Scott Clifford - kurifodo@u - 0425963
Final Proposal - Part I

Overview

Analyzing the stock market trends and individual stock trends can be a beast of a task. With thousands of stocks trading every week day and six hours a day, it would be difficult to compute immediate results given the plethora of data produced by companies on the stock market. This would be where the distributed system aspect of solving the problem would come in.

Some of the objectives of analyzing stock market data might be to produce trends, averages, moving averages, or predict the direction of a stock or index (though this might be reaching too far for four weeks =D). The scope of data could be limited to a certain select number of indexes such as the Nasdaq or S&P 500 to provide a base focus of the objectives.

Suggested Solution and System Architecture

The first essential feature of the system is the stock market data. Whether this is publicly available in a nice plain text format, I do not know. However, if need be, it can be parsed off of news sites like Yahoo! or Google. Thus, one component of the project might be simply crawling the web for the data and formatting it.

After acquiring the stock market data, another phase might be to compute statistical measures from the data and outputting each of them in a formatted way for use of analysis later. For instance, the output of a MapReduce phase on Hadoop could contain files which contain:

Stock Name -> Time Period ; Average ; Moving Average ; etc

After computing these statistics, graphs and charts could be generated with a basic reporting application while more data is collected and analyzed (since the stock market is always trading). For this project, one snapshot might suffice though since this would likely "hog" the resources of the Hadoop cluster.

Development Plan

Data Collector: Have this completed within the first week (if the data is not readily available already).

Algorithms and Statistical Measures: Code these within the next week after the data collector. The algorithms might touch on the prediction portion of the objectives mentioned above.

Application/Graphical Interface for displaying results: With the remaining time, a UI might be developed to display the results of manipulating the data.

Feasibility Rationale

The greatest risk is not being able to come across available stock data and then not being able to parse it easily from the internet. This would be the show stopper if anything for this project since the rest seems doable after having the data on-hand.