JOiV

Big Data and Shipping-managing Vessel Performance

Mandeep Virk[#], Vaishali Chauhan[#]

[#] *Department of Computer Science Engineering, Chandigarh University Gharuan ,S.A.S Nagar Mohali, Punjab, India*
E-mail: mannuvirk.virk27@gmail.com, vaishali.cse@cumail.in

Abstract— **Shipping business is staggering the trade by a substantial number which portrays the usage of leading technologies to deliver formative and reliable performance to deal with the increasing demand. Technologies like AIS, machine learning, and IoT are making a shift in shipping industry by introducing robots and more sensor equipped devices. The hitch big data originates as a technology which is proficient for assembling and transforming the colossal and divergent figures of data providing organizations with meaningful insights for better decision-making. The size of data is increasing at a higher rate because of the procreation of peripatitic gadgets and sensors attached. Big data is accustomed to delineate technologies and techniques which are used to store, manage, distribute and analyze huge data sheets with a high rate of data occurrence. This gigantic data is allowing to terminate the business by developing meaningful and valuable insights by processing the data. Hadoop is the fundamental basic for composing big data and furnishes with convenient judgments through analysis. It enables the processing of large sets of data by providing a higher degree of fault-tolerance. Parallelism is adapted to process big size of data in the efficient and inexpensive way. Contending massive bulk of data is a determined and vigorous assignment that needs an enormous crunching armature to guaranty affluent data processing and analysis.**

Keywords— **big data, Hadoop, big data analytics.**

I. INTRODUCTION

With the onset of automation and machinery the cruise freighting industry (shipping) has also jumped into the era of digitization where ships are equipped with different types of sensors to auditor climate situations, vessel movements and much more. Big data in shipping is playing a frolicking role where valuable insights are generated after analyzing a huge extent of data collected from sensors/satellites. Big data analytics is helping the industry by providing better explanations for regulating marine traffic, fuel efficiency, vessel safety and security, energy management and optimization etc.

The intent of this investigation is to enforce an inclusive exploration on the use of big data in monitoring and analyzing vessel performance in shipping. The appearance of big data in shipping is also discussed. Moreover, several challenges corresponding to gigantic amount of data in shipping are discussed in this research. Few case studies that have been operated in the shipping sector are also redefined in the paper. Further, the research paper is arranged in the order such as: Section II presents the big data and its discrimination. The purpose of big data/big data analytics in Shipping is explained in Section III. Hadoop solution to the problem is explained in Section IV. Case studies that have been operated in the shipping sector are redefined in Section V. Section VI introduces the challenges. Section VII presents the conclusions.

II. BIG DATA AND ITS DISCRIMINATION

Big data is a hitch that relates to extremely large and varied data volume which is difficult in storing and processing. The conventional database technologies are not competent to handle the increasing load of data. The action of big data is fuzzy which demands substantial ways to distinguish and restate the data into novel and meaningful insights. Several researchers have narrated big data in different manners in the earlier literature. For instance, [2] mentioned big data as the hefty figure of data which can be utilized for visualization in scientific areas. There are different definitions for big data. For example, [3] mentioned big data as the volume of data that is on the far side of technology's capability to handle. Meanwhile, [4] and [5] mentioned big data as described by three Vs: variety, volume, and velocity. These terms (variety, volume, and velocity) were initially presented by Gartner as a starting point to define the reasons for different challenges in big data. IDC also stated big data techniques as "a birth of novel technologies which are designed to collect, process and investigate the gigantic bulk of data " [1] interpreted big data as a methodology that demands substitutive forms of consolidation to unveil values from massive data sets which are complex, multiple and hefty. [6] Defines big data in five

distinct phases (generally referred to 5 Vs) that are variety, volume, velocity, value, veracity.

(1) The volume deals with the expansion of data which is getting generated from several sources and is not easy to handle.
(2) Variety concerns with the distinct structures of data generated from different sources like web, enterprises, customer relations, weather, farm etc [7].
(3) Velocity concerns with the speed at which data transfers. [8].
(4) Value is the indispensible facets of big data; it deals with discovering concealed assessments from extensive datasets [9].
(5) Veracity indicates the deformity of data. The veracity of data sources results in the accuracy of the analysis.

What is a big data problem?
To describe a data problem three factors should be considered:

The problem with gigantic bulk of data depends on three factors: Volume, Velocity, Variety. When the enormous bulk of data structured/unstructured gets generated at a faster rate that becomes unable to handle the traditional systems contributes to a big data problem.

III. BIG DATA IN SHIPPING

To find the motive of big data in shipping, it is necessary to discriminate between big data and lots of data. Traditional techniques such as warehousing and distributed systems have been helping organizations to analyze the enormous data sets to find valuable insights for better decision making. However, big data has taken the industry to the next level. Big data allows the companies to utilize immense bulk of data from business organizations in a more suitable and economical way. It is analyzing the data in both ways i.e. batch/real-time modes. The first major leading task in influencing big data is to discover the problem which requires to get solved and then impose it on necessary data required to solve that problem.

The shipping industry has astonished the world's trade by a massive amount which is around 90% of the total trade. But still, the shipping industry lays a way beyond in digital market comparative to other transport and logistics industries. The shipping industry is generating millions of data points daily from seaports, social media feeds, transferring of vessels and parcels, logistics, couriers and much more. Big data is helping the companies to enhance the potential by gathering the data and analyzing it to find valuable and useful insights. Also, big data is making enhancements to monitor the vessel movements on the deck by capturing the data. It is also helping the shipping business to find shortest routes to deliver vessels.

One of the main problems that the shipping industry is facing is the cause of misplacement of vessels on their route from source to destination. This problem has leads the industry to turn into smarter technologies like big data analytics. The one solution to this problem is a proper experimentation of data captured from the sensors connected to the vessels. This data will help to find the reasons of missing the vessels on their way to the destination by finding the cause that may be weather conditions, theft, improper management, improper positioning, and rate of tides and many more. This can be done by using a proper technical solution like Hadoop that analyze the data to discover results that explain the origin of missing vessels and help with better solutions to manage and increase the performance.

IV. HADOOP SOLUTION

Hadoop framework [11] and MapReduce paradigm [12]. Map Reduce [16] is one of the leading techniques for managing big data in a cloud environment; it provides an environment for large datasets which are stored in the cluster to get processed. The foundation of cloud framework can act as a suitable environment for executing big data analysis by contending the required data storage [1]. The main purpose of Hadoop was the space utilization.[14] Hadoop is an open source framework developed by using JAVA platform. Hadoop is constructed of:
• Hadoop File System (HDFS)
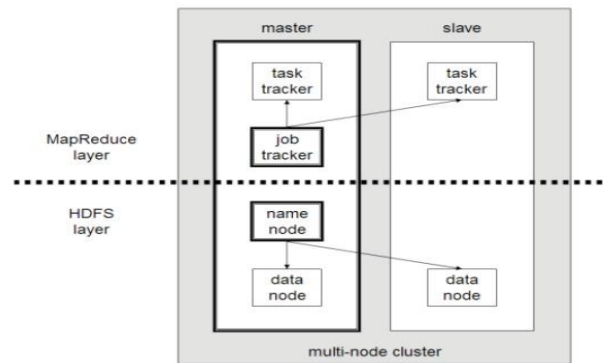• Programming Paradigm (Map Reduce).

Fig.1. Hadoop architecture

Hadoop utilizes a distributed file system to store and manage a large amount of data by providing the fault-tolerance environment. The data is analyzed by making use of technologies like 'HIVE'/pig. A query language 'HIVE' which is used for summarizing data and querying.
The system works as data from various sources is loaded into the Hadoop engine from different sensors attached on vessels placed on the ship then by making use of several queries like hiveQL, pig. Analysis is done to find the proper reason behind the missing of vessels and also the performance is monitored through the use of Hadoop. Fig.2 shows the examination of data in Hadoop cluster.
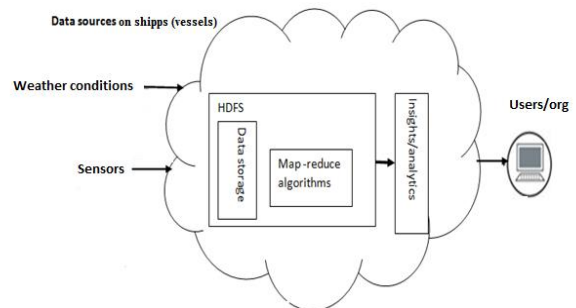

Fig.2. analysis of vessel data

HDFS is a platform-independent file system formed using Java. In this large sized query is divided into fixed size blocks to elevate the data processing rate. It provides well organized data storage. MapReduce is associated with the processing of huge data sets by using different algorithms on the cluster.

The analyzing process of sizeable data sets that get stored in HDFS is supported by technologies like Pig and Hive.

This is some of how big data analytics help in improving performance by minimizing the risk and helping in better decision-making.

V. CASE STUDIES

Discussion on the operation of big data is complemented by reported case studies: marineTraffic, Genscape

MarineTraffic [13]

The organization utilizing big data to identify vessel movements, find destination addresses and speed of transferring vessels. The company uses AIS technology that will enable maritime to introduce itself in the digital market. This helps the company with improved decisions and better analysis.

Genscape [10]

The firm is mastering the shipping industry utilizing big data by immolating with real-time insights resulting in meaningful and efficient decisions. The company provides the industry with most recent data that helps in improved performance.

VI. CHALLENGES

The most prodigious problem with big data in shipping is the availability of gathered data is often poor so it is required to be assured before any decision-making [15]. Lack of predictability is also the significant reason while commanding big data application performance. Data loss while various hardware failures are also addressed. Computational efficiency is also addressed as the major issue as it deals with storage and analyzing data i.e. what is the purpose of storing data if you cannot analyze it in a reasonable time? Analyzing data in reasonable time is a challenge. Cost for commanding big data is also a big issue as it requires a lot of storage and a lot of computational power. So the solution to the problem should be cost-effective.

VII. Conclusion

Technology has remade shipping multifariously by emerging different techniques in this sector. The expansion of sensors and wireless devices are continuously increasing the data. Demand is propagating for analytics that help the business to explore, visualize different type of data discovering hidden opportunities. New opportunities can be acquired from this data for shipping/business by gaining real-time insights. This study has contributed a proposal for improving the performance by finding major reasons of risk. This will help in increasing business and reducing shipping costs. Also for the future optimization of vessels in this industry, more sophisticated tools/applications need to be developed.

REFERENCES

[1] [1] Hashem, I.A.T. et al., 2014. The rise of "big data" on cloud computing: Review and open research issues. Information Systems, 47, pp.98–115.

[2] M. Cox, D.Ellsworth, Managing Big Data for Scientific Visualization, ACM Siggraph, MRJ/NASA Ames Research Center, 1997.

[3] J.Manyika, M.Chui, B.Brown, J.Bughin, R.Dobbs, C.Roxburgh, A.H. Byers, Bigdata: The next frontier for innovation, competition, and productivity, (2011).

[4] P.Zikopoulos, K.Parasuraman, T.Deutsch, J.Giles, D.Corrigan, Harness the Power of BigData The IBM BigData Platform, McGraw-Hill Professional, 2012.

[5] J.J. Berman, Introduction, in PrinciplesofBigData, Morgan Kaufmann, Boston, 2013,xix–xxvi (pp).

[6] Sakr, S. & Gaber, M.M., 2014. Large Scale and big data: Processing and Management Auerbach, ed.

[7] D.E. O'Leary, Artificialintelligenceandbigdata, IEEEIntell.Syst.28 (2013)96–99.

[8] J.J. Berman, Introduction, in PrinciplesofBigData, Morgan Kaufmann, Boston, 2013, xix–xxvi (pp).

[9] M.Chen,S.Mao,Y.Liu,Bigdata:asurvey,Mob.Netw.Appl.19(2) (2014)1–39.

[10] https://www.genscape.com/solutions/maritime-freight/maritime-services-overview#tabs-About-Genscape-Maritime-Services_panel

[11] J. Gantz, D.Reinsel, Extracting value from chaos, IDCi View (2011) 1–12.

[12] Hadoop Distributed File System: Architecture and Design http://hadoop.apache.org/common/docs/r0.18.2

[13] http://www.marinetraffic.com/en/ais/home/centerx:-12.1/centery:25.0/zoom:4

[14] Kaisler, S., Armour, F., Espinosa, J. A., Money, W. (2013). Big Data: Issues and Challenges Moving Forward. International Confrence on System Sciences (pp. 995-1004). Hawaii: IEEE Computer Soceity.

[15] Lesser, A., 2014. Big Data and Big Agriculture. Gigaom Research, p. 11.

[16] J. Dean, S. Ghemawat, MapReduce: Simplified data processing on large clusters, Commun. ACM51 (2008)107–113.