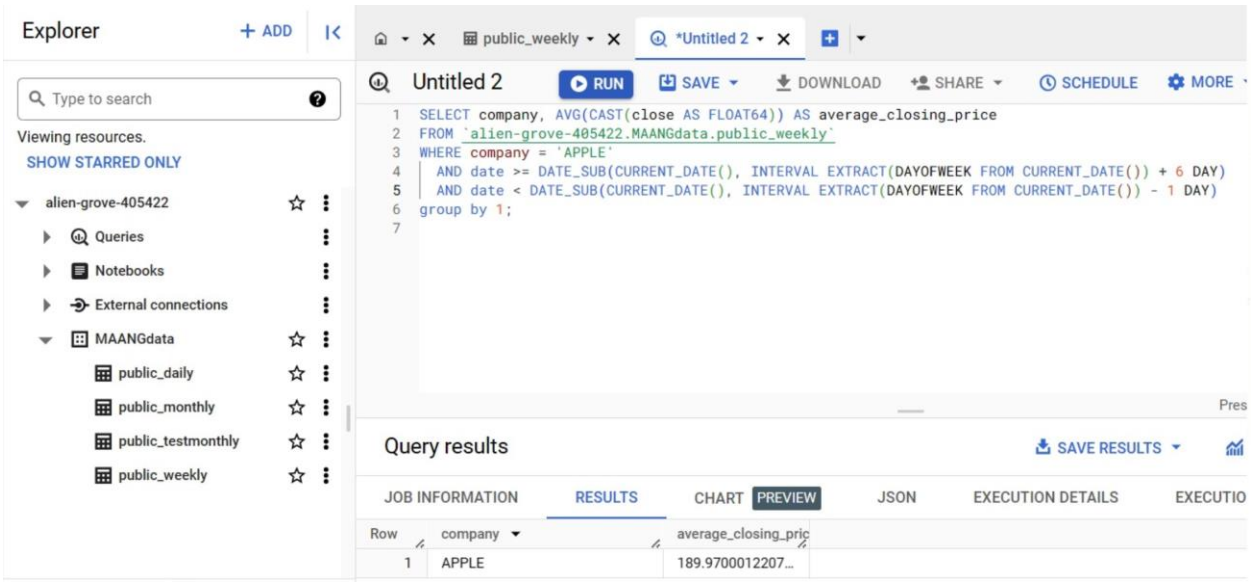

Google Cloud
My First Project
Search (/) for resources, docs, products and more
Search
BigQuery
Explorer
+ ADD
Type to search
Viewing resources.
SHOW STARRED ONLY
alien-grove-405422
Queries
Notebooks
External connections
MAANGdata
public_daily
public_monthly
public_testmonthly
public_weekly
BigQuery Studio
Data transfers
Scheduled queries
Analytics Hub
Dataform
Partner Centre
Assessment
SQL translation
Monitoring
Capacity management
Release notes
SUMMARY
public_weekly
alien-grove-405422.MAANGdata
Last modified
29 Nov 2023, 23:31:57
UTC-5
Data location
us-central1
public_weekly
QUERY
SHARE
COPY
SNAPSHOT
DELETE
EXPORT
REFRESH
SCHEMA
DETAILS
PREVIEW
LINEAGE
DATA PROFILE
DATA QUALITY
NEW
Row
date
open
high
low
close
1 1997-05-12 0.1218750029802322 0.125 0.0854170024394989 0.0864579975605011
2 1997-05-19 0.0880210027098655 0.0885419994592666 0.0656249970197677 0.0750000029802322
3 1997-05-26 0.0755209997296333 0.0822919979691505 0.0729169994592666 0.0750000029802322
4 1997-06-02 0.0755209997296333 0.0854170024394989 0.0687500014901161 0.0828130021691322
5 1997-06-09 0.0828130021691322 0.0854170024394989 0.0765630006790161 0.0791670009493827
6 1997-06-16 0.0802080035209655 0.0802080035209655 0.0747400000691413 0.0763019993901252
7 1997-06-23 0.0770829990506172 0.0770829990506172 0.0739580020308494 0.0744789987802505
8 1997-06-30 0.0755209997296333 0.0958330035209655 0.0739580020308494 0.0955730006098747
9 1997-07-07 0.0916669964790344 0.1286460012197494 0.0916669964790344 0.1145830005407333
10 1997-07-14 0.1161459982395172 0.1247399970889091 0.1057289987802051 0.1078130006790161
11 1997-07-21 0.1088540032505989 0.1166670024394989 0.1031249985098838 0.1114580035209655
12 1997-07-28 0.1114580035209655 0.1252599954605102 0.1109379976987838 0.1208330020308494
13 1997-08-04 0.1187499985098838 0.1205729991197586 0.1052080020308494 0.1145830005407333
14 1997-08-11 0.1145830005407333 0.1166670024394989 0.0968749970197677 0.1057289987802051
15 1997-08-18 0.102604001760482 0.1104170009493827 0.0984380021691322 0.1062500029802322
16 1997-08-25 0.1052080020308494 0.1197919994592661 0.1052080020308494 0.1169269979000091
17 1997-09-01 0.117187999189000 0.1333329975605011 0.1156250014901609 0.125
18 1997-09-08 0.1265629976987838 0.1848960071802139 0.125 0.1843750029802322
Results per page: 50
1 – 50 of 6363
Job history
REFRESH

Monthly:



Analy&cal Queries:

1. Calculate the average closing price for last week for Apple:



2. What is the opening stock price for Meta in the month of October?

```sql
SELECT company, open
FROM `alien-grove-405422.MAANGdata.public_monthly`
WHERE company = 'META'
  AND EXTRACT(MONTH FROM date) = 10
  AND EXTRACT(YEAR FROM date) = EXTRACT(YEAR FROM CURRENT_DATE());
```

Query results

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION

Row	company	open
1	META	302.739990234375

3. Which is the lowest trading price and company in the month of November?

```sql
WITH MonthlyLowestPrices AS (
  SELECT company, MIN(low) AS lowest_trading_price
  FROM `alien-grove-405422.MAANGdata.public_monthly`
  WHERE EXTRACT(MONTH FROM date) = 11
    AND EXTRACT(YEAR FROM date) = EXTRACT(YEAR FROM CURRENT_DATE())
  GROUP BY company
)

SELECT company, lowest_trading_price
FROM MonthlyLowestPrices
ORDER BY lowest_trading_price ASC
LIMIT 1;
```

Query results

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUT

Row	company	lowest_trading_price
1	GOOGLE	124.9250030517578

4. On the 27th of November, what were the top 3 stocks sold and by which company?