ABSTRACT

Finance was among one of the earliest domains to deal with Big Data. Stock market traders work in "millisecond environment" where financial decisions need to be made in milliseconds to be able to deal with the innate volatility of the stock market. With the ever-increasing velocity, frequency and volume of big data in finance due to rapidly rising use of algorithmic trading, a specialised low-latency computing architecture is needed to be able to store, process and extract information from this data with minimal decision latency. Currently, stock traders make use of physical, local computing systems and geographical closeness to the trading venue to reduce the latency of their financial decision-making.

Big Data technologies are used to provide distributed computing and parallel processing capabilities. They provide a way to process huge quantities of data efficiently, increase responsiveness to the changes in the data and are scalable.

Through this research, I aimed to address the gap between real-time data processing and market risk estimation of financial high frequency big data and big data architecture. I evaluated the impact on latency for market risk estimation of 3 high frequency stock market datasets. I used statistical methods with financial market algorithms in conjunction with Big Data architecture - Spark and Hadoop, for this study. The latency study showed that Hadoop Map-Reduce and Spark both outperformed the baseline system in decision latency and resources utilization by 40.6% and 20% respectively while Hadoop outperformed Spark in each of these aspects on all datasets. The market study showed that over a period, equity firms start behaving like the overall market.