Name : Aditya Shethiya

Id:1034791

Email:ashethiy@my.bridgeport.edu

Project phase 4 part\_2

Spark and Big Query

Spark

In spark we calculate the standard devotion for this purpose I used spark sql langue it is avalible package in spark

Their instruction as fallows

spark implemention:

in spark they directly create object of spark as sc sparkcontent and hive as hc

from pyspark.sql import SQLContext

sqlContext = SQLContext(sc)

lines = sc.textFile("/home/training/Desktop/nse-india-stocks-companies/nseComp.1.txt")

parts = lines.map(lambda l: l.split(","))

df.show()

from pyspark.sql import SQLContext

sqlContext = SQLContext(sc)

df = sqlContext.sql("SELECT companyName,Stddev(high),count(high) FROM nsestock group by companyName ")

also we can directly use hive table as input to spark

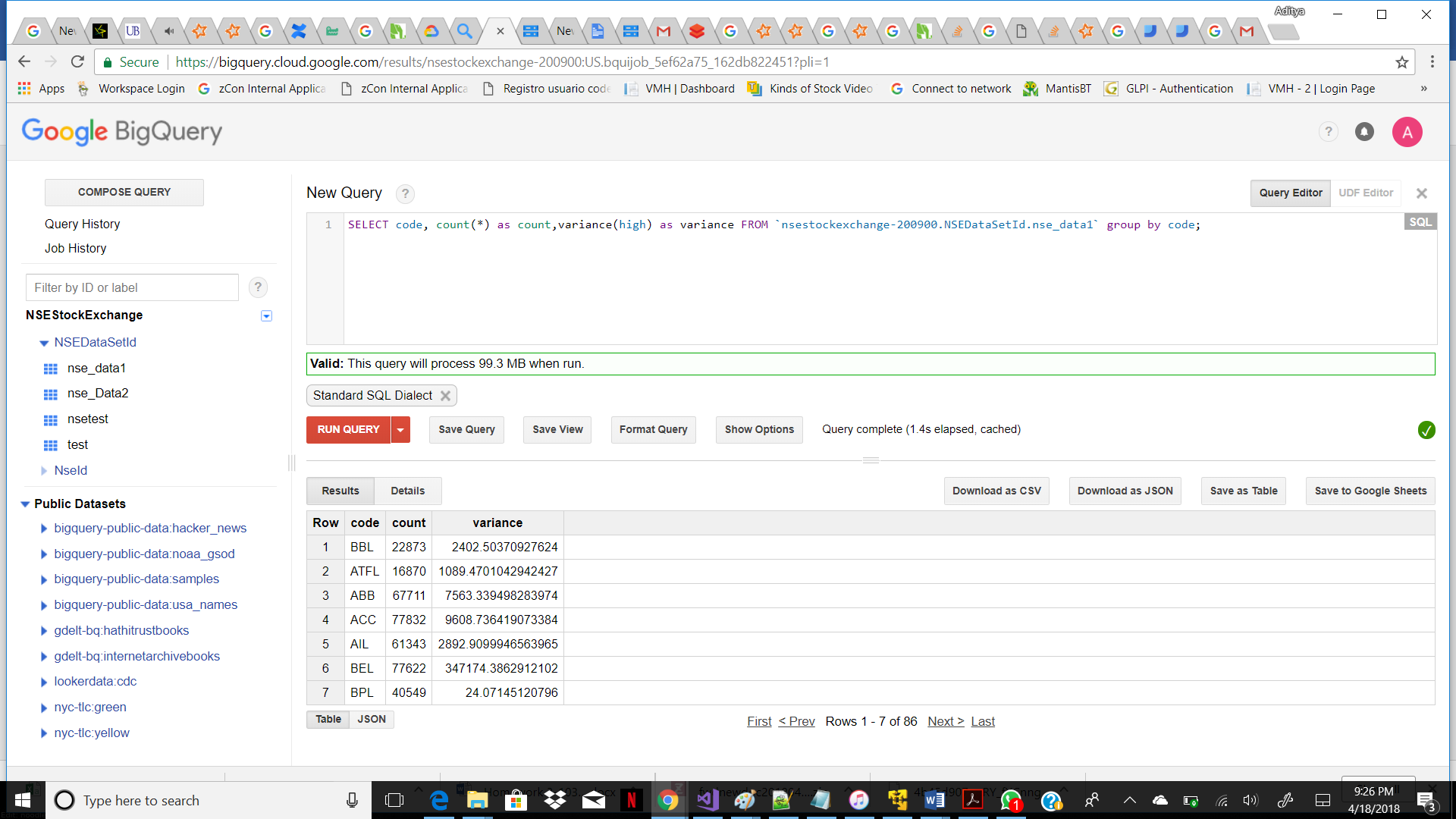
Big Query:

In big query we calculate the variance of the data which we are taking input from spark first we upload our data in google cloud storage

SELECT code, count(\*) as count,variance(high) as variance FROM `nsestockexchange-200900.NSEDataSetId.nse\_data1` group by code;

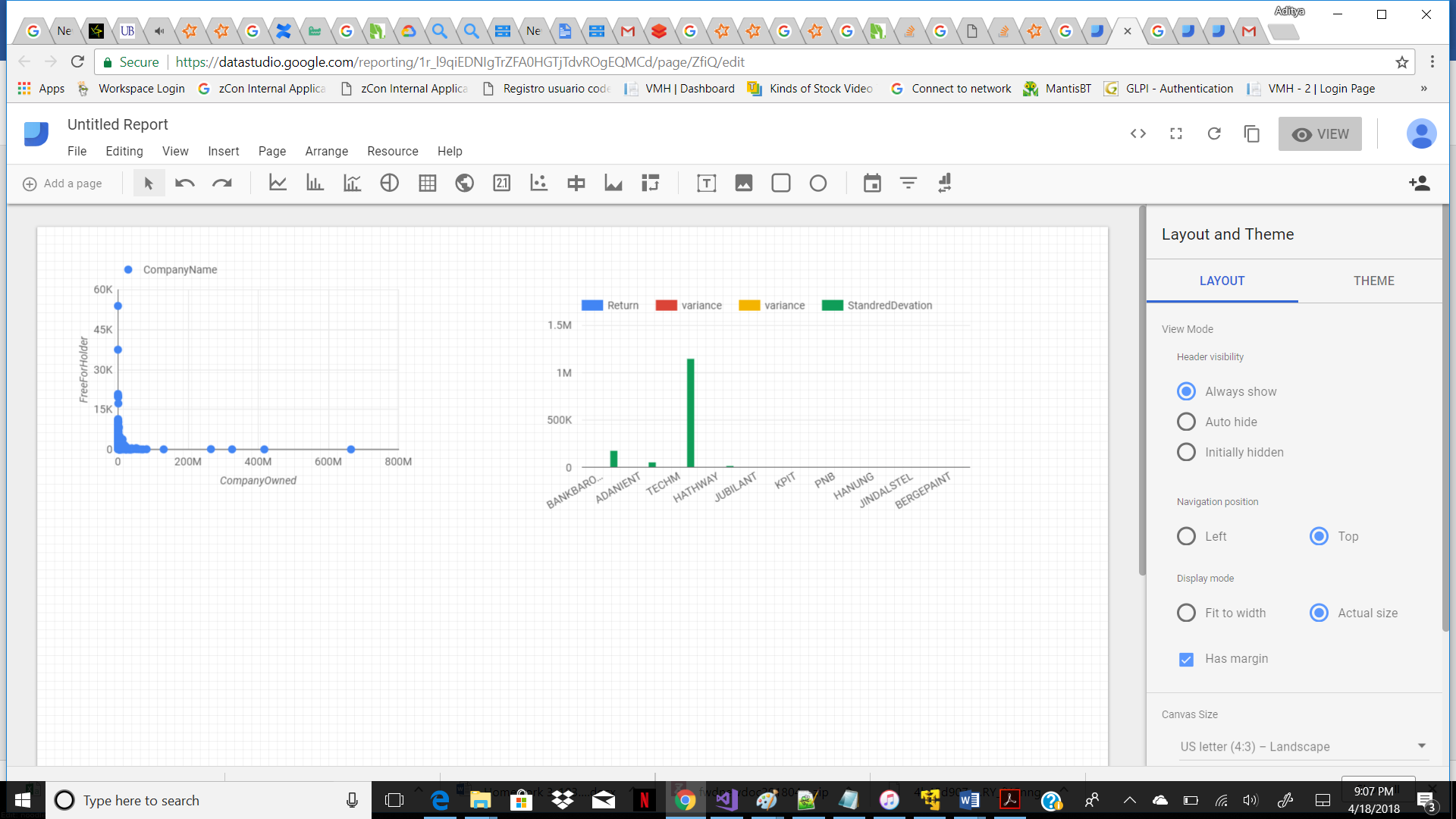
Link for big query:

<https://bigquery.cloud.google.com/results/nsestockexchange-200900:US.bquijob_5711701_162d6971dff?pli=1>



Google Studio:

Then we are giving this output to google studio for visual representation which is given below:



Link: https://datastudio.google.com/reporting/1r\_l9qiEDNIgTrZFA0HGTjTdvROgEQMCd/page/ZfiQ/edit