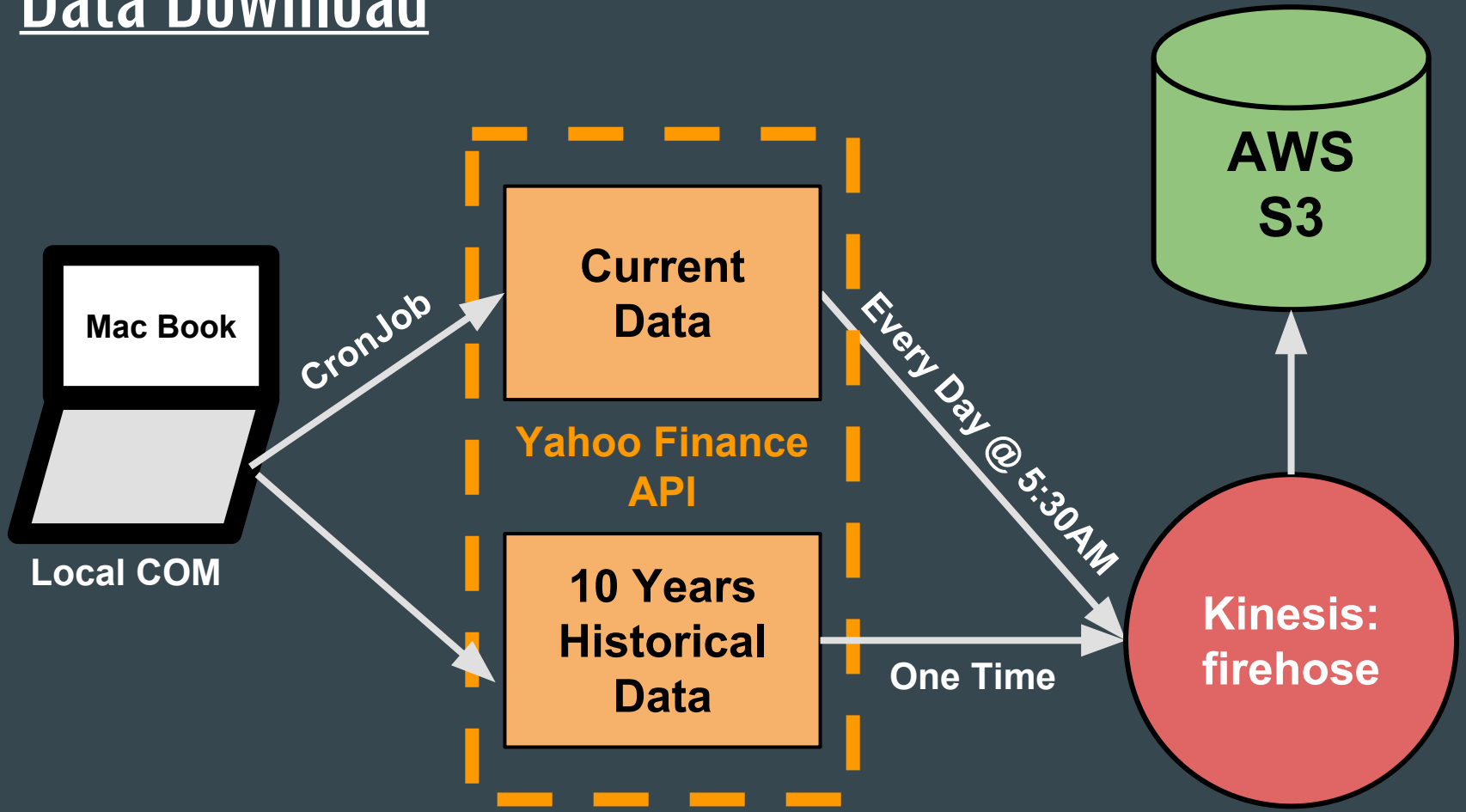
Stock Market Analysis

Yahoo Finance API

By gU Data Science Student:

Yunsup Jung

Data Download



Data Download

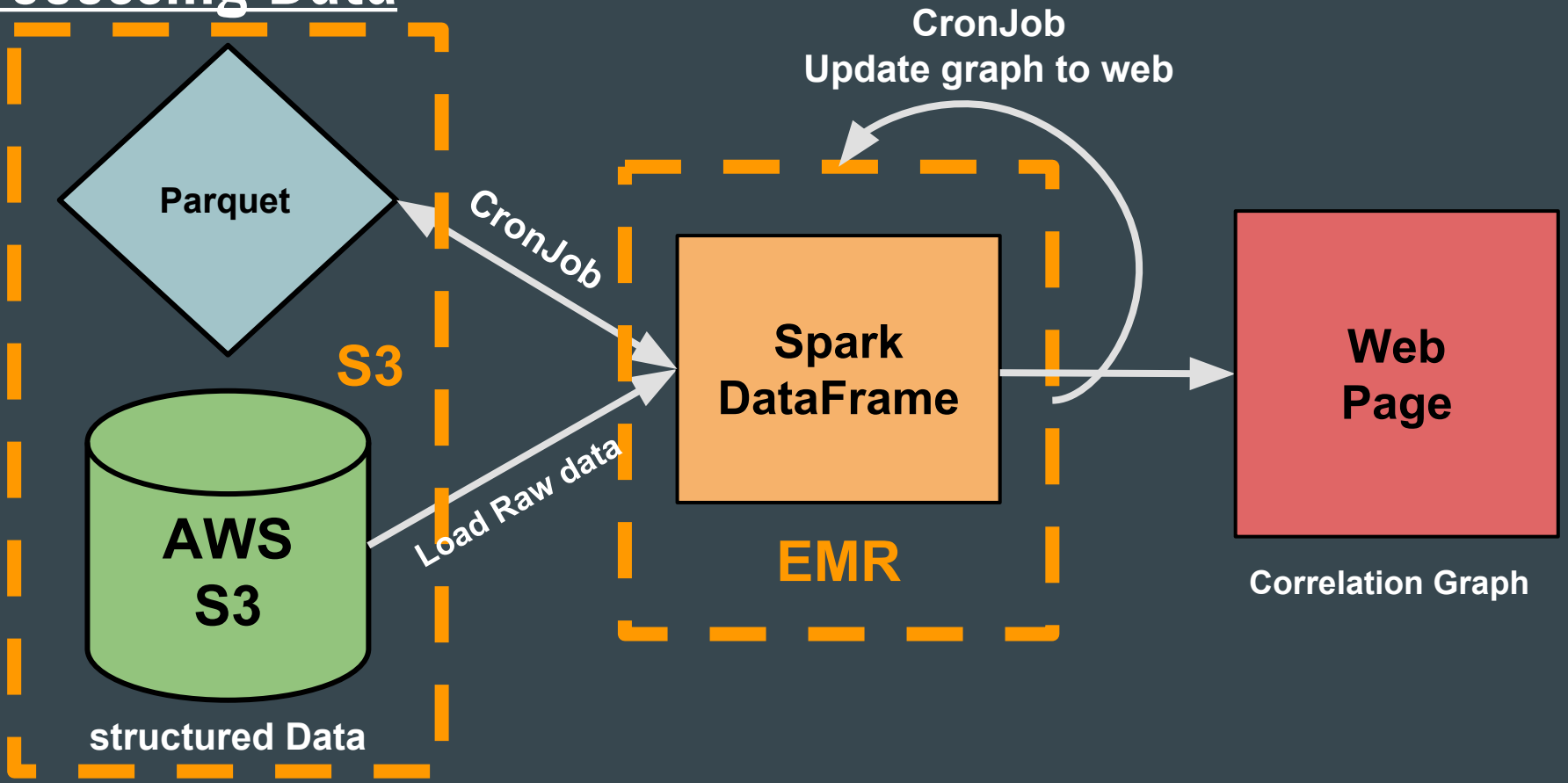
```
Inserted 2489 days Alphabet Inc. data.
[{'Volume': '3749600', 'Symbol': 'googl', 'Adj_Close': '845.030029', 'High': '867.00', 'Lo
w': '841.900024', 'Date': '2017-01-27', 'Close': '845.030029', 'Open': '859.00'}]
Inserted 2490 days Alphabet Inc. data.
It is Holiday or Weekend.
It is Holiday or Weekend.
```

- **Download 10 years data**
  - **python Local_yahoo_S3.py AMZN googl 3650**

- **Download current data**
  - **nohup python Local_yahoo_S3.py AMZN googl 0 &**

```
if days_before == 0:
    #schedule to update data every day at 5:30 PST
    schedule.every(1).day.at("05:30").do(yahoo_current, symbol1, symbol2)
    yahoo_current(symbol1, symbol2)
    #schedule.every(1).minutes.do(yahoo_current, symbol1, symbol2)
    while True:
        schedule.run_pending()
```

Processing Data



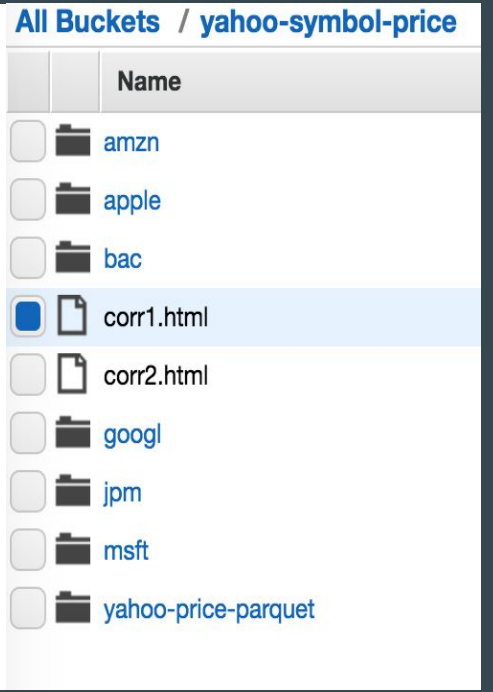
Processing Data

```
msft = yahoo_price_df.selectExpr("price", "Date AS date").where(yahoo_price_df.Symbol == 'msft')
msft.write.mode('overwrite').parquet("s3a://yahoo-symbol-price/msft")
```

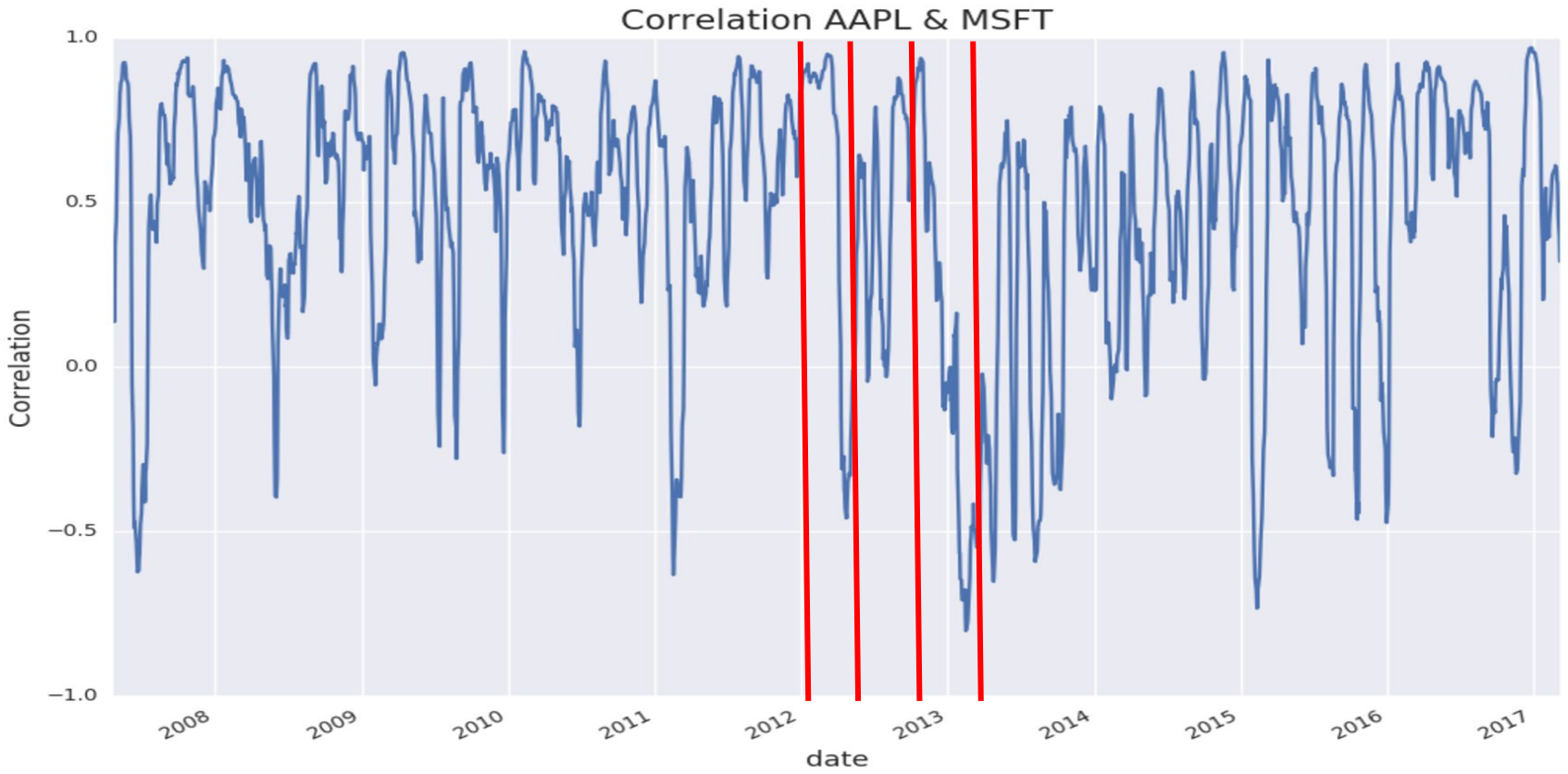
```
[hadoop@ip-172-31-53-12 ~]$ crontab -l
0 6 * * * python EMR_update_current.py
```

- **Update data everyday at 6AM, and store into S3 with partition each symbol.**

- **On EMR, make a data frame with selected symbols, then draw correlation between two symbols.**
  Ex) draw correlation graph aapl & msft



Result


Correlation AAPL & MSFT
1.0
0.5
0.0
−0.5
−1.0
Correlation
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
date

Conclusion

- **With this tool and architecture, people can easily get their favorite stock data include price and volume.**

- **Also, it can be use stock recommendation tool with time series analysis (ARIMA predict - the range will be too wide).**

- **With correlation graph, it is possible to classify type of business. Also, it can realize seasonal rival relationship between two companies.**
  ex) aapl and msft

Next

- **I tried to predict the future price range for each symbols with ARIMA time series model, but it didn’t work well.**

- **I suggested to use spark-ts (time series for spark) package or use toPandas() commend in order to apply ARIMA with pandas.**