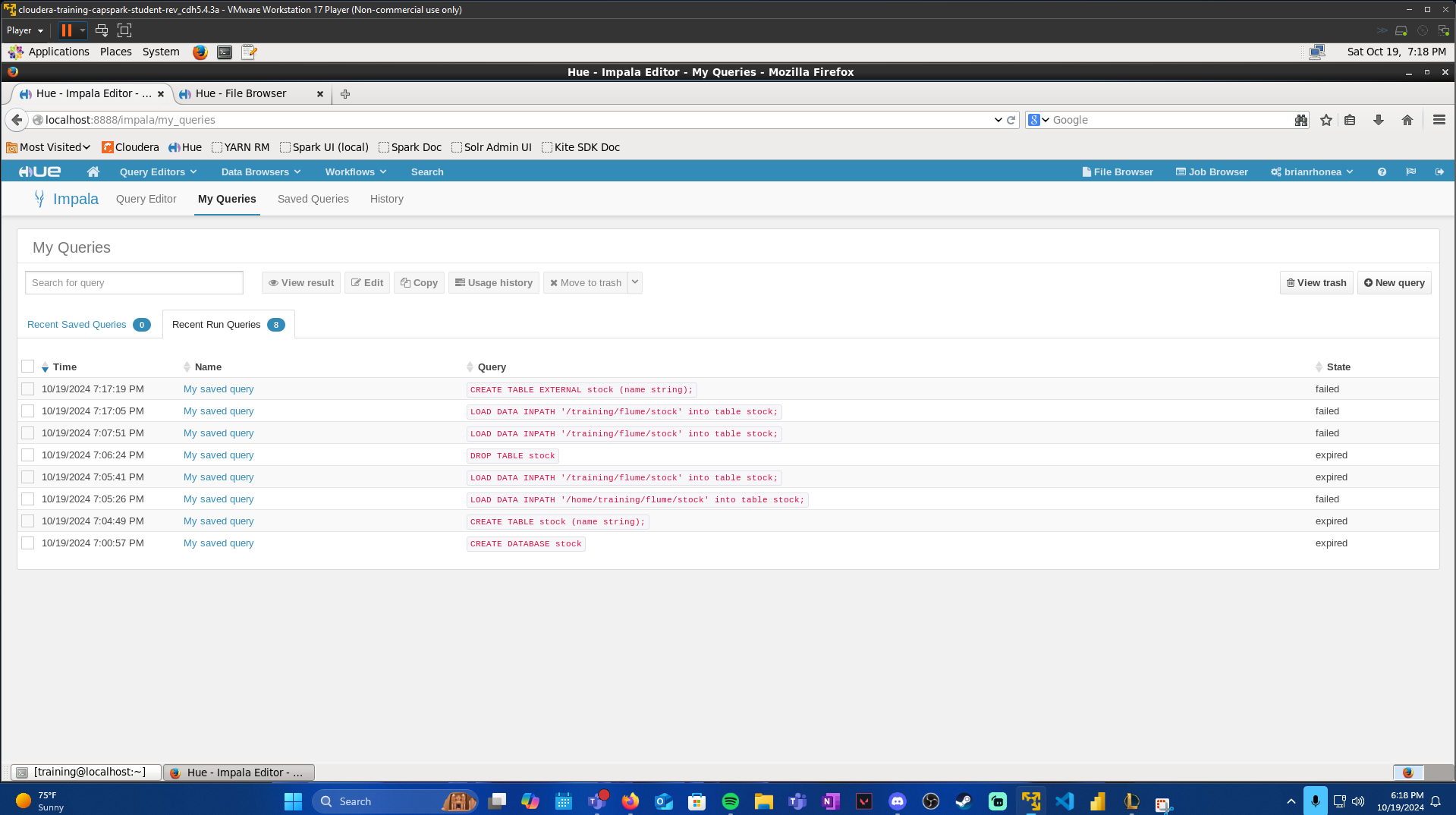
Project 2 Part B:



This was on October 19th. I created a table and imported the data; however, it did not import how I wanted it to. Without thinking about the consequences, I dropped the table and lost all data. I tried to find it in HDFS trash to restore but could not. So, for this reason, I will be gathering and using data from October 20th to October 30th for this project. The dates will not be matching from Project 1. The dates will not be 5 consecutive days because there were days, I was unable to run my script.

**Project 2: Part B**

**Create table in Hive to import data from file. With sample data shown as well.**

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

1. How many records are there in the table?

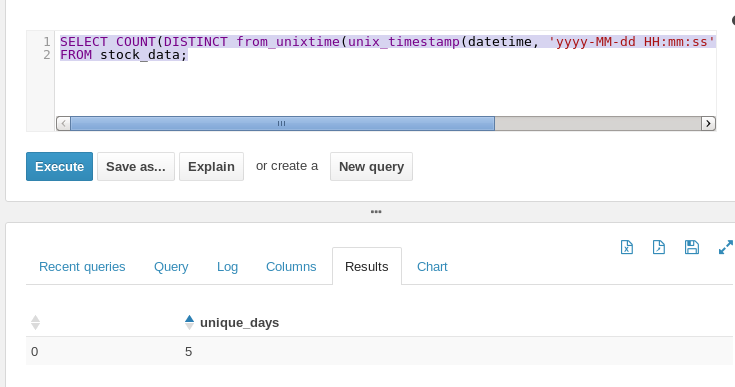
**SELECT COUNT(\*) from stock\_data; There is 3111 records in the table.**

A screenshot of a computer

Description automatically generated

1. How many different days are there in the table?

**SELECT COUNT(DISTINCT from\_unixtime(unix\_timestamp(datetime, 'yyyy-MM-dd HH:mm:ss'), 'yyyy-MM-dd')) AS unique\_days**

**FROM stock\_data; There are 5 unique days.** 

1. How many records per each day are there in the table?

**SELECT from\_unixtime(unix\_timestamp(datetime, 'yyyy-MM-dd HH:mm:ss'), 'yyyy-MM-dd') AS day, COUNT(\*) AS records\_per\_day FROM stock\_data GROUP BY from\_unixtime(unix\_timestamp(datetime, 'yyyy-MM-dd HH:mm:ss'), 'yyyy-MM-dd') ORDER BY day; There is some discrepancy in the days because I had forgotten to run the script some days.**

A screenshot of a computer

Description automatically generated

1. What are the symbols in the table?

**SELECT DISTINCT stock AS unique\_symbols**

**FROM stock\_data;**

A screenshot of a computer

Description automatically generated

1. What is the highest price for each symbol?

**SELECT stock, MAX(CAST(high AS DOUBLE)) AS highest\_price**

**FROM stock\_data GROUP BY stock;**

A screenshot of a computer

Description automatically generated

1. What is the lowest price for each symbol?

**SELECT stock, MIN(CAST(low AS DOUBLE)) AS lowest\_price**

**FROM stock\_data**

**GROUP BY stock;**

A screenshot of a computer

Description automatically generated

1. What is the average price for each symbol?

**SELECT stock, AVG(CAST(close AS DOUBLE)) AS average\_price**

**FROM stock\_data**

**GROUP BY stock;**

A screenshot of a computer

Description automatically generated

1. What is the range of price for each symbol?

**SELECT stock, (MAX(CAST(high AS DOUBLE)) - MIN(CAST(low AS DOUBLE))) AS price\_range**

**FROM stock\_data**

**GROUP BY stock;**

**A screenshot of a computer

Description automatically generated**

1. What is the date on which each symbol experienced the highest price?

**SELECT sd.stock,**

**sd.datetime,**

**CAST(sd.high AS DOUBLE) AS high\_price**

**FROM stock\_data sd**

**JOIN (**

**SELECT stock, MAX(CAST(high AS DOUBLE)) AS max\_high**

**FROM stock\_data**

**GROUP BY stock**

**) max\_price ON sd.stock = max\_price.stock AND CAST(sd.high AS DOUBLE) = max\_price.max\_high;**

