

Prominence of Big data and its related issues, challenges and applications

R.GOKILABHARATHI,

M.Tech (Network),
Kalasalingam University, Krishnankovil,Tamilnadu.
E-mail id:r.gokilabharathi@gmail.com

K.GURULAKSHMI

M.Tech (Network),
Kalasalingam University, Krishnankovil,Tamilnadu.
E-mail id:guruoct19@gmail.com

HATIM GASMEL SEED

M.Tech (Network),
Kalasalingam University, Krishnankovil,Tamilnadu.
E-mail id:hatim.gasmsabri@gmail.com

K.SOWNDARIYA

M.Tech (Network),
Kalasalingam University, Krishnankovil,Tamilnadu.
E-mail id:sowndariya25@gmail.com

Abstract—Imagine that the data coming fast and furious from many sources is the foundation of your big data strategy. This data is saturated with valuable elements that you're probably not using. Large amount of data will be coming from plenty of resources is the beginning of big data policy. This data is sudden with valuable elements that probably not using often. It is mainly used for analysis and transforming the datasets. It allows to access unstructured format to the distributed file system across thousand's of datasets. Many applications and companies assist to filter and convert the huge number of elements into useful groups of details that can examine those subgroups to produce latest and, more useful details for decision making. Now-a-days big data is all around us including smart phones with its application, websites and companies produce a huge amount of data for analysis and making a critical decision. In spite of current solution that offered for big data, still, it faces many challenges and issues. We focus on discussing the important challenges and issues in big data and how it would be reflect decision making. Many issues, challenges must be taken into consideration for dealing with big data. We explained what the big data is and what are the important challenges and their related issues in big data, starting from capturing the generated data, storing it and analyzing it to be used for specific applications. We also focus how to use the result of analyzed data in specific applications.

Keywords-component; Big Data, Analytics, Hadoop, HDFS(Hadoop Distributed File System)

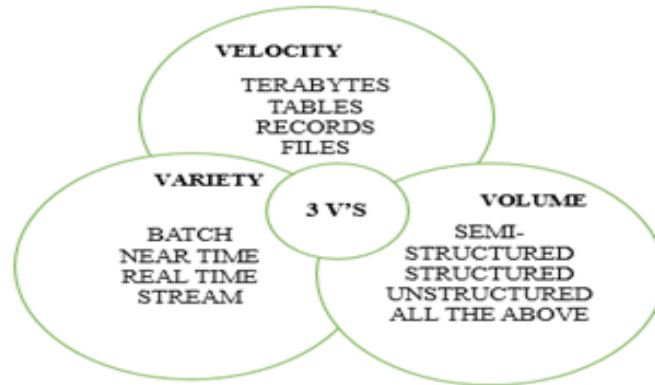
I. INTRODUCTION

Big Data is an data sets that are huge and complex. It is a traditional data processing application. It also defines the use of predictive analytics and user behavior analytics. Big Data is characterized by Velocity, Variety, and Volume.

Velocity: It observes and tracks the object.

Variety: It is available in Real-time.

Volume: It gets the information from texts, images, audio and video.



“Fig.1 “3 V’S” OF BIG DATA”

Big Data increases both the number of data sources and the variety, volume of data is useful for analysis. A non-relational system can be used to produce analytics from big data or to pre-process big data before it is consolidated into a data warehouse.

A. Cloud Storage Capability:

Organizations that want to utilize large sets of data sets should consider third-party cloud service providers, which can provide the storage and the computing power for a specific period.

B. Data Analysis Methods & Capability:

A Big Data initiative is going to require next-level data visualization tools, which present BI data in easy-to-read charts, graphs and slide shows. Text, audio, videos files can also provide valuable in sight. The right tools can even recognize specific pattern based on predefined criteria.

II. BIG DATA ANALYTICS

Big data analytics is used to test the raw data for the motive of getting the closure about that information. It involves applying an algorithmic or mechanical process to derive unknown perceptions. Big data analytics involved in the process of gathering, classifying and examining large sets of data to locate patterns and other helpful details. It will aid organizations to better understand the information carry within the data and will also identify the data that is most important to the business and future predictions.

III. USE CASES FOR BIG DATA ANALYTICS



Fig.2 Big data analytics

A. *The following four types of big data analytics:*

1. Descriptive – It denotes the level of what is going on now based on incoming data.
2. Diagnostic – A view to determine what happened and why in the past performance. The result will be often an analytic dashboard.
3. Predictive – An analysis of possible framework of what might occur. The deliverables are usually a predictive forecast.
4. Prescriptive – An analysis of recommends what actions should be taken and it focuses in rules and recommendations for next steps.

IV. HADOOP

Hadoop is an Apache Open source framework. It supports the processing and storage of large amount of datas.It allows to use multiple distributed processing of huge datasets across group of systems.By using basic programming models. The Hadoop framework applications works in an environment.It also provides distributed storage across clusters of systems. From single server to thousands of machine each one offering local computation and storage.

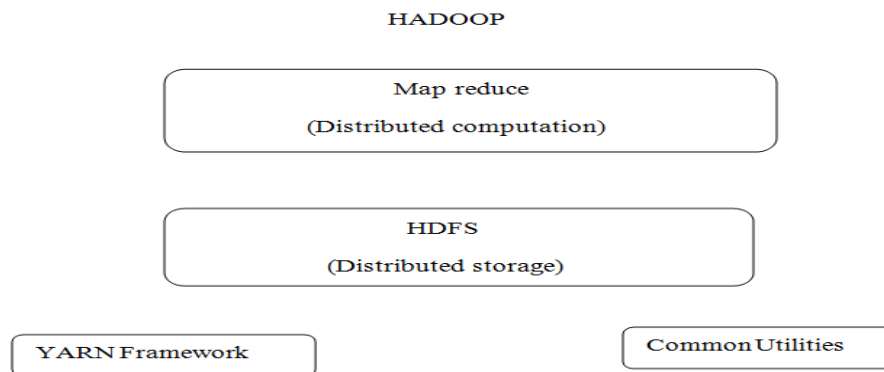


Fig. 3 Hadoop Architecture

A. *Hadoop Distributed File System:*

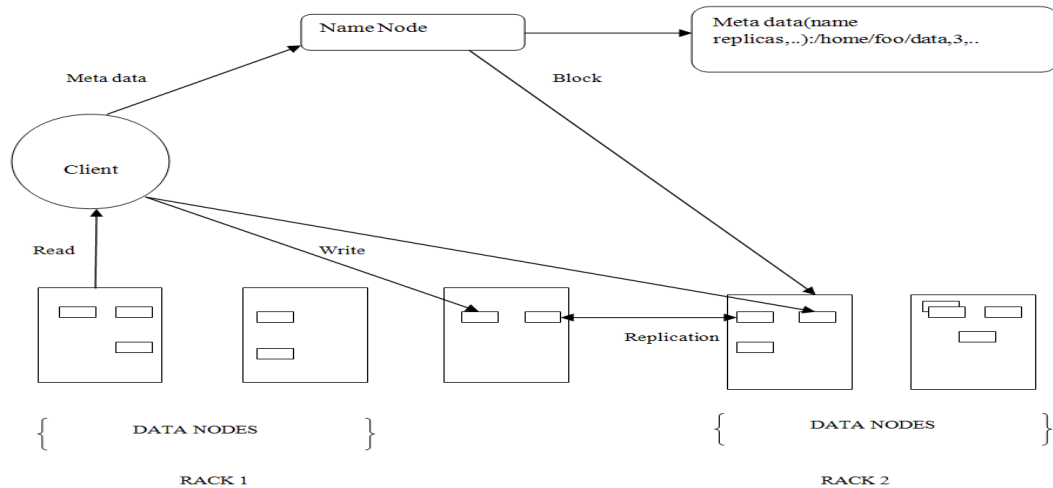
It is the primary storage systems.Hadoop file system was developed using distributed file system. Hadoop creates a common interface to communicate with HDFS. HDFS provides authentication,authorization and file permissions. It runs on commodity hardware like GNU/Linux operating system.There is a datanode in HDFS. Each cluster having having each node. On that node,it manages,stores the datanode into their system. Datanodes done read-write operations from client request. It also performs block, creation, deletion and replication.

B. *Goals of HDFS:*

a) *Fault detection and recovery:* HDFS runs on hardware.It consists of large number of nodes,sometimes it occurs some errors. HDFS helps to find the fault detection and try to recovery the error as soon as possible.

b) *Huge datasets:* In each cluster,it have hundred's of nodes.It is helpful to arrange and managing the application because of huge amount of datasets.

c) *Hardware at data:* A requested task can be done efficiently,when the computation takes place near the data. Due to the large amount of datasets are involved, it reduces the network traffic and increases the throughput.



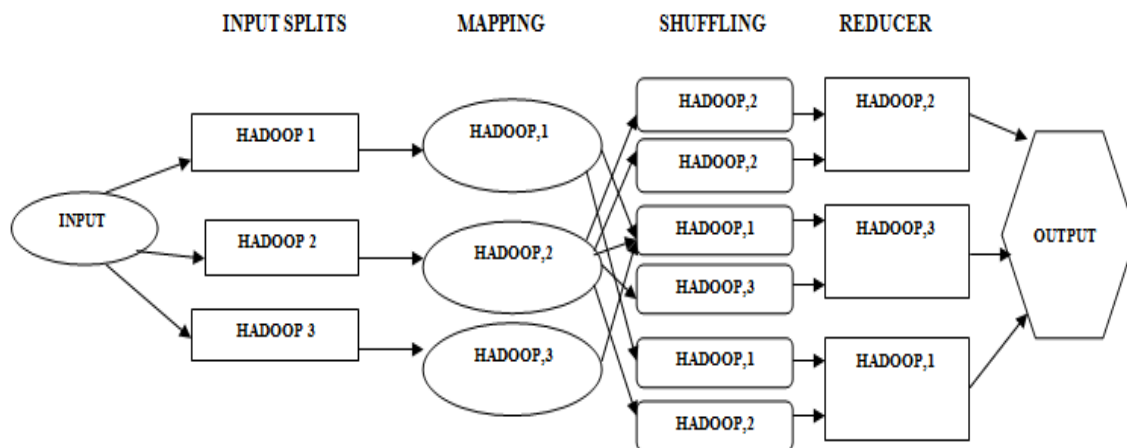
“Fig. 4 HDFS Architecture”

V. MAP REDUCE

Map Reduce splits the input data into small chunks and process it parallel and finally produce a shuffled and reduced output. It consists of single master and slave.

MASTER's work is Job Tracker

SLAVE's work is Task Tracker per Node



“Fig.5 Mapreduce Architecture”

VI. CHALLENGES IN BIG DATA

A. Heterogeneity:

Big data related to the huge and a large amount of resources; this means a different type of data captured and collected for analysis process, the algorithms that used for big data analysis process always dealing with homogenous data and cannot understand heterogeneity. For example, the audio and video data with normal data. This leads to sort and construct the data before data analysis carefully.

B. Size Of The Data:

The name itself “Big Data”, the volume of the data represent a major challenge since the data are growing rapidly. The process of managing large size of data requires an effective solution to handle this challenge.

C. Processing Time:

Big data processing takes more time than normal data processing. The successful system means to provide the data to the user correctly and on time. With a big data, it's difficult to handle a large amount of data in term of processing time.

D. Sharing of data:

Data sharing would support the availability of the information in public, but users refuse to share their big data for competitive purpose. Suppose that each user share their own big data more accurate result can be obtained in order design new and common standardization, to handle each data imported from different sources and platforms.

E. Infrastructure Fault:

Big data storing and analyzing requires complex and high-speed hardware infrastructure. Large and complex data means large and complex hardware infrastructure needed. Hardware reliability to the certain period of time, for this reason, system, might fail at any time. To avoid these matter companies uses a backup system solution for storing all data to achieve continuity, in addition to that company starts to use software solution for hardware backup by directing traffic to another system in case of hardware failure.

F. Big Data Output Displaying:

Analyzing big data becomes very complex when dealing with huge amount of information as well as a variety of categories of information obtained. These leads to display all these data in good and useful manner as a large amount of data cannot be understood without efficient sorting of it. To overcome this, big data must be grouped or clustered into hierarchy approach starting from higher level view. By doing this solution data can be visualized and display output effectively.

G. Privacy:

One of the critical challenges that must be carefully managed is privacy. Big data stores different types of information including public, private, and personal information. In addition to no one like to share their personal information like Password, Credit card number, health status report...etc. to avoid misuse of this valuable information.

For this reason, big data must find an efficient solution to protect user privacy, especially for personal information it should provide multi lines of defense regarding any expecting attack or misuse of user personal information.

VII. ISSUES IN BIG DATA

While capturing the cost, it's need too fast and also difficult to capture 100%. There are '3' fundamental issues.

Storages Issues

Management Issues

Processing Issues

A. Storage issues:

If the section data wants to store large amount of data but it suffers from storage medium. E.g.: To store the data and convert 1 Hexa-byte would require 25,000 disks. The Hexa-byte of data could be processed on single computer system. It means, unable to directory attach the requisite no of disks.

To solve the issue

- a) Data should transmit only result information or transfer the data by using coding technology.
- b) It can transfer the Meta data along with actual data.

M. Management issues:

Management has difficult to address with Big Data. To some these issues of access Meta data, utilization, updating, governance and prove to major stumbling blocks.

B. Processing issues:

Assume the data is chunked into blocks of 8-words. So one Hexa byte = one K peta byte. One block at 5- Giga hertz required for end-to-end processing would be 20 Nano sec. Thus effective process of Hexa byte of data will require extensive parallel processing and also analytics algorithm to provide timely and actionable information.

VIII. APPLICATIONS OF BIG DATA

A. Communications, media and entertainment:

Many organizations, industries require Big Data Support. In Wimbledon Championship (You Tube Video) Big Data will take full advantage to bring analysis on the tennis, cricket, reality shows, and movies in Television, and Mobile in real time. A music service is served on-demand, used to collect all relevant data from its billions of users in real time worldwide. These data will be analyzed and informed music recommendations to individual users.

B. Transportation:

As in present days, large amount of information (data) from local-area networks, excessive speed data have most affected. So Government uses Big Data to control network traffic, Route planning, intelligent transport systems, congestion management. Private sectors also use Big Data for revenue management, technological enhancements, logistics etc.

C. Government:

In government the huge Summons are the unification and interoperability. Big data is used to collect the huge amount of social disability claims. It changes from unstructured data format to correct structured data format. By using Big Data information the decision making can be done quickly and effectively and also used to predict the deceit claims. THE FOOD AND DRUG ADMINISTRATION (FDA) is applying Big Data to find food related complaints. It detects response faster which aims to treatment and reduces death.

D. Banking and Securities:

The Security Exchange Commission (SEC) is applying Big Data to observe the activity of financial market. It is helpful to catch the illegal trading activity. High frequency trading, pre-trading decision support analytics, sentiment measurement, predictive analysis are also use Big Data.

E. Education :

By using Big Data, students have deployed a learning and management systems. Using this scenario when the student or faculty logs on to the their portal or system, it will show the details about their working hours and pages.

TABLE I. Applications of Big data

BANKING/FINANCIAL	RETAIL	HEALTHCARE	TRANSPORTATION
EDUCATION	ENTERTAINMENT	LIFE SCIENCES	CHEMICALS
GOVERNMENT	AUTOMOTIVE	AEROSPACE	ENERGY&UTILITIES

IX. CONCLUSION

Big Data from its beginning until its current state, it expands the abstraction of big data come behind the applications and the issues faced by it. The main aim of this paper is to explore the role of Big Data in various fields. Big data used in various applications including banking, agriculture, government sector, banking and securities, chemistry, data mining, transportation, cloud computing, finance, marketing, education, stocks, BDA, health care etc. Big data is efficient because which will make easy the things to work.

References

- [1] Arash Baratloo, Mehmet Karaul, Zvi Kedem, and PeterWyckoff. Charlotte: Metacomputing on the web. In Proceedings of the 9th International Conference on Parallel and Distributed Computing Systems, 1996
- [2] Baker, Malcom and Jeffrey Wurgler,2007. "Investor Sentiment in the stock market", Journal of Economic Perspectives, vol.21(2),pages 129-152.
- [3] "Big Data: Trends, Strategies and SAP Technology" by Carl W.Olofson, Dan Vesset August 2012.
- [4] "Big Data is Future of Health care" by Cognizant 20-20 insights September 2012.
- [5] "Brandon Bunker, Senior Director of customer Analytics and intelligence, Vivint, "Big Data insights Platforms for Rapid Data Discovery", 2013.
- [6] Davenport, T.H. and J.G. Harris (2007) Competing on Analytics: The New Science of Winning, Boston:Harvard Business School Press.
- [7] David Loshin, "Addressing Five Emerging Challenges of Big Data", Progress Software, 2014.
- [8] "Data-driven healthcare organizations use big data analytics for big gains" by IBM software.
- [9] "Deep learning applications and challenges in big data analytics" by Maryam M Najafabadi, Flavio Villanustre,Taghi M K hoshgtaar, Naeem Seliya, Randall Wald and Edin Muharemagic.
- [10] D.Fisher, R.DeLine, M.Czerwinski and S.Drucker, "Interactions with Big Data Analytics," interactions, vol.19, no. 3, pp.55-59, May 2012.
- [11] Eckerson, W. (2004) "Gauge Your Data Warehousing Maturity," DM Review (14)11, pp. 34.
- [12] Eckerson, W. (2011) "Big Data Analytics: Profiling the Use of Analytical Platforms in User Organizations,"TDWI, September.
- [13] Formerly Booz Company, Benefiting from Big Data: A new approach for the telecom industry",in 2013.
- [14] Jefry Dean and Sanjay Ghemwat, MapReduce:A Flexible Data Processing Tool, Communications of the ACM, Volume 53, Issue.1,January 2010, pp 72-77.
- [15] L. Huston, R. Sukthankar, R.Wickremesinghe, M. Satyanarayanan,G. R. Ganger, E. Riedel, and A. Ailamaki. Diamond:A storage architecture for early discard in interactivesearch. In Proceedings of the 2004 USENIX File and Storage Technologies FAST Conference, April 2004.

- [16] MarcinJedyk, MAKING BIG DATA, SMALL, Using distributed systems for processing, analysing and managing large huge data sets, Software Professional's Network, Cheshire Data systems Ltd.
- [17] Peter Groves, Basel Kayyali, David Knott, Steve Van Kuiken, "The big data revolution in health care," enter for US Health System Reform Business Technology Office, published in January 2013.
- [18] Puneet Singh Duggal, Sanchita Paul ,— Big Data Analysis: Challenges and Solutionsl, International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV
- [19] Remzi H. Arpaci-Dusseau, Eric Anderson, Noah Treuhaft, David E. Culler, Joseph M. Hellerstein, David Patterson, and Kathy Yelick. Cluster I/O with River: Making the fast case common. In Proceedings of the Sixth Workshop on Input/Output in Parallel and Distributed Systems (IOPADS '99), pages 10.22, Atlanta, Georgia, May 1999.
- [20] Thomas H. Davenport, Jill Dyche, "Big Data in Big Companies," in International Institute for Analytics May 2013.
- [21] Victor L. Voydock and Stephen T. Kent. Security mechanisms in high-level network protocols. ACM Comput. Surv 15(2):135–171, 1983.
- [22] William Gropp, Ewing Lusk, and Anthony Skjellum. Using MPI: Portable Parallel Programming with the Message-Passing Interface. MIT Press, Cambridge, MA, 1999.

AUTHORS PROFILE

R.Gokilabharathi completed her B.Tech Computer Science and Engineering in the year 2015 from Parisutham Institute of Technology and Science, Thanjavur. She pursuing her M.Tech in Kalasalingam University, Krishnankovil. Her area of interest in Networks.

K.Gurulakshmi completed her B.Tech Computer Science and Engineering in the year 2015 from Kalasalingam University, Krishnankoil. She pursuing her M.Tech in Kalasalingam University, Krishnankovil. Her area of interest in Internet of Things and Networks.

Hatim Gasm Elseed completed his B.S (Honors) in Network Management in the year 2015 from Sudan International University, Khartoum-Sudan. He pursuing his M.Tech in Kalasalingam University, Krishnankovil. His area of interest in Network Security.

K.Sowndariya completed her B.Tech Computer Science and Engineering in the year 2016 from Theni Kammavar Sangam college of Technology, Theni. She pursuing her M.Tech in Kalasalingam University, Krishnankovil. Her area of interest in Data Structure and Networks.