Map Reduce for Algorithmic Trading

Akshaan Kakar, Alex Berard
Dept. of Computer Science
Columbia University
Email: ak3808@columbia.edu, alice.berard@columbia.edu

***Abstract*—Algorithmic Trading is an extremely competitive sector of financial markets. Developing trading algorithms involves the pivotal step of backtesting, where the performance of the algorithm is validated against large amounts of historical securities pricing data. These time series require large amounts of computational resources for storage, processing and visualization. In this paper, we describe the implementation of an algorithmic trading backtest engine that uses ?Big Data? tools as a backbone to backtests for trading algorithms in an efficient and scalable fashion. We also show how this system can be extended to support live trading as well as a plethora of other features.**

***Keywords*-Algorithms, Finance, Trading, Big Data, Hadoop, Spark, Backtesting**

I. Introduction

With the computerization of order flows in financial markets, traders were afforded the ability to have computers place buy and sell orders on securities according to predefined strategies. The advent of such Algorithmic Trading strategies eventually developed into an extremely competitive sector of financial markets. The development of trading algorithms is now a formalized process that involves intricate research and mathematical models. Although a significant portion of this development effort is invested in the underlying mathematical theory, backtesting of algorithms is the most pivotal step in the process. Backtesting is the phase in which a trading algorithm is validated against large amounts of historical pricing data. Since trading algorithms may place orders extremely frequently and involve multiple securities in a single portfolio, large amounts of data must be stored, read and processed in order to run these tests. Moreover, the result must be visualized so that developers can gauge the performance of their algorithms quickly and conveniently.

There is an abundance of well developed and well maintained tools which have been built expressly to grapple with the large amounts of data that have now become commonplace. These tools leverage novel algorithmic and system level techniques to to deal with the space and time bottlenecks that come with large datasets. Since the problem of backtesting algorithmic strategies is inherently a 'Big Data' shaped problem, we decided to use a combination of such tools in order to to build an efficient and flexible system for the task.

II. Related Works

Since the advent of algorithmic trading, a variety of trading platforms have been developed. These platforms offer a wide variety of backtesting capabilities

III. System Overview

the main objectives in our implementation were to keep the platform dataset agnostic and easily extendible. In order to achieve these goals, we partitioned the system into independently functioning subsystems which we then integrated to realize the final backtesting engine. These systems are described in detail in the following subsections.

A. Hadoop Data Warehouse

B. Spark Algorithm Processing Engine

C. Javascript Server Back End

D. Visualization

IV. Algorithm (you can change this section name)

Show algorithm and tools you have used to solve your problem.

V. Software Package Description

Describe the software package that is going to be open sourced. Show some screen shots of how user use it and or UI.

VI. Experiment Results

Describe the experiment results of your algorithm. Show how did you evaluate the performance of your algorithm.

VII. Conclusion

Drew a conclusion of your project, describe the contributions of each team member in percentage and discuss future works.

Acknowledgment

The authors would like to thank... more thanks here

APPENDIX

REFERENCES

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.