

ANALYSIS AND PREDICTION OF STOCK MARKET USING BIG DATA TECHNOLOGY

Akriti Tyagi, Anjali Sharma, Karan Pratap Singh and Purwa Maheshwari

Department of Computer Science & Engineering,

ABES Institute of Technology, Uttar Pradesh, India

ABSTRACT

Stock Market is the financial ground on which the high amount of data is released at every single point of time, which is very complex and nonlinear in nature. These data sets are used to predict the high profits and risks that a user can have in investing in a company. To predict such an analysis there should be adjacent to the accurate result which is one of the crucial challenges. In this paper, the goal is to forecast the prediction through analyzing the stock data for those clients who are interested in investing in a company which is done via Big Data technology using Hadoop, HDFS, MapReduce, Sqoop, and Hive. This is a process of financial decision making for the investments. This model will help the shareholders to know the current scenario of any company and the market.

KEYWORDS: Stock Market, Hadoop, HDFS, Sqoop, and Hive

INTRODUCTION

The Stock Market Analysis and the prediction model is the model through which the shareholders can be benefited by knowing all the important aspects which can give a profit. The investment is completely a profit and risk affiliated criteria. So this model tries to bridge this gap between the shareholder and the investor. The data sets releases are very complex and nonlinear in nature. To manage this unstructured and heavy data sets Big Data is used because it is the technology which is used to analyze the heavy data sets. There is a fundamental decision process before investment as needs an adjacent value to the accurate result.

Stock Market: Stock Market is a platform where the companies (i.e., the shareholders) and the investors come together to participate in investing IN the companies shares

Big Data: Big Data is the technology to analysand the large and massive data sets which are having a huge amount of data, this may be structured or unstructured. The data can be retrieved from Facebook, Twitter, or real-time data. Most of the data sets are analysand on the singer server environment but whenever the data set increases there is a need for increased infrastructure to handle the data sets with high memory speed and storage drives. The data sets are in Hera-bytes, Pentax-bytes or ea-bytes.

Hadoop: Hadoop is an open source software platform which provides high availability, reliability, flexibility for data partitioning and managing. It processes the data and stores it in the distributed environment.

HDFS: HDFS is Hardtop Distributed File System, it is based on master and slave architect. There are two types of nodes on HDFS Namenode and Datanodes, Name node works as master and the Anodes works as workers. All the metadata (the data about the data) information is stored on Ndjamena and the original data is

stored on the anodes.

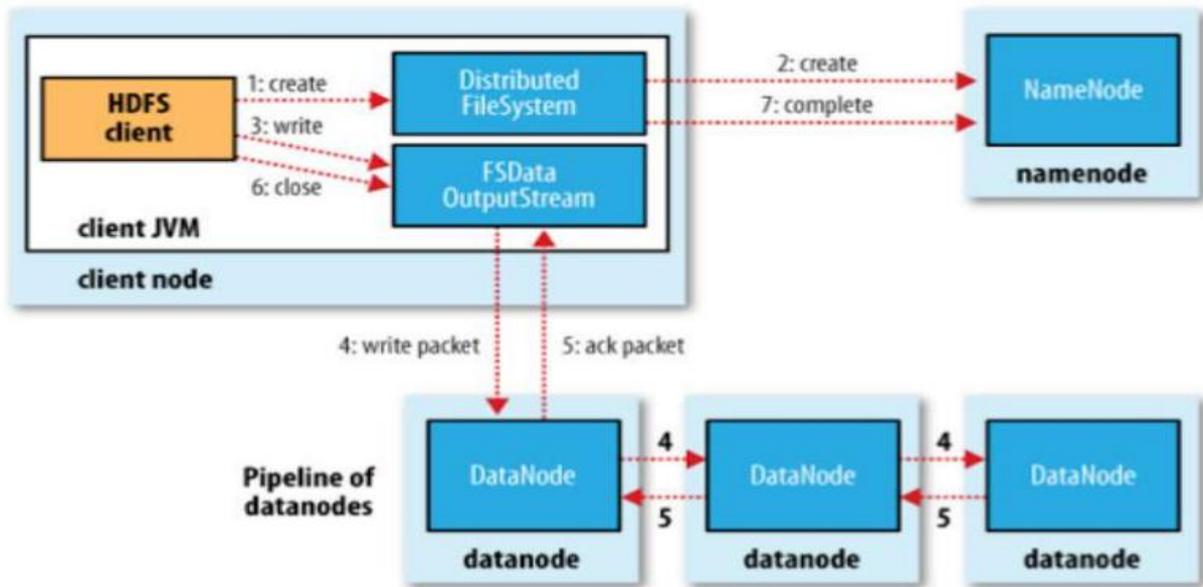


Fig1. HDFS Architect

MAP-REDUCE: MapReduce is the framework for writing applications that process huge amounts of structured or unstructured data stored in the Hadoop Distributed File System (HDFS). Hadoop Yarn opened Hadoop to other data processing engines that run alongside with the existing MapReduce jobs to process data simultaneously in many different ways.

The **Mapper function** divides the input into ranges by the InputFormat and then creates a map task for each range of the input. The JobTracker distributes those tasks to the data nodes. The output of each map task is partitioned into a group of key-value pairs for each reducer.

The **Reducer function** then collects all the results and combines them to answer the larger problem that the name node needs to solve. Every reducer pulls the relevant partition from the machines where the maps executed, then writes its output back into the HDFS. Thus, the reducer will be able to collect the data from all of the mappers for the keys and combine them to solve the problem.

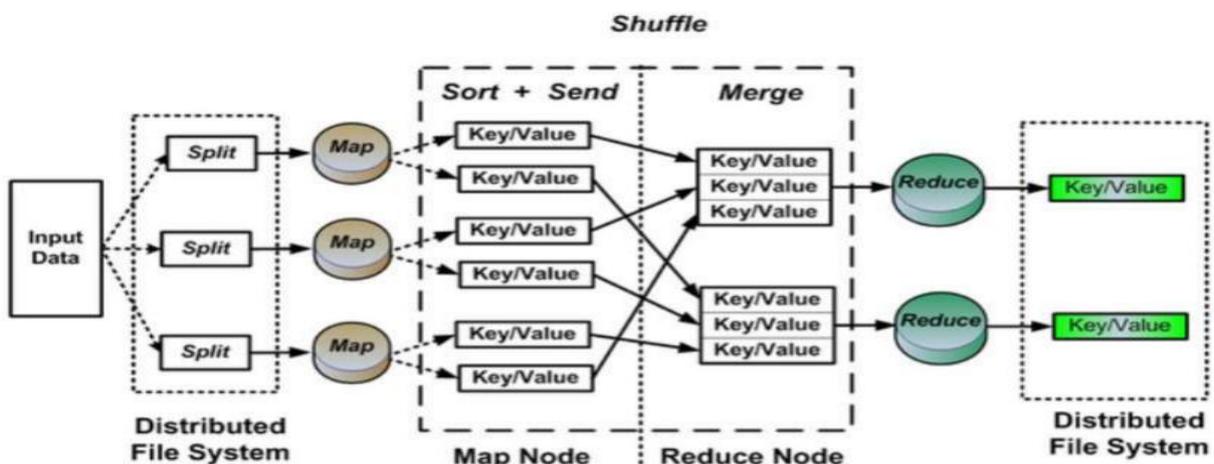


Fig2. Architecture of MapReduce

Sqoop: Sqoop vigorously transfers the bulk of data between Hadoop and structured data stores such as relational databases. Sqoop can also be used to extract data from Hadoop and export it into external structured data stores. Sqoop works with relational databases such as Teradata, HSQLDB, Netezza, Oracle, Postgres, and MySQL.

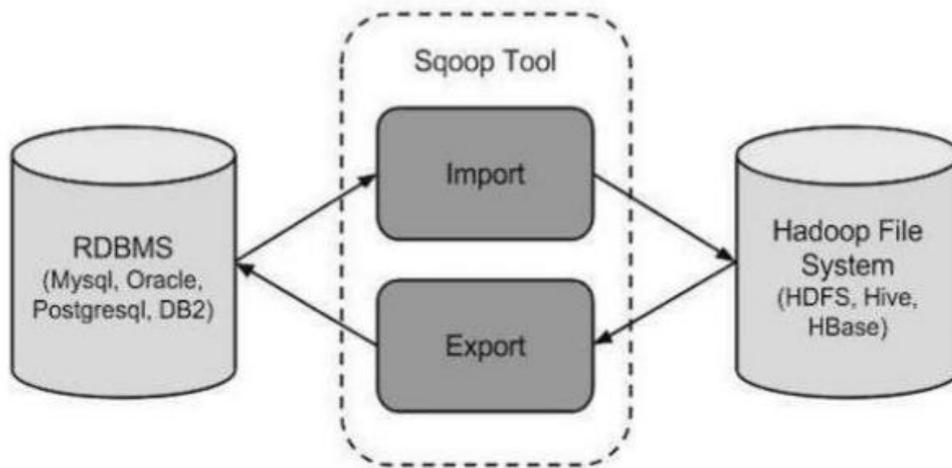


Fig3. Workflow of Sqoop

Hive: Hive is a data warehouse which is used for managing and analyzing the data. The Hive provides SQL like structure to query the data through which the data is processed.

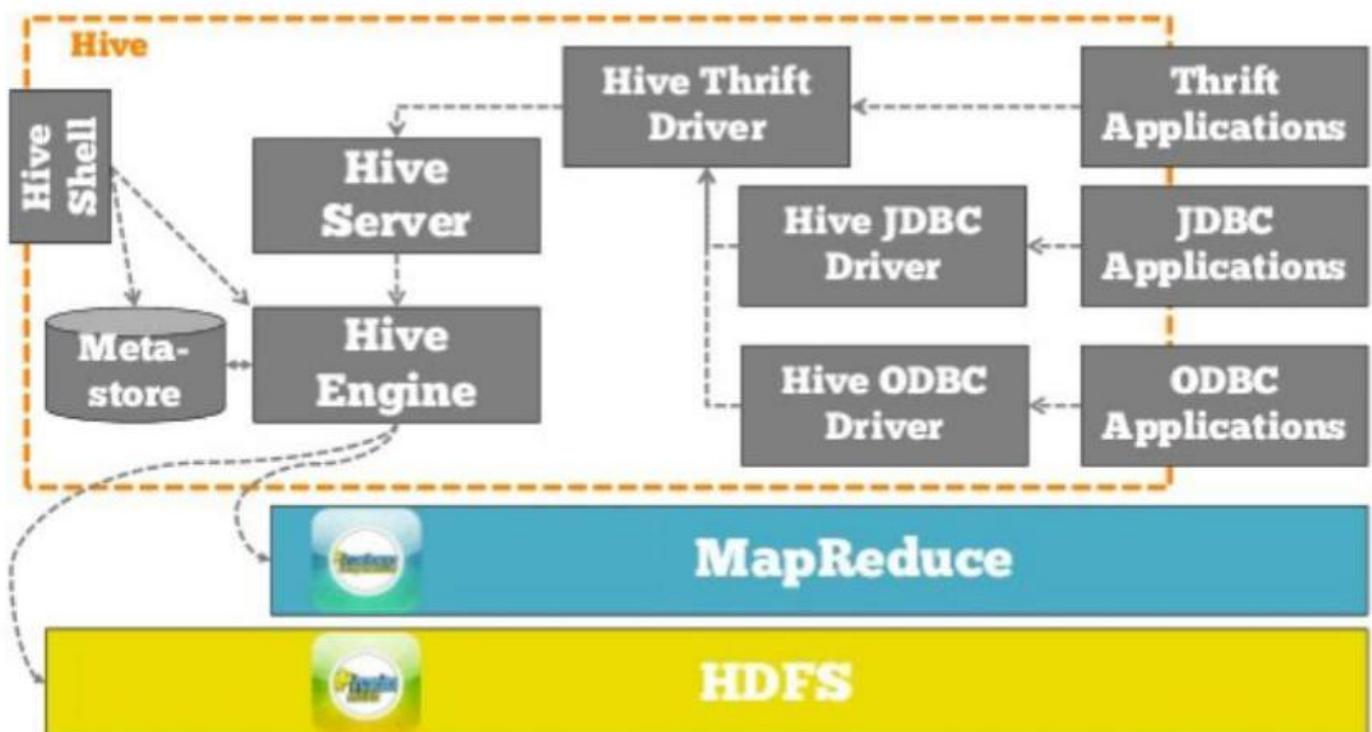


Fig3. Hive Architecture

OBJECTIVE

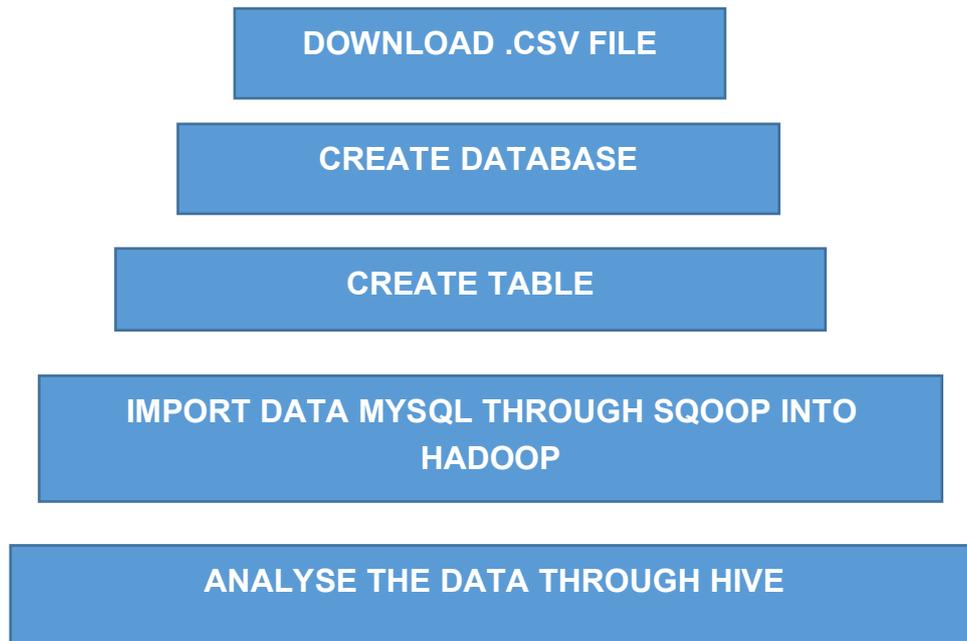
The objective of this paper is to provide a sight of vision towards the condition of a company on which an investor is interested to invest in. In share market, every investor should have the right to know all the ups and downs of a company through which the company is going, so that one can be prevented from the loss and also be benefited through the idea of analyzing.

Here the plan is to have such a framework which provides the transparency to the investor and the market.

PROBLEM DOMAIN

According to the research analysis of the stock exchange data through big data, technology is only possible with the heavy data sets. The analysis of data is done as per the prices (low price, high price, etc.) of the company which should correspondingly contribute in the analysis of a company stock data so that on one have all the profit and loss aspect on a share investment.

METHODOLOGY



Here the data is downloaded in dot CSV format from the NSE (National Stock Exchange). Stock data set is taken to 5 years of the software companies by covering the range of 1 day, 2 day, 1 week, 3 months, 1 year, 2 year and 5 years.

As the big data is itself means the massive data that is why the data here is streamed for the model so that the analysis can be done the heavy data set.

CONCLUSION

This research paper contributes to the analysis of stock data through big data technology to benefit the long-term investors. The analysis is done on the basis of the past and the current data with respect to the open price, close price, high price and low price of a stock. The investors are getting to know about every company in detail and also about the ratings to get a profit on each and every share investment.

FUTURE SCOPE

For the further consideration, the plan is to work on the live data streaming. So, that not only the long-term investors but the short term investors can also attain the profit. The companies shares are going high and low

on the regular premises. Therefore for the better surety and support to the user, they have the right to get to know about each and every important norm of the company.

It is also needed to manage the analysis for short-term investors so that they can also extract the profit from share market.

REFERENCES

[1] Kushagra Sahu, REVATI PAWAR, SONALI TILEKAR, RESHMA SATPUTE, Stock-exchange forecasting using Hadoop MapReduce technique, International Journal of Advancements in Research & Technology, Volume 2, Issue04, April 2013.

[2] Jeffrey and Sanjay Map-Reduce: Simplified processing on the large cluster, Google Research Publication 2004.