Financial Risk Management with Apache Spark

Requirements Document

CS Capstone Project
Saint Louis University
October 1, 2017

Kamran Madatov
Bao Thai
John Mitton

Under Guidance from:
Dr. Goldwasser & Dr. Chambers

Mentor:
Dr. Ted

I. Project Description

Value at Risk (VaR) has been widely adopted in the financial industry to measure risk. It is used for regulatory compliance, understanding the risk characteristics of large portfolios, and and making informed trading decisions. Three common methods of calculating Value at Risk are variance-covariance, historical simulation, and Monte Carlo simulation. Monte Carlo simulation can be more accurate than the simple models, but it requires more computational power. Fortunately, Apache Spark provides an easy way to scale statistical problems beyond what a single server can handle. Using Spark and historical stock data, we will calculate VaR of stocks with Monte Carlo Simulation in less time.

II. Project Objective

The objective of this project is to help investors make appropriate decisions with their portfolio to reduce loss when holding, buying, and selling investments. A mobile application will be created to track listed investments and inform the user based on news, history of stock prices, trend volume, social media, and market stock exchange. By using statistical data and probability calculation, we would like to ensure accuracy of estimates and allow users to be more informed on changes with a stock.

III. Usability Requirements

Case	Actor	Action	Priority
1	Server	Calculate the VaR for a particular stock from historical data	5
2	Server	Must have a Monte Carlo model visualization of individual stocks in portfolio	5
3	Server	Calculate probability and possible percent change of a stock through model	5
4	Server	Sends a notification to alert user of severe changes	4
5	User	User can log into portal and view portfolio	5
6	User	Can view particular articles regarding changes and calculation of VaR for certain stocks	3
7	Server	Possibly inform of new stocks with high chance of growth	1
8	Analysts	Can manually perform and input measures	1

IV. **Target**

With a variable at risk tool, we can use this to analyze performance of any type of investment. We are potential targeting hedge funds and investment firms such as Wells Fargo, Charles Schwab, Goldman Sachs, and etc. If the product will be commercialized, it will most likely be integrated into big corporation to allow easier management of risk among their large clients' database. This could reduce operation costs and allow for more productivity among business analytics.

This can also be a normal commercialized licensed software, where users can look up particular stocks to gain insight without having to manually do any calculations for Monte Carlo. And also will do all the background work of gaining insight on particular stocks through using the big data that our tool is connected to, so the software will be much easier to distribute among individuals.

V. **Cost & Reliability**

To maintain the cost of high data computation using Sparke, as well as using Elastic mapping to store our big data from the web crawler, we are looking at a monthly budget of $300 per month, according to the cost that was charged on the team's AWS account for leaving Elastic and Apache Spark running for the month of September. However, if we have multiple users at a flat rate price, we can make up that operational cost. And after that, we can look at cheaper alternatives to handle more requests as we expand and scaling into a higher capacity of information processing.

Depending on how large of the market exchange we want to look at, our storage capacity can be the most costly thing in our operations, which will greatly add to the additional $300 per month cost. The more data we want to store, the higher storage capacity we will need, which we will possibly look into expanding Elastic data or using Glacier to only read data per day.

There will be a option of purchasing individual and organization accounts. The price for all individuals will be a flat rate (which will be determined when the product is ready to be deployed to production), the cost for the organization will depend on the size of members in the organization and amount traffic that will be utilized. In organizational situations, various options will be available and the price will be negotiated.

VI. **Maintainability**

In an ideal scenario, we will always have a development operations team to continuously ensure the tool is working smoothly for clients. These engineers will also resolve any issues and tickets being brought up among clients and users. We want to continuously improve the program as well, for possibly more in-depth research or additional features. This will allow for better estimation within the program to ensure higher profit return for our clients.

We would also like to have a business analyst to ensure the algorithms and Monte Carlo modeling are correct and ensure there are no flaws as the software is in use. This will prevent any miscalculation or technical error that can result in damage for a corporation's revenue or reputation, or even our own tool's reputation.

Moving forward, the application will still take agile requirements and requested updates to improve the accuracy of the application and objective. R&D team will be created to research potential future solutions to compete with other competitors and stock market.