Analysis of NYSE Data for Stock Trends

Bhavesh Reddy Merugurreddy[1,*] **and Niteesh Kumar Akurati**[1,**]

[1]*School of Informatics and Computing, Bloomington, IN 47408, U.S.A.*
Corresponding authors:bmerugur@iu.edu
**Corresponding authors: akuratin@iu.edu*

A detailed analysis of the New York Stock Exchange(NYSE) data by deploying Hadoop and Spark technologies on multiple clouds in a distributed fashion[1]. The analysis results are used to observe stock trends. The results include highest change and lowest change in stock prices for a given day and percent increase or decrease in the stock prices which might be used by investors to make smart decisions and observe the movement of a particular stock in the market over a period of time[2].

Keywords: Hadoop, Spark, Ansible, OHLC, Python

https://github.com/cloudmesh/classes/blob/master/project/S17-IR-P014/report/report.pdf

1. INTRODUCTION

Newyork Stock Exchange(NYSE) is an American Stock Exchange. NYSE records lots of data regarding trading and status of all the stock information. OHLC(Open High Low Close) data is one of the popular form of data in the stock market[3]. Creating charts using the OHLC data available from NYSE we intend to analyze stock trends and correlations among the attributes like volume and price trends of various stocks. The analysis being done as part of the project will help us to look into the insights of the stock market and understand the dynamic patterns of stocks and the trends of success and failure of stocks[4].

2. MILESTONES

- Performing Analysis on local VM

- Deploying Spark and Hadoop on FutureSystems or Chameleon Cloud

- Analysis on the distributed cloud environment

- Deploying Spark and Hadoop on one more cloud(Chameleon or FutureSystems)

- Benchmarking

- Final update with report

3. TECHNOLOGIES

- Distributed Computation and Storage:- Spark and HDFS

- Development:- Python and Java

- Deployment:- Ansible

4. DEPLOYMENT

Ansible Playbook is used as the application and configuration deployment tool. Deploying the hadoop and spark framework into the cluster environment. Ansible will help push configurations to the environment automatically based on playbooks written for various configurations

5. BENCHMARKING

TBD

REFERENCES

[1] D. Gadiraju, "nyse," Github repository. [Online]. Available: https://github.com/dgadiraju/data/tree/master/nyse

[2] johnsinit, "Open high low close (ohlc) chart type," Webpage. [Online]. Available: http://www.brighthub.com/money/investing/articles/33137.aspx

[3] "New york stock exchange," Webpage. [Online]. Available: https://en.wikipedia.org/wiki/New_York_Stock_Exchange

[4] C. Miller, "An introduction to stock market data analysis with python," Webpage. [Online]. Available: https://ntguardian.wordpress.com/2016/09/19/introduction-stock-market-data-python-1/