Asset Management in Manufacturing Industries Using Big Data Analytical Architecture

Dr.T.R.Saravanan[1], R.S.Vaibhave[2]

[1]Assistant Professor, Department of Computer Science and Engineering, JEPPIAAR SRR Engineering College, Padur, Chennai -603103.

---***---

Abstract - *The asset management is for managing company's details by storing and sorting them. The company which wants to have a good place in market can make use of these ideas of storing data and further analyzing them. The storage is about details of the workers, machine and their functionalities, stock market details which require huge storage and therefore big data concept is employed. Handling with the data includes Hadoop, data mining technologies which are essential for this architecture. The structured, semi-structured and unstructured data has to be managed. The major purpose is to make use of the unused data of the company. The unused data are just the ones which have low necessity for the company but maybe of some use to the employee such as the machine details and the instructions of operating them. The employee can also get access to their own salary details by having authentication. By giving more features for the user the company can be more organized and easy for the employee at the same time. Here the data scientist plays an important role for visualizing the data which are to be given to the employee. The clustering and classification algorithms are used for sorted storage. The prediction process is carried after the report of all the data has been collected. Resultant graph is being generated so that it is easy for the authorities to analyze the growth and their position in market.*

Key Words: **authentication, big data, classification, clustering, Hadoop, market**

1. INTRODUCTION

A huge amount of data for operating and maintaining is required and the asset of the company serves this purpose. Nowadays, every industry uses asset management functions for storing large amount of data. But they aren't using the best of it. Some companies don't even process the stored or existing data. This results in the degradation of the company's performance and its efficiency. The Information and Communication Technologies have become dynamic and the industries should develop their standards by using these technologies efficiently [1]. Many applications and systems which are governed by these ICT's are implemented in some industrial areas. Though they make use of them, these technologies do not yield a complete overview about the system. Here the users and the data scientists are the major part of the system, primarily when it comes under the tree of big data. The big data objective is to process and handle data which is of high velocity, volume, variety and veracity [2]. In the view of Information and Communication Technology, Big Data is the best technology to be used in industries [3]. The data scientist must be an expert in solving the problems which arises during the understanding of data, preparing the data, modelling the data, evaluating the data and deploying it efficiently. Managing the assets in an efficient way with many facilities is the major goal. This includes maintaining the company's standards and also making the company more organized. The asset management system provides an easy way for storing the company's data and managing it. The architecture paves way for more accurate grouping of data and storing them in ordered way. The Hadoop distributed file system is used so that any type of data regarding the company can be handled. The data may be text, images and other formats also. The system overall generates a prediction report for the growth of the company.

The upcoming part of the paper is arranged in the following manner. In section 2, this paper reviews about how the uploading of data and pre-processing is carried out. In section 3, the classification and clustering of the data is discussed. In section 4, the analysis and the prediction of the growth is explained. In section 5, the aspects of the user interface is presented. Section 6 depicts the overall structure of the system by explaining the architectural design. The section 7, describes the tools involved to achieve the successive environment. The following section consists of the future work to be carried out. Finally the section 8 summarizes the outcomes. These all operates in a single system and provides efficient management through big data technology.

2. FILE UPLOAD AND PREPROCESSING

File uploading is the chief process which is carried out in every big data projects in order to work with it. Here in this system, the data stored up to date is to be considered for processing. Data gathering is the major step taken to perform stock market analysis. In the Manufacturing industrial sector, monitoring of various assets usually the sensors are used to keep the machineries in an optimal state and to reduce the maintenance downtime. This converts the physical machinery property into an electrical signal. Then the manipulation is performed using the filters to reduce the noise generated. And these are sent to a database where liable information can be obtained. This is also a part of asset of the company.

Each and every data obtained can be of several formats. Namely unstructured, semi structured or structured. All these data are structured finally for further processing. The complexity theory of big data helps in understanding complex formation of patterns, simplified implications, abstracting the knowledge better and provides guidance for designing algorithms and models for computing. Here the each and every data obtained is stored in the Hadoop Distributed File System (HDFS). Where the HDFS instances are divided into two components: name node and data node. Where the name node tracks the storage of data and the data node are used to actually store the data. All the data are loaded into the HDFS using the Uniform Resource Locator (URL) and the raw data are maintained by the data nodes. The admin can login into the system and can upload the data. For the uploading process, the admin will request to the name node. And the name node will check whether the user is authorized or not. After validating the user, the URL is set to keep in track of the data nodes and finally the data is stored in 'n' number of data nodes.
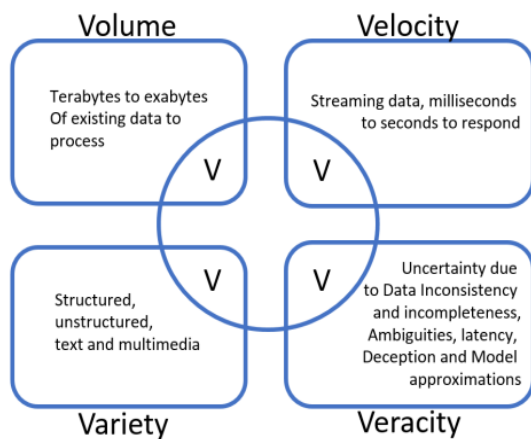

Fig -1: 4V's of Big Data

Data preprocessing is where transformation of raw data to understandable format is done. The obtained raw data is more often not clear and hence preprocessing is done. And it is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. The uploaded data is first retrieved from the HDFS. Once after retrieving it, the data is passed to the MapReduce algorithm. This process helps in organizing the unstructured or semi structured data into structured format. This process removes all the unusable values from the data sets. And finally the data reduction process is carried out and this produces the reduced representation of the data in a data warehouse.

3. DATA CLASSIFICATION AND CLUSTERING

Naive Bayes learning models have learning algorithms that have capacity to perform data analysis and pattern recognition. This algorithm is mainly used for classification and regression analysis. Naive bayes are based on the concept of decision planes that defines the decision boundaries. Automated machine learning for consistency of data is important to be considered. Of all these, clustering process involved on the data sets is the most concerned thing [8]. The pre-gathered or processed stock data attributes of companies and its marketing status will be grouped into small parts. The data grouping is performed based on the price, open, high, low, and close. The attributes used for naive bayes training were then chosen, based on the result of the "stepwise regression" analysis. The Naive Bayes classification algorithm is a probabilistic classifier. It is based on the probability models that incorporate strong independence assumptions.

The independence assumptions often do not have an impact on the reality. Therefore they are considered as naive. Probability models can be derived by using Bayes' theorem (credited to Thomas Bayes). Depending on the nature of the probability model, we can train the Naive Bayes algorithm in a supervised learning setting. The Predictive Model Markup Language standard is been confirmed by the Naive Bayes model generator. A Naive Bayes model consists of a large cube that includes the following dimensions: Input field name, Input field value for discrete fields, or input field value range for continuous fields. The Naive Bayes algorithm divides the continuously appearing fields into discrete bins. Target field value, in Naive Bayes model records how often a target field value syncs with a value of an input field. We can activate the Naive Bayes classification algorithm by using the following command: DM_Clas Settings (). DM_set Algorithm ('N aive Bayes'). To group these data into those clusters, whose stock data category has already been defined. Thus it builds up a strategy to predict the marketing of the upcoming days. The grouping of data should be composed of points separated by small distances, relative to the distances between the clusters. This clustering will apply the mapreduce with naive bayes approach to give more efficient results in high volume data clustering process.

4. ANALYSIS AND PREDICTION

Knowledge discovering includes retrieval, representation, archiving, management and preservation. Tools for knowledge discovery includes fuzzy set, rough set, soft set and near set [6] [7] [9]. Machine learning technique aims to reduce complexities and computational cost [4].

Therefore the emerging processing technologies should be used which has more cores embedded in it [5]. Since the stock market value is varying, it is important to set a benchmark for the company. Stock analysis is a key process followed to achieve this. This system analyses the previous stock data of the industry with the help of certain parameters which affects the stock value. Data mining algorithms are used to implement these values. This helps us in determining the month's high and low. Analysis of stocks using the data mining technique will be helpful for the new investors to invest in the stock market. The stock

market includes some activities like Sensex calculation and exchange of shares. The exchange of shares provides an efficient market for trading. This system analyses the Sensex based on the company's stock value. The company's stock value depends directly on some factors like: Firm-related factors, monetary policy of RBI, demand and supply, popularity, corporate results, investor confidence and political and economic shocks.

4.1. Firm-related factors

Hike in sales revenue, fall in the cost of operations, repaying debts and product launch increases future cash flows of the company. Negative factors like change in top management, high employee turnover, high manufacturing cost and product failures adversely impacts company's future earnings.

4.2. Monetary Policy of RBI

Since RBI reviews its monetary policy for every couple of months, any increase or decrease in Repo or reverse Repo rates changes the stock prices.

4.3. Demand and Supply

It is the major reason for price change in stock. When the demand increases, the supply becomes less and therefore the price rises. And if the demand decreases, the supply becomes more and therefore the price decreases.

4.4. Popularity

Popularity of the company has a direct effect on the buyers. The main strength is in the hands of the share buyers. If the company performs well, then it may result in the rise of stock price, and vice versa.

4.5. Corporate Results

These will be regarding to the profits or progress of the other companies over a span of time.

4.6. Investor Confidence

Even the confidence of the investor can regulate the market value to go up or down. It is said to be bull market, when the investor confidence is growing. Here the stock prices raises. And it is called as bear market, when the investor confidence is fading. And this subjects to fall in the stock prices.

4.7. Political and Economic shocks

Any notable change around the world could affect the economy and the stock price of an industry. An act of violence can also lead to dramatic fall in stock prices.

The price of the stock rises or falls due to different factors. Knowing the relationship between stock and other variables will ensure us to stay at a safer side. Based on

these above factors, the incrementing and decrementing cost of the stock can be predicted.

5. DATA ACCESS INTERFACE

Here in this system, the employee is also given access to the data stored in the Hadoop environment. Since the data stored about the company are under used, some of these data can serve as a guide to the employees. This retrieval is limited to some extent that the employee can only get the details about the machineries and their functionalities. The employee can also give a simple alert during emergency situations. And it is the duty of the data scientist to sort out the data which is to be used by the employees. This interface consists of a user login. Once after logging in with authorization, the employee can choose the following options. Machine details – To get the details of the machinery (date of its purchase, date of last service, next service date, up time, down time). New user – This helps the new employee to register in and virtually visit a tour about all the departments the company has and the machineries it runs. Emergency alert – Is used at some emergency situations when any damage is noted. This alerts the admin or the manager who is in charge.

6. ARCHITECTURAL DESIGN

The overall design of the project is pictured above. It starts with the data sources which is acquired from the sensors (transducers) fixed in the machines running. These data may be structured, semi structured or unstructured in nature. These are processed under the roof of Hadoop technologies and are sent to the data marts. The data scientists excelled in data mining and big data analytics are employed here. They retrieve the unstructured data from the data marts and process them. This processing uses some algorithms like Naive Bayes and Map Reduce to perform data clustering and classification. This yields the structured data as an output. And finally these data are generated into reports and are transferred to the admin's desk by using MIMOSA standards. At last the company's previous stock values are taken into consideration to analyze company's profit and predict its future holdings in the stock market.
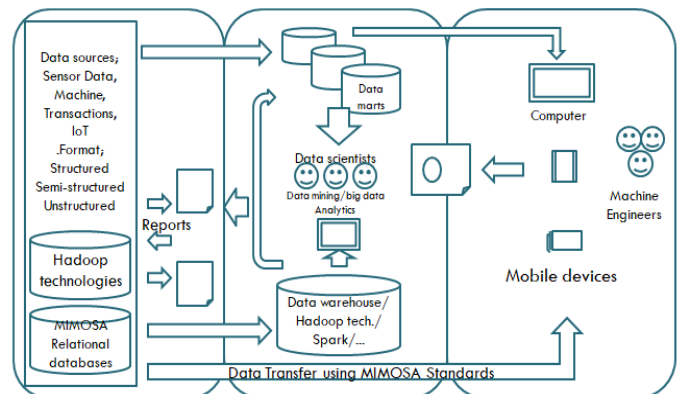
Fig -2: Architectural flow of the project

7. TOOLS INVOLVED

Many tools are available to process big data. Considering the major emerging tools like Map reduce, Apace Spark and Storm. Map reduce and Hadoop are the prevalent software platforms. Map reduce is a process of divide and conquer. The algorithm goes like master and worker node. Hadoop and Map reduce are the best in the field of Big data for solving problems. Apache Spark is for fast processing of data and to carry out analysis in a better way.

8. FUTURE WORK

The data obtained from any field doubles every two years. These data are to be either analyzed or sorted out. All these works efficient with the automated systems. It is not easy to transform data into useful knowledge without the means of parallelism. Before every analysis of data, the values which are missing after the data been fed should be filled or the tuplets have to be removed. This approach leads to loss in data more often. There is a need to balance between these issues. Processing the data at high speed and achieving high performance, throughput at the same time is another issue. Along with this, efficient programming is necessary. There is a need for expressing data access requirements and abstracting programming languages.

9. CONCLUSION

Hereby the system simply starts with the uploading of data acquired from the sensors and organizing them in the Hadoop environment. The data mining technique is important for this asset management since it sorts large data sets into structured data. The data scientists are involved to give a clear visualization for the employee's need of data. The system overall gives an improvising method for achieving profit in the company through stock market analysis and prediction. Hence the current work gives an understanding for the organizations that are in the process of implementing big data solutions. The 4V's and big data technologies are essential here. For a company to increase its business value, it has to convert the data into information and knowledge. The big data field analysis usually involves statistical analysis which is much used for the stock market analysis, machine learning and data mining. These in combination can give an efficient system. Combining the techniques for processing, storing and retrieving in the best way will yield path for emerging big data researches.

10. REFERENCES

[1] Jaime Campos, Pankaj Sharma, Unai Gorostegui Gabiria, Erkki Jantunen, DavidBagleee, "A Big Data Analytical Architecture for the Asset Management," in 9th CIRP IPSS Conference: Circular Perspectives on PSS, vol. 64, Denmark, 2017, pp. 369-374.

[2] X.Jin, B.W.Wah, X.Cheng, Y.Wang, "Significance and challenges of big data research," Big Data Research, vol. 2(2), pp. 59-64, 2015.

[3] M.K.Kakhani, S.Kakhani, S.R.Biradar, "Research issues in big data analytics," International Journal of Application or Innovation in Engineering and Management, vol. 2(8), pp. 228-232, 2015.

[4] O.Y.Al-Jarrah, P.D.Yoo, S.Muhaidat, G.K. Karagiannidis, K.Taha, "Efficient Machine learning for big data : A review," Big Data Research, vol. 2(3), pp. 87-93, 2015.

[5] A.Jacobs, "Pathologies of big data," Communications of the ACM, vol. 52(8), pp. 36-44, 2009.

[6] J.F.Peters, "Near Sets. General theory about nearness of objects," Applied Mathematical Sciences, vol. 1(53), pp. 2609-2629, 2007.

[7] D.Molodtsov, "Soft set theory first results," Computers and Mathematics with Applications, vol. 37(4/5), pp. 19-31, 1999.

[8] Z.Huang, "A fast clustering algorithm to cluster very large categorical data sets in data mining," SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 1997.

[9] Z.Pawlak, "Rough Sets," International Journal of Computer Information Sciences, vol. 11, pp. 341-356, 1982.