# Stock Market Prediction System using Map Reduce Apriori Algorithm

Dr. M. Gobi[1], A.Kowshika[2]
Assistant Professor[1], Research Scholar[2]
Department of Computer Science
Chikkana Government Arts College, Thirupur, India

**Abstract:**
Stock Market contains very huge amount of data that is working in terabytes and peta bytes, these are very complex and can be learned by a data mining methods. Stock Market has high profit and high risk features that's why its prediction must be in parallel of accuracy .It has very large amount of investors who invest their money into shares and selling shares or buying the shares. Modified Map Reduce Apriori algorithm with Top Down approach is implemented on datasets.

**Keywords:** Map-Reduce, Apriori algorithm, Data mining Support, Association rules, Top down Approach.

## I . INTRODUCTION

Data mining and knowledge discovery Process are used to extract useful, hidden and unknown patterns and knowledge from large database. The stock market has by high risk, high-yield, so investors are concerned about the analysis of the stock market and trying to get prediction of the stock market. However, stock market is impacted by the politics, economy and many other factors, coupled with the complexity. Analysis of large volume of data and to process it, is difficult and challenging.  To analyze this data we use the map reduce framework it analyze the scattered data of stock market and predict the future trends and solutions which would benefit to the investor. It has a file system that provides an interface between the user's applications and the local file system, which is the Hadoop Distributed File System (HDFS).Map Reduce Apriori Algorithm is modified using top down Approach and implemented on datasets.

## II  STOCK MARKET PREDICTION SYSTEM

User or New stock holder is the end user of the application and searches for the company shares to invest money. And can also sell the shares. Money control is the leading financial information source. Managing your finance with online investment portfolios Live Stock Price, Stock trading News Live etc. Server act as  intermediate to communicate with the database and end user. It is responsible for generating proper query to retrieve information from the database. It consist the data of the reviews that are fetched from the Money Control Website. In categorizing the stock market prediction systems different dimensions can be considered**:**

**Input data:**
Some prediction methods are based on historical market prices and use technical analysis to predict the market. Various methods are based on analyzing the news content; however combination of historical market data and news can also be used.

**Prediction goal:**
The  market prediction goal can be the future stock price or the volatility of the prices or market trend. Market trend is the general direction of the stocks prices which is upward or downward. Market volatility is defined   as: the amount of uncertainty or risk  above the size of changes in a security's value. A higher volatility means the higher fluctuation of their corresponding stock prices.

**Prediction horizon:**

prediction horizon is  time span in which the prediction would be valid. This can be short-term or long-term prediction which starts from 5 minutes to 1 hour after the news release and long term starts from 24 hours and can last longer. stock marketing is a network which provdes a platform for almost all major economic transactions in the world at a dynamic rate called the stock value which is based on market equilibrium. Predicting those stock value offers enormous  profit opportunities which are a huge motivation for research in this area. Knowledge of a stock value can change even a fraction of a second result in high profits. Similarly, a correct prediction result can be extremely profitable in the amortized case. This eager of finding a solution has prompted researchers, in both industry and academia fields to find a way past the problems like volatility, dependence and seasonality on time, economies and rest of the market.

## III  PROBLEM STATEMENT

Through this system the investors are able to obtain the time stock information, finding stock chart, news and research on internet that can help users to find the right investments strategies with good profit. Map reduce is the framework where we can access large amount of data in parallel manner. (HDFS) Hadoop distributed file system is used to manage the file system. The  Map Reduce engine uses to distribute work around cluster. The people using the Internet to invest is growing fast. Stock Market has high profit and high risk features that's why its prediction must be in parallel of accuracy.It has very large amount of investors who invests their money into shares and selling  or buying the shares. The stock value runs up and down of record territory, investors are increasingly turning to the Web to research; discuss the trade stocks and securities. By using different technique can get accurate reliable prediction result which give consumer better

solution for where to invest their valuable money. The stock markets has become an integral part of the global economy. Any fluctuation in this market could influences our personal and corporate financial lives, and the economic health of a country. The stock market has been one of the most popular investments due to its high returns. There is always some risk to investment in the Stock market due to its unpredictable behaviour. So, an 'intelligent' prediction model for stock market would be highly desirable and would of wider interest. Stock price prediction is the most important topics in finance and business.The stock market domain is dynamic and unpredictable. Several research efforts have been carried out to predict the market profit using different techniques. These techniques cannot provide deeper analysis which is required and therefore It is not effective in predicting market prices. To analyze large volume of data and process it is difficult and challenging. Analyze this data using map reduce framework it analyze the scattered data of stock market and predict the future trends which would benefit to the investor.

Lots of investors are involved in stock marketing and they are interested to know more about the future of market which can make successful investments. Effective market prediction can help investors with trade advices or can be used as component inside automatic trader agents. Stock value prediction systems it indirectly help traders by providing supportive information such as the future market direction progress. E.g.If the direction of selected stock during 24 hours is predicted to be up, buy those stocks would be a profitable trading action.

## IV. MAP REDUCE ARCHITECTURE

Map Reduce framework manages scheduling tasks, monitoring them and re-executes the failed tasks .It has a single master Job Tracker and one slave Task Tracker per cluster node. Apriori Algorithm is also used for generating association rules on the basis of future trends. In the proposed system, map reduce component is used to allows work on small modules and works parallel. DFS provides data replication upto three times to avoid loss of data in case of media failures. The Master node stores the huge data in HDFS and runs parallel computations on all the data i.e. Map Reduce. (DFS)Distributed file system that provides a limited interface for managing the file system to allow it to scale and provide high.
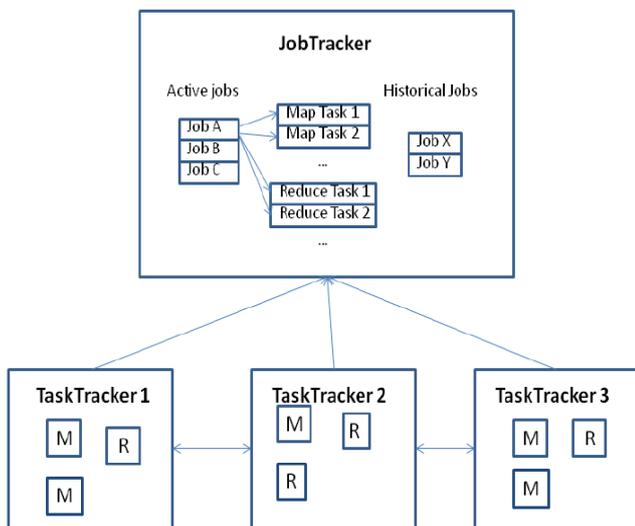


HDFS creates multiple replica of each data block and distributed those blocks over computers throughout a cluster to enable reliable and rapid access.
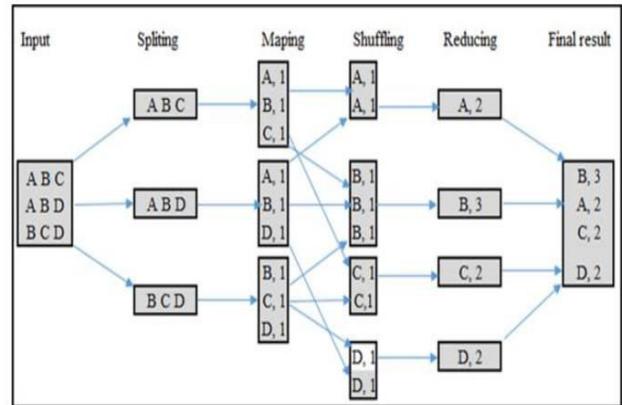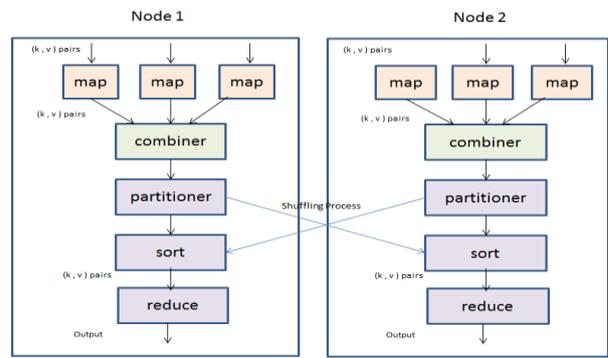


When a file is loaded into HDFS, it is replicated and fragmented into blocks of data, which is stored across the cluster nodes.

### Map Reduce Apriori Algorithm

i. Scan the dataset to calculate support S of each item.
ii. min _supp=number/total number of items.
iii. If support S is greater than min _sup then add an item to frequent 1-itemset.
iv. Compute frequent item set for each map node using min _sup and collect all together in reduce phase.
v. Remove items that do not meet the min _sup.
vii. Collect the frequent item set at the reduce node and count item frequencies compared with min _sup.
viii. Remove the items that do not meet the min _sup in Reduce Node using prune().

**Table.1.Input Transactions**

| TID | ITEMSET |
|-----|---------|
| 200 | K,E,Y,M,K,N |
| 201 | D,O,N,K,E,Y |
| 202 | C,O,K,E,V |
| 203 | K,Y,C,M |
| 204 | L,O,O,E,M |

**Table.2. Support Count of Each Item Set**

| ITEMSET | SUPPORT COUNT |
|---------|---------------|
| K | 5 |
| E | 4 |
| Y | 3 |
| M | 3 |
| N | 2 |
| D | 1 |
| O | 4 |
| C | 2 |
| L | 1 |
| V | 1 |

Support count of each item is calculated and compares min _sup. min _sup=60/100*number of transactions. Remove the items which are less than min _sup.
Min _sup=60/100*5=3

**Table.3.Frequent item set**

| ITEMSET | SUPPORT COUNT |
|---------|---------------|
| K | 5 |
| E | 4 |
| Y | 3 |
| M | 3 |
| O | 4 |

## FREQUENT ITEM SET GENERATION
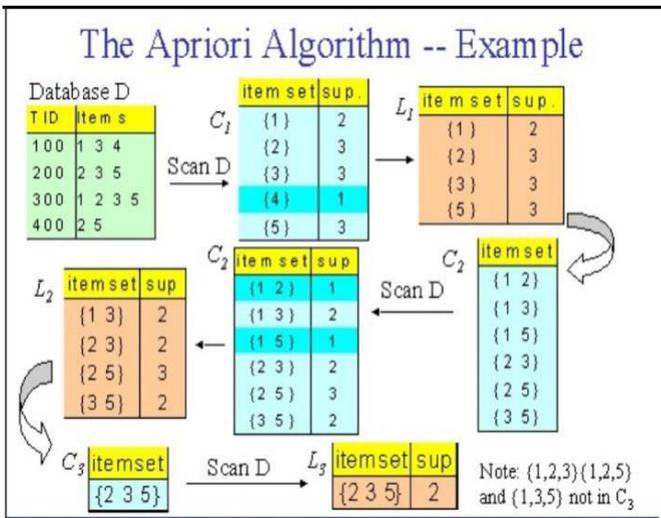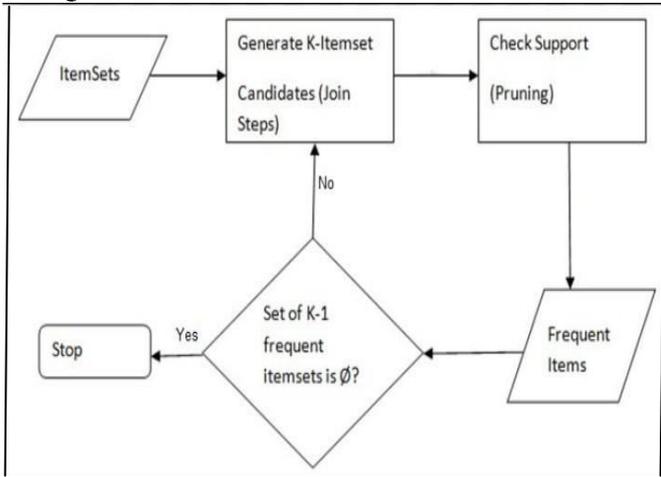




New Proposed Top down approach overcomes the deficiency by reducing number of database scan also useful for large amount of data base scan. Implementation of previous Algorithm based on MR Apriori algorithms. Three algorithms are implemented MR Apriori and existing algorithms (one phase and k-phase) based on hadoop map reduce programming model. Hybrid approach takes less time than previous algorithm because it takes all item set at once in map which satisfies the min-sup criteria. Parallel execution of input occurs in Map Reduce which takes less time as compare to previous algorithm.

## V. CONCLUSION

In this paper an algorithm is implemented which is hybrid approach of modified Apriori with map reduce for generating stock market prediction. It will overcome the deficiency of classical previous algorithm hence reduces the number database scan and it is useful for large amount of database scans. By using different technique can get accurate reliable prediction result which give consumer better solution for where to invest their valuable money. In future it can be enhanced for variety of datasets directly for feasibility. Parallel execution of input occurs in hybrid approach which takes less time and easy to implement as compare to previous method. This research intends that can help users to find the right investments strategies with good profit. The results show that strategy designed has higher efficiency and takes less time for execution.

## VI. REFERENCES

[1].Sudhakar Singh, Review of Apriori Based Algorithms on Map Reduce Framework.ICC-2014,pp 593-604.

[2].Kyuseok Shim, MapReduce Algorithms for Big Data Analysis, DNIS 2013, LNCS 7813, pp. 44–48, 2013.

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

[4].Sruthi M,Parallel Implementation of modified Apriori Algorithms on Multicore Systems, Proceeding of IMCICICSIT 2016

[5].Apache hadoop.http:/hadoop.apache.org.

[6].Anil R.Surve, Hadoop-HBase for Finding Association Rulesusing MapReduce Apriori Algorithm,2016 IEEE.

[7].Othman Yaha,An Efficient Implementation of Apriori Algorithm based on Hadoop-MapReduce Model. Vol. 12 © 2009 - 2012 IJRIC & LLS.

[8].Xueyan Lin.MR-Apriori:Association Rules Algorithm Based on MapReduce.978-1-4799-3279-5/14/$31.00

[9].M.Afzali,Hadoop-MapReduce:A Platform for finding Large datasets,978-9-3805-4421-2/16/$31.00 c 2016 IEEE".

[10].T. Kimoto, K. Asakawa, M. Yoda, and M. Tajeoka, Stock Market Prediction System with Modular Neural Networks, In Proc.of IEEE International Joint Conference, on Neural Networks, 1990, pp. 1-16.

[11].K.Kamijo,& T.Tanigawa StockPrice Pattern Recognition A Recurrent Neural Network Approach. In Proc. of IEEE International Joint Conference on Neural Networks, 1990, pp. 215-221.

[12]. M. Yoda, Predicting the Tokyo Stock Market. In: Trading on The Edge, Deboeck, John Wiley & Sons Inc., 1994, pp. 66-79.

[13].Zarowin, P., 1988, Size, seasonality and stock market overreaction, Manuscript (New York University, New York, NY).

[14].Oppenheimer, H.R., 1981, Common stock selection: An analysis of Benjamin Graham's 'intelligent investor' approach (YMI Research Press, Ann Arbor, MI).

[15].W. Paik, M. H. Kyoung, K. S. Min, H.R.Oh, C. Lim and M.S.Shin, Multi-stage News Classification System for

Predicting Stock Price Changes, Journal of the Korea Society for Information Management, Vol.24 No.2, 2007, pp.123-141.

[16].Yu, Y., Y. Kim, N. Kim, S.R. Jeong, Predicting the Direction of Stock Index by Using Domain-Specific Sentiment Dictionary, Journal of Intelligence and Information Systems, Vol.19, No.1, 2013.3, pp.92-110.

[17].V. Sehgal and C. Song, "SOPS: Stock Prediction using Web Sentiment," Seventh IEEE International Conference on Data Mining Workshop, 2009, pp.21-26.

[18].Shunli Ding,A comprehensive Evaluation System of Association Rules on Multi-Index,(978-1-4673-6593- 2/15) 2015 IEEE .

[19].Kim, Y., N. Kim, S.R. Jeong, "Stock index Model Using News Big Data Opinion Mining," Journal of Intelligence and Information Systems,Vol.18, No.2,2012.6, pp.143-156.

[20].H. Chen and D. Zimbra,Artificial Intelligence and Opinion mining, IEEE Intelligent Systems, May/June 2010, pp. 74-80.

[21]J. Bollen, H. Mao and X. Zeng, Twitter mood predicts the stock market, journal of Computational Science, Vol.2, 2011, pp.18.

[22].L. Zhuang, F. Jing and XY. Zhu, Movie Review Mining and Summarization, Proceedings of 15th ACM International Conference on Information and Knowledge Management, November 2006, pp.43-50.

[23].S. Ahn and S. B. Cho Stock Prediction Using News Text Mining and Time Series Analysis, Proceedings of KIISE 2010 conference, Vol.37, No.1, 2010.6, pp.364-369.

[24].R. P. Schumaker and H. Chen, Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZ Fin Text System, ACM Transactions on Information Systems, Vol.27, No.2, Article 12, February 2009.