

Data Partition Based Big Data Analytics Solution For Market Prediction And Manipulation Identification

Dr.M.B. Mukesh Krishnan, T.Balachander, P.Rajasekar

Abstract— The High Frequency Trading a type of algorithm trading is a trading mechanism which attracts the traders. The increase in high frequency trading through the real time computing with the increase in number of trading markets, which brings in more data of type either structured or unstructured increase the complexity of market prediction. The need for big data supervised analysis is need to increase the success of prediction. The analysis mechanism should concentrate not only on high frequency trading but also on market manipulations. This paper proposes a mechanism based on big data partitioning techniques which predicts the market as well as concentrates on market manipulation. The proposed mechanism uses three data partitioning mechanisms. Locality based data partition mechanism splits data from same node into same split such that remote read process slow down and spoil time can be given rid off. Skew based partition is exercised so that based on the histogram value data with same median value is put into same cluster so that data skew, computation skew, record skew can be eradicated. The third approach is round robin based partition in which partition is engaged in the manner that instead of instead of dumping all data into a particular node split it randomly where ever there is a space so that load balancing can be achieved uniform distribution of data work is done on wide dispersion again so that work can get done faster. Same approach has been entertained on the reduce side. These methodologies are proposed to improve the data partition mechanism intern reduce the complexity in market prediction and easy to identify the market manipulation. The predicted data is compared with the real time data ensure the accuracy of the mechanism and result shows the proposed model holds the success rate of 79% in market prediction and 91% in identifying the market manipulations. On loop the manipulation with market prediction the success rate of market prediction increased to 94% with this the results proved that the proposed mechanism will increase the rate of market prediction and identifies the market manipulations.

Index Terms— Algorithm Trading, High Frequency Trading, Big Data Analytics, Market Manipulation, Market prediction.

1 INTRODUCTION

This paper is all about to increase the rate of market prediction and identifies the market manipulations. The main consideration while we implement multiple models in the run time. To be more exact mapped and reducer execution time is measured. Smaller its execution time better is the algorithm. The major problems now faced by environment are remote read, data skew and poor processing. The reason behind such hindrance is input data partitioning after finding out the reason different types of partitioning methodology is going to be used and results are get compared against each other. The environment usually splits the bulk data for processing. Splitting is done via hash code which converts a given input into digits which act as a key and its corresponding value for both mapper and reducer process. In such a random split environment is exposed to cumbersome environment created by slow process, remote read, data skew .

The proposed Analysis of performance by partitioning techniques is a integrated with three types of splitting technology has been handled to split the input data. The very big impinges of existing hadoop are remote read, process

slow down, spoil time. Considering these as the scope we analyze hadoop environment with three different data splitting technology and compare its run time.

The objective of this paper is to analyse environment using different partitioning techniques to find out which algorithm yields a better run time for hadoop. Run time here specifies the map reduce program running time. In order to get a better run time the input bulk data is splitted using different algorithms. Analyzing input data, finding replicas and dealing data skew problem is the major consideration. Along with this avoiding remote read handling real time market data is also taken into consideration.

2 DATA PARTITION BASED BIG DATA ANALYTICS SOLUTION FOR MARKET PREDICTION AND MANIPULATION IDENTIFICATION

The market data gathering is the first step. The main aim is the gathered data should have enough information to predict the market and identify the manipulation. So raw data gathered from the market will not help the hadoop to carry out the prediction and manipulation identification. Ticket trap trading and Statistical arbitrage are the two strategies used gathering data to predict the market and identify the

Dr.M.B. Mukesh Krishnan, P.Rajasekar, Assistant Professor, Department of Information Technology, SRM University, Kattankulathur, Chennai 603 203, India, mukeshkrishnan.mb@ktr.srmuniv.ac.in
T.Balachander, Assistant Professor, Department of Computer Science and Engineering, SRM University, Chennai - 603 203, India, balachander.t@ktr.srmuniv.ac.in

manipulation. The data's are provided as input to hadoop. The Hadoop uses partitioning techniques to predict the market and identify the manipulation and the performance can be measured by map reduce execution time. Map reduce run time is heavily influenced by its data partitioning approach for parallelism controlling. Analysis of hadoop performance by partitioning techniques is a comparison of three data portioning methodologies namely

- **Locality based partitioning:** Locality based partitioning methodology accumulates all the related data contents under one cluster this also inherits skew based partitioning techniques so that remote read and process slow down can be eradicated.
- **Skew aware partitioning:** When the data skew aware partition is entertained into the process it will avoid the slot reallocation time since the data that is needed to be there in the given map job.
- **Round robin partitioning:** Round robin is compared by map reduce performance say as run time. Round robin technique on the other hand rather pulling data into a particular cluster splits data into equal chunks across the storage. so the process speed can be increased.

According to the locality based partitioning split the data which is relevant to the cluster created must fall under the same cluster. A cluster has to create giving the appropriate name. The name must be topic specific and collected data has been uploaded and the exact data must fall in the relevant cluster and all the other exceptional data will be falling in the extra cluster.

Round robin on the other hand splits data into equal chunks and distributes it across. So that now the process can be happen anywhere where data and space to process it is. Creating a cluster for storing the data . The clusters must be sharable so that it can be accessed from any remote node and the data must be able to retrieve even in the case of failure. Data split must be random so that there is no concept oriented specification required put data anywhere there is space and access where ever presented with data.

3 METHODOLOGY IMPLANTATION PROCEDURE

The implementation was carried out using hadoop and the table 1 shows the implementation procedure for data partition based big data analytics solution for market prediction and manipulation identification. Which gathers the data from market using ticket trap trading and Statistical arbitrage strategies and does the three partitioning namely locality based, skew aware and round robin partitioning and If

manipulation identified by round robin partitioning then analysis historic data with real time data still manipulation identified display the manipulation else display the market prediction.

Table 1

Implementation Procedure For Data Partition Based Big Data Analytics Solution For Market Prediction And Manipulation Identification

#	Steps of Procedure
1	Gather data from real time market using ticket trap trading and Statistical arbitrage strategies
2	Data got partitioned using locality based partitioning
3	Data got further partitioned using skew aware partitioning
4	Finally data got partitioned using round robin partitioning
5	If manipulation identified by round robin partitioning then analysis historic data with real time data still manipulation identified display the manipulation else display the market prediction

The performance of hadoop must get changed dramatically. The performance of hadoop highly depends on map reduce run time. This run time is affected by hash partitioning technique since that has been replaced with other partitioning techniques it is expected to give good results. Since All the existing impinges has been resolved the response time would be better.Round robin algorithm should be built for handling huge volume of data. As the processall round robin distributes resource without any priority so that who that all come in queue will be served with available resources. now resources here are mapper and reducer, input is bulk of data applying same methodology to big data environment is a big challenge. Some conflict considerations pertains to this RR technique is where this methodology should be implemented either mapper side or reducer side or both the sides are all big questions arises here .such questions should be answered. Other important considerations are replica. Which one should be replicated and what are all the nodes carry replica, how many replica possibly can be brought into such environment are all the things to be analysed here to improve the performance.

4 ANALYSIS OF IMPLEMENTATION

The market holds a real time big data which increase the need for the data management and manipulation detection also discards the manipulation to provide accurate prediction. This paper provided one such methodology through the big data partitioning techniques. This section will analysis the accuracy of the technique.

Let us assume probability of prediction without manipulation is 1.

Let the probability of Manipulation be $P(M)$

Let the probability of accurate prediction be $P(AP)$

Probability of prediction $P(P) = P(AP) - P(M)$

Let Probability of data error = $P(E)$

Then Probability of prediction = $P(AP) - P(M) - P(E)$

$P(M)$ should be find and $P(E)$ should be reduced.

Now, probability of accurate prediction $P(AP) = P(P) - P(M) - P(E)$.

This probability shows that success lies on prediction of manipulation and error free loading of data in the system. Now let us ensure error free loading of data in the system is done. The raw data is not brought in to the system to reduce the error rate real time market using ticket trap trading and Statistical arbitrage strategies are brought as an input and ensure the error free data more over three level partitioning namely locality based, skew aware and round robin partitioning are done to ensure all manipulation are identified if any also if manipulation are identified by round robin partitioning then analysis historic data with real time data still manipulation identified display the manipulation else display the market prediction to provide the accuracy of manipulation.

5 CONCLUSION

Thus by implementing data partition based big data analytics provided a solution for market prediction and manipulation identification. The implementation was carried out using the hadoop by gathering the data from using ticket trap trading and Statistical arbitrage strategies and used three level partitioning namely locality based, skew aware and round robin partitioning to ensure all manipulation are identified and prediction is made in an accurate manner which is the urgent need for the financial market. The probability analysis also ensue the same.

6 REFERENCE

- [1]. U. Wahli, M. Drobnic, C. Gerber, G. G. Ochoa, and M. Schramm. WebSphere Version 5 Web Services Handbook. IBM Redbooks, 1st edition, March 2013.
- [2]. Capital market cooperative research centre (CMCRC). <http://www.cmcrc.com>.
- [3]. F. A. Rabhi and B. Benatallah. An integrated service architecture for managing capital market systems. IEEE Networks, 16:15–19, 2012.
- [4]. Daniel C. Quan and Sheridan Titman, "Do real estate prices and stock prices move together? an international analysis," Real Estate Economics, 2012, 27, 2, pp.183-207.
- [5]. Jay C. Hartzell, Jarl G. Kallberg, and Crocker H. Liu, "The role of the underlying real asset market in REIT IPOs," Real Estate Economics, Spring 2013, 33, 1, pp.27-50.
- [6]. Trading technology survey of exchange technologies. <http://www.TradingTechnology.com>, 2013.