Workflow process

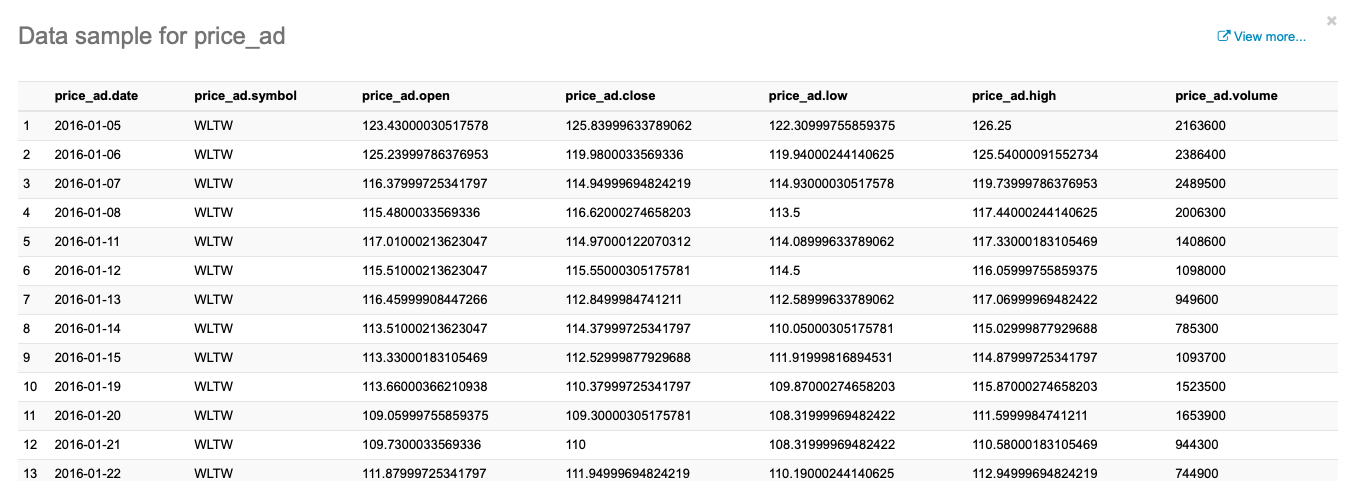
Group 2 - Stock market prediction

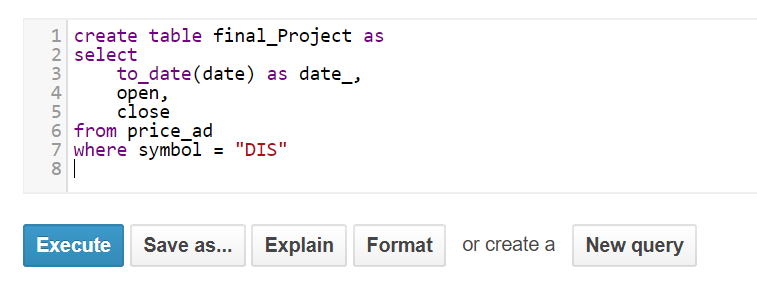
The video for our project can be found here –

<https://youtu.be/v7VBTBaZpc8>

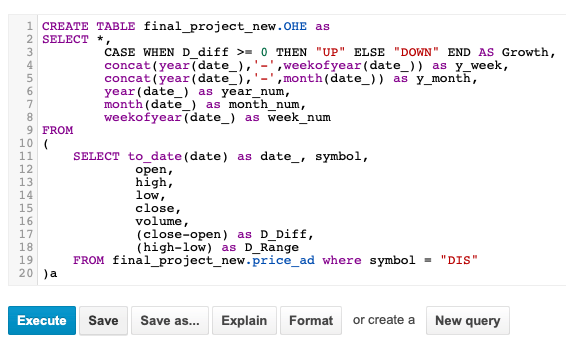
**Data pipeline**

1. Download data from Kaggle NYSE - <https://www.kaggle.com/dgawlik/nyse>
2. Push data to Hive and create a hive table
   1. Create a Hive script to splice data for just Disney stock prices

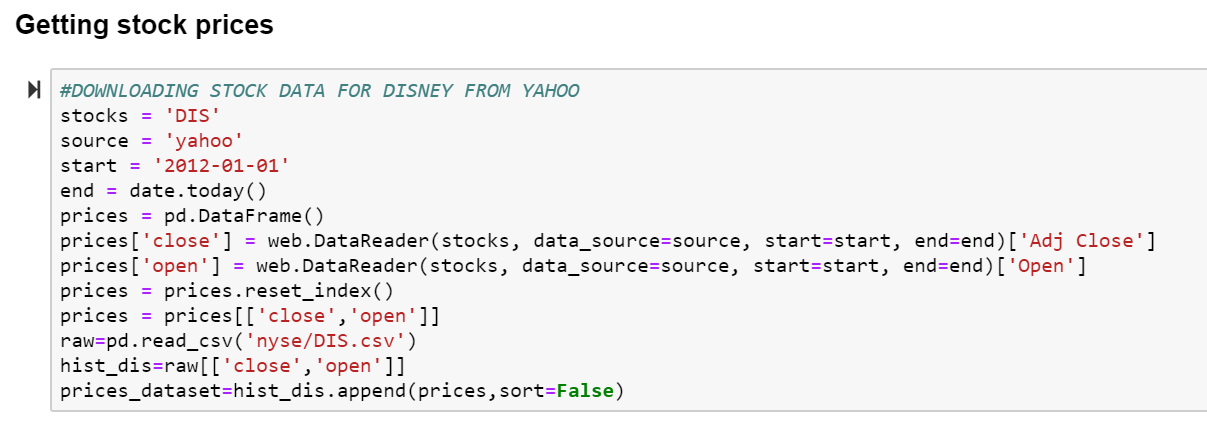




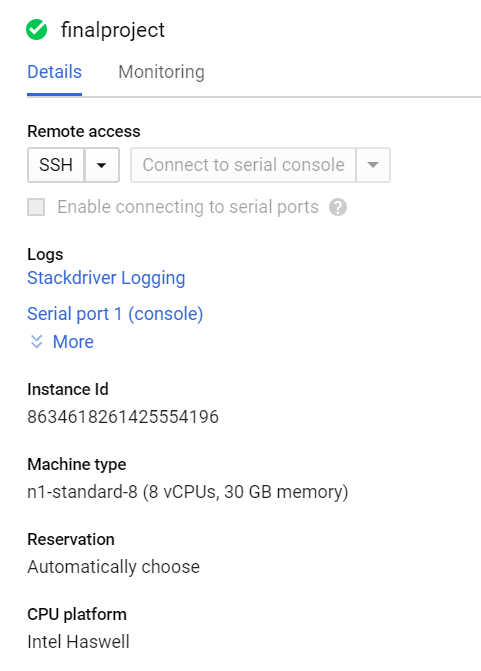
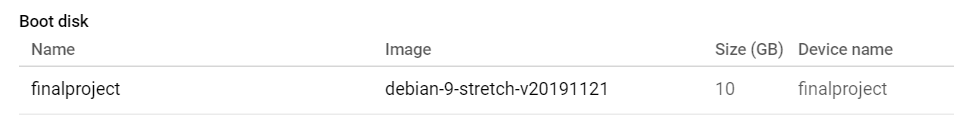
* 1. Push this data to Tableau and download in CSV for modelling



1. Use this data and append recent stock data extracted using Yahoo finance API



1. Run the LSTM model (code filename: group2\_stockprediction\_code.ipynb)and send the predictions for the next 8 days to a google sheet
2. This google sheet then feeds into the Tableau dashboard
3. The modelling python script is hosted in a GCP VM instance running Debian/GNU 9



1. Using crontab, we have scheduled the python script to run every day at 1am
   1. During every run, the script runs LSTM modelling on training data that includes opening and closing prices from January 1, 1962 to the day before and predicts the open and close prices for the next 7 days.

