Master Thesis Proposal

Ville Vainio

February 19, 2019

1 Context

The modern economy revolves around stock market. Stock market is way for companies to obtain capital which they can invest into their own business. In exchange, person who invests into the companies stocks technically owns a piece of the company. The investor can make profit by selling these stock in a higher price or by receiving dividends from the company itself.

The price of the stock is simply determined by the law of supply and demand. If somebody is willing to pay a higher price for the stock then the price of the stock can grow. Because of this the stock market is in continuous fluctuation where people are selling and buying the stocks with the price they think the stock is worth using stockbrokers as the middleman. [1]

There are many strategies on how to invest into these stocks which depend on multiple factors such as; how much do you expect to profit with your investment, how much are you willing to take risk, do you want to make money by selling the stocks or by receiving dividends and so on. The underlying principle with every strategy is to minimize the risk you need to take in order to gain as much as profit as possible. Some of the strategies are based on subjective evaluation of the companies, but more technical strategies use metrics that are calculated from the financial statistics or the real-time market values. Strategies using the former data are called fundamental analysis and the strategies using latter data technical analysis. Neither of these approaches can predict the future of the market, but can statistically decrease the probability of larger losses in the market for the investor altough the probability of large losses is still not zero with these methods. [2]

Fundamental analysis is based on the idea that each stock has a intrinsic value that can be larger than the actual price of the stock in the market and buying these will eventually lead to profits.[3] The fundamental analysis focuses on the financial metrics that consist of companys overall statistics. These are for example how much the company has made profit, how much the company has paid dividends and what is companys cash flow. These tell a lot about the growth of the company and how the future of the company looks like. These metrics are usually published quarterly four times a year and present more long-term statistics about the company. Because of this, the amount of data these values present is quite small in terms of space.

The technical analysis that focuses on the real-time market values, on the other hand, needs new data almost daily. Stock exchanges are usually open from morning, opening around 8 to 10am, until evening, closing around 5 to 7pm on weekdays. Before and after this there are more limited pre- and after-hours trading which lasts usually around 1 to 2 hours depending on the exchange in which more limited stock trades can be made. During these hours multiple values are recorded on the prices of the stock from which the most important ones being: the highest price the stock was sold, the highest price the stock was sold and the number of stocks traded during the time interval.

The technical analysis focuses on finding recognizable patterns through this data. [4] Where the data used by the fundamental analysis was relatively small, these values can generate gigabytes of raw data in a week.

2 Research Questions

Orchestrating data fetching for both of these analyses is common task on every system that provides automatic calculation on the metrics that the analyses use to calculate predictions. When stock exchanges consist of thousands companies, which is the case e.g with the american stock exchanges, the total amount of data grows enormously. Because of this, the process of ingesting the data becomes harder to implement efficiently so that the user does not notice notable delay in the calculated values. Other problems arise when this kind of system is scaled to ingest market data from other than just for example markets in US.

This thesis would be focusing on the part of getting this data efficiently and safely to the parts of the software that calculate these user metrics from the data. This kind of system should be scalable and stable. The main questions this thesis would try to answer would be: What are the state of art techologies to implement this kind of system efficiently, how to implement this kind of system in practice and how does this compare with an existing solution.

3 Expected Outcome

The expected outcome of this thesis would be an efficient and stable system that fetches and stores stock market data from an open source endpoint. State of art technologies to implement this would probably be Apache Spark and HDFS database clusters [5] which would hopefully lead to a significant performance and stability gain for the system. The expected performance would be the best possible performance so that the bottleneck of the system would mainly be on the open source stock data provider.

4 Approach

The thesis would do a literary research on the different big data solutions. The thesis would briefly analyze what are the bottlenecks in the existing prototype system and introduce technologies and methods that are available that can make a cloud system more scalable and efficient concerning these bottlenecks. These include methods and technologies for example storing and fetching data, which is seemingly one of the major possible bottleneck in this kind of system.

Based on this research, the thesis would try to implement the seemingly best solution in practice using a existing prototype system as a point of comparison. After this the thesis would conclude with an analysis on how did the system improve compared to the prototype system performance-wise, what are possible downsides of this developed system and what could possibly be implemented as further improvements. Perfomance would be measured on metrics such as the throughput of the system, the time system takes to complete one iteration of fetching and the latency of the system.

References

[1] John L. Person. *Mastering the Stock Market: High Probability Market Timing and Stock Selection Tools.* John Wiley & Sons, Incorporated, 2013.

[2] Justin Fox. *Myth of the Rational Market.* Harper Business, 2009.

[3] Söhnke M. Bartram and Mark Grinblatt. "Agnostic Fundamental Analysis Works". In: *Journal of Financial Economics (JRE)* (June 2017).

[4] John J. Murphy. *Technical analysis financial markets: A comprehensive guide to trading methods and applications.* New York Institute of Finance, 1999.

[5] Mohiuddin Ahmed, Abu Sina Md. Raju Chowdhury, Mustaq Ahmed, Md. Mahmudul Hasan Rafee. "An Advanced Survey on Cloud Computing and State-of-the-art Research Issues". In: *IJCSI International Journal of Computer Science Issues* 9.1 (2012).