Market Analysis: A Big Data Solution

Mukkul I Jayhne

Computer Science, Galgotias University, Uttar Pradesh, India

Abstract: In this paper we present the Bigdata infrastructure for handling the large amounts of data processed in the Invest-ment industry. Various bigdata tools can be used for this purpose. But in this case, we are going with Hadoop. Hadoop is frame-work which is developed using Java programming language. It is a framework which uses various concepts of parallel and distributed computing to make the computational speed faster. This causes the programs to execute at a much larger speed with the help of few normal speed computers. This increases the affordability rate and makes it much more efficient. It uses its own file system called as Hadoop distributed file system, that is HDFS. HDFS is known for its security and high risk control. Since it runs in cluster there becomes absolutely no use of a Super computer to process data faster. Hadoop is the widely used big data processing engine with a simple master slave setup. One of the most common place where bigdata is most commonly uses is the share market industry. There are various reasons why bigdata is used in this field. The most common one being to increase the profits by understanding the pervious data. The analytics and understanding of data can only be possible if the large amounts of data is handled in a proper way. Suggesting the shares to users is one of the main concept of this paper. But rather than focusing on the analytical part of the framework our main aim is to make it easier for the admin to use bigdata so that the large amounts of data sets can be easier to process. This application can have a lot of advantages in the algorithmic trading. Algorithmic trading is a type of trading where different algorithms are used for buying and selling the shares. It can also have various use cases in stock brokering firms for processing large amount of data quickly. The main contribution of this paper is to integrate the cloud computing and Hadoop framework. The cloud computing in this project is a web based application which is directly connected to the Hadoop system. The parallel pipeline is developed for the purpose of easy handling of data by the admin. We also have developed a communication protocol upon TCP/IP for the purpose of the pipeline. The share market industry involves are huge amount of data and exabytes of data is processed every minute for various purposes. The investors usually go through all the data that is involved in their research purposes and try to analyse it. There are various factors and also attributes that the investors try to take into account when going through the data. Analysing data involves a large amount of data and very well built platform for it to support the data. In this paper, we have Identify applicable funding agency here. If none, delete this also built a web based platform which helps user analyse the data in a much simpler and graphical notation. The attributes include financial transaction patterns by the investor, market conditions and sentiments, macroeconomics variables, scheme level features, and demographic factors. Predicting the redemption behavior requires sophisticated platform that can capture multiple factors that affect the redemption behavior. However, these big data infrastructure provide us with various use cases, with tools like Hadoop and spark it becomes even more easier to find use cases at macro levels. This platform can investigate these factors for near real-time data and can provide highly accurate predictions for the redeeming investors in the future at a investor-level. Our results show that by implementing cloud computing, bigdata analytics and sofisticated algorithmic trading the results which are data driven can be used to generate a resonable amounts of profits and also the data could be processed in a much simple and faster way.

Keywords: Big data, Hadoop framework, BSE, NSE, FMCG

I. INTRODUCTION

In this current framework the examination patter shows up on most of the stock market exchange for the most part they centred in the market methodologies and antici-pate singular organization benefit by utilizing Algorithmic exchanging system. It is the way toward programming PCs to put electronic exchanges as indicated by predefined systems. Algorithm exchanging handles high-volume huge information, and furthermore high-speed information preparing. In the present markets, a stock can encounter 500 statement changes and 150 exchanges one microsecond. Thusly, costs vary by the milli-or even miniaturized scale second. Putting in an extensive number of requests at high speeds in view of customized systems is essential revenue driven age. Specifically, the rate of market information access and information driven method-ology preparing nearly decides algo-exchanging improvement as such, more effective huge information foundation can make productive openings. The execution of information handling frameworks in light of framework setup, such CPU, memory, system and capacity. Headways away have slacked because of restrictions displayed by dormancy and throughput. Present day strategies, for example, in-memory databases, which de-pend on primary memory for an information store medium, are quicker than circle enhanced database frameworks, however are as yet constrained by the present memory ability. What's more, in-memory database still do not have a non-unstable

capacity medium to give long haul constant stockpiling. To deal with QF huge information, SSD upheld capacity could be more proficient than HDD used capacity.

II.   PROPOSED SYSTEM

A. Problem solution

In our proposed framework we are going to take two unique sorts of Indian market, for example, National Stock Exchange (NSE) and Bombay Stock Exchange (BSE) inside the market having different sort of part, for example, Banking, Automobiles, Energy and Power, Software and Fast Moving Consumer Good (FMCG).where as every division having numerous organizations, for example,A few of top companies is the banking sectors are Axis Bank, Punjab National Bank, Bank of India, etc. Now we can process this information by directly creating a pipeline with the user interface. This helps the admin to directly send the data to the Hadoop Distributed File System. And once the data is processed it returns the required details that the user is looking for.

III.   SYSTEM DESIGN

We will analyze three years of stock market points of interest of rundown over the organizations, our framework will centre around market, division and sectors independently, we discover which showcase performs well in most recent three years by utilizing the datasets of five companies in each and every sector. On the off chance that performs the operation resulting in the outcome demonstrating NSE showcase after that framework discover which is the best performing division in NSE and then examine all the different companies under the market to find the most profiting company. This is the total operation examination in the offer market dataset base paper idea finished at this level. We are not doing just examination work and handling large amounts of datasets, our application for the most part centred around giving suggestions to the user about the investments he can make in the stock market to obtain the best profits according to the type of risk the user can take. Therefore, we have to look at every individual segment and companies for a long time at long last we go out on a limb factor for every division and individual company. This hazard factor is only the normal mean estimation of every last year. That is the risk factor (Hazard factor) is nothing but the average of each companies for every year. Now that since we have datasets for three years. We will calculate the risk factor of all the companies by analysing the average for every year. Depending on the increase or decrease in the individual company average the risk factor of that company is determined.

IV. SYSTEM SPECIFICATIONS

A. Admin and Data Management

The application framework has two different types of login, Admin and User. The Admin login serves the purpose by providing the necessary data that is required for the user and the application. The Admin has the authority to upload the required data and files. The admin can change the data according to the users requirement, that is, after a particular amount of time if the user wishes to see the latest data the admin can be contacted and the latest data can be uploaded by the admin.
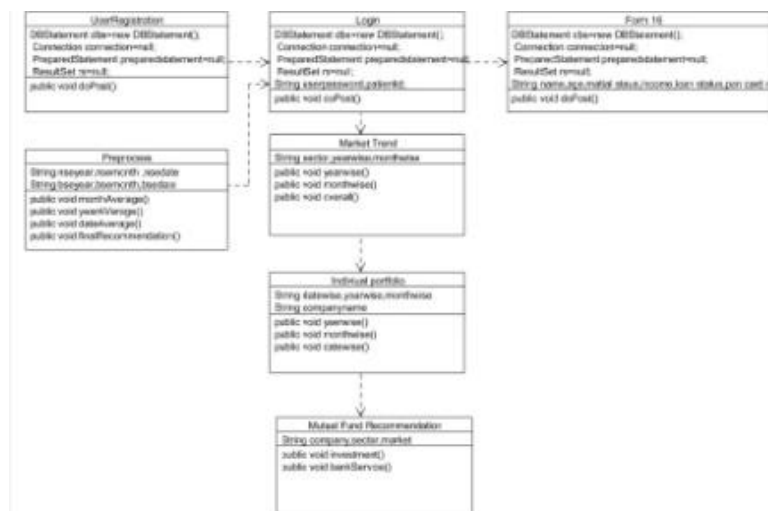

Fig. 1.

B. BigData Processing

Immense Collection of information is recovered from open source datasets that are freely accessible from significant Application Providers like website Money Control. Enormous Data Schemas were broken down and analysed a basic working idead of the Schema is understood. The CSV and TSV documents are Stored in Hadoop or HDFS and were controlled by Master for reading and writing purposes and is manipulated utilizing Java Application that is also created by us which is developer benevolent, light weighted and effortlessly modifi-able.

C. User data collection

The user data that is collected in the front end undergoes preprocessing for the framework to use it. The preprocessing is a backend work running in Hadoop groups and furthermore called as long running job as it is booked to process large amount of information with the goal that the application would makes utilization of the outcomes created for updation. The client shared assets venture in view of the client certification, for example, Aadhar number, Account points of interest, yearly wage, working status, loan status, and so on every one of these fields will be accessible on form16. Dataset mapping process is done in the preprocessing stage the whole area both National Stock Exchange and Bombay Stock Exchange organizations will be mapped date savvy, month insightful and year astute information. The calculation of risk it is to invest on every sector and companies will be finished over the preprocessing time.

D. Market Analysis

The main purpose of the application framework is to give a detailed history about the companies in the three years period. The application gives history of market information in NSE and BSE, so client will get through knowledge regarding market volatility, the expected rate of returns, the risks and the assets distribution. Client can see the historical backdrop of all the division by utilizing our right tools. There are various tools that we have set up in our framework program, tools like graphical representations of every sector/company, detailed rate of returns on very company, the top performers/losers, etc. Among the segments like automobiles, Energy, FMCG, Banking we have diverse organizations client can see history of every organization in the premise of date astute, month and year insightful. This gives the user to choose the right type of investment he wants to make after considering all the possibilities.

E. User Portfolio Management

The main purpose of the application framework is to give a detailed history about the companies in the three years period. The application gives history of market information in NSE and BSE, so client will get through knowledge regarding market volatility, the expected rate of returns, the risks and the assets distribution. Client can see the historical backdrop of all the division by utilizing our right tools. There are various tools that we have set up in our framework program, tools like graphical representations of every sector/company, detailed rate of returns on very company, the top performers/losers, etc. Among the segments like automobiles, Energy, FMCG, Banking we have diverse organizations client can see history of every organization in the premise of date astute, month and year insightful. This gives the user to choose the right type of investment he wants to make after considering all the possibilities.

V. PROJECT IMPLEMENTATION

The bigdata infrastructure solution for this project is im-plemented on JAVA. Various other technologies are used in the parallel processing in between the client and server. The communication between the client and ther server is done using TCP/IP. The hardware components can also be used to improve the latency and the efficiency. The compare the betterment of the storage devices we can use SDD and HDD storage systems. SDD, solid state drive is very commonly used in cut throat timing business applications where as small as a millisecond also counts. As we know that the prices in share markets fluctuate every millisecond it becomes really crucial to have a fast performing storage system. Also a scalable database management system is implemented for the purpose of storing users banking details which is later used for the user to buy the suggested stocks by the framework or any other stock as the user requests.

VI. FUTURE ENHANCEMENT

The Indian Share market are still at their growing stages - NSE and BSE, when compared to the other large scale share markets all around the world. The data generated by Indian share markets is increasing second by second. The analysing of share market data problem (because of large scale)can be solved with either by infrastructure or by using proper big data tools.
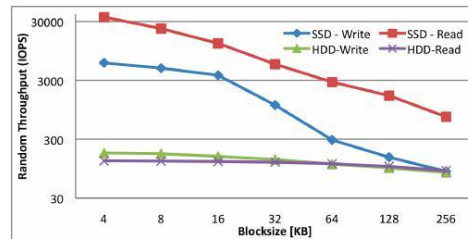
Fig. 2.

This project solves a small problem when looked at the bigger picture. The stock suggestion algorithm can be improved with various data analytics tools. Improving the stock recommendation program can lead to better data driven results.

REFERENCES

[1] A. J. Menkveld, "High frequency trading and the new market makers," Journal of Financial Markets, vol. 16, pp. 712-740, 2013.

[2] M. OHara, "What is a quote?," The Journal of Trading, vol. 5, pp. 10-16, 2010.

[3] A. P. Chaboud, B. Chiquoine, E. Hjalmarsson, and C. Vega, "Rise of the machines: Algorithmic trading in the foreign exchange market," The Journal of Finance, vol. 69, pp. 2045-2084, 2014.

[4] E. Boehmer, K. Y. Fong, and J. J. Wu, "International evidence on algorithmic trading," in American Finance Association (AFA) 2013 San Diego Meetings, 2014.

[5] M. Lewis, "An Adaptation From Flash Boys: A Wall Street Revolt," in The New York Times, ed, 2014.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: Univer-sity Science, 1989.