A Study On Big Data Analytics

K.Chandra Sekhar[1] ,J.Savithri Sowmya[2],V.Umadevi[3].

[1]Assistant Professor, Department of CSE, Pragati Engineering College,Surampalem, India,
[2]Department of CSE, Pragati Engineering College,Surampalem, India,
[3]Department of CSE, Pragati Engineering College,Surampalem, India.
Corresponding Author: K.Chandra Sekhar

ABSTRACT: Big data is a new technology of the world economic and social changes. The world's data collection turns into major technological changes that can acquire new ways in decision making, managing our health, cities, finance and education. data complexities are expanding including data's volume, variety, velocity and variability, the real impact hinges on our capacity to reveal the 'value' in the data is through Big Data Analytics technologies. Big Data Analytics represents a grand challenge on the outline of highly scalable algorithms and systems to merge the data and show large hidden values from datasets that are complex and of a gigantic scale. Potential breakthroughs include new algorithms, methodologies, systems and applications in Big Data Analytics that find helpful and concealed information from the Big Data efficiently and successfully.
KEYWORDS: Complexdatasets, Googlefilesystem, Hadoop, LargeDatasets, low-cost commodity hardware, Map Reduce, Parallel DBMS.

I.   INTRODUCTION

Have you ever wondered what it is Big data is about and how it can useful to us? Here we are going to learn the fundamental concepts of Big data analytics. Big data analytics can be called as a procedure of inspecting expansive large data sets. We use foremost analytics techniques against the expansive information to reveal the concealed examples, client inclinations and other helpful data. This encourages the associations to settle on educated choices.

1.1 Values of Big Data:

" Figure 1. Values of Big Data".

1.1.1 Cost reduction:
Big data innovations, for example, Hadoop bring huge cost points of interest when it comes in putting away gigantic measures of data – in addition to they can distinguish more effective methods for working together.

1.1.2 Faster, better decision making:
With the speed of Hadoop and in-memory analytics, Integrated with the capacity to dissect new wellsprings of information, businesses that can break down data quickly and settle on choices in light of what they have realized.

1.1.3 New products and services:
With the capacity to check user necessities and fulfillment through analytics comes the ability to give client what they need. organizations are making new items to address clients' issues are fulfilled through Big data analytics.

II. CHARACTERISTICS OF BIG DATA:

2.1 Volume:
'Big Data' is related to a size which is large. Size of data plays very important role in determining value out of data. Also, if a particular data can actually be considered as a Big Data or not, depends upon volume of data. Hence, 'Volume' is one of the characteristic which needs to be considered while dealing with 'Big Data'.

2.2 Variety:
Variety in 'Big Data Refers to heterogeneous sources and the nature of data, both structured and unstructured. In Olden days, spreadsheets and databases were the only sources of data considered by most of the applications In any case, Now a days, information as messages, photographs, recordings, checking gadgets, PDFs, sound, and so on. is also being examined in the analysis applications.

2.3 Velocity:
Velocity in 'Big Data' refers to the speed of generation of data. How fast the data is generated and processed to meet the demands, decides real potential in the data. Velocity in 'Big Data' deals with the speed at which data flows from sources like business processes, networks and social media sites, sensors, Mobile devices, etc.

2.4 Variability:
Variability in 'Big Data Refers to the irregularity which can be appeared by the information now and again, thus hold up the process of being able to handle and manage the data effectively.
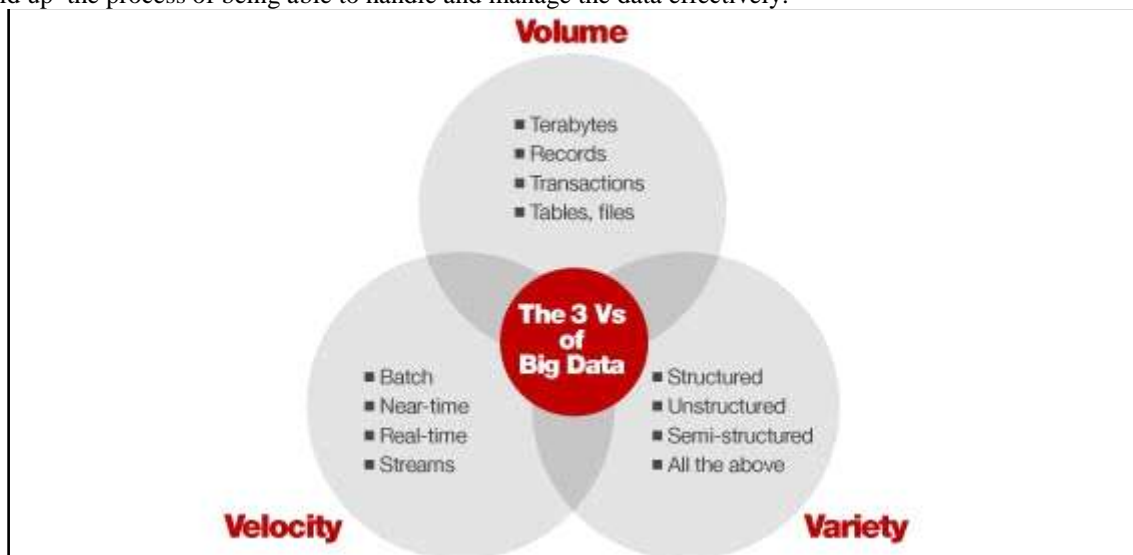

"**Figure 2.**Characteristics"

III. ARCHITECTURE

Big data architecture is designed to handle the processing and analysis of data which is too large or difficult for traditional database systems. Organizations enter into the enormous information real may contrasts, contingent upon the abilities of the clients and their devices.For some, it means hundreds of gigabytes of data, while for others it means hundreds of terabytes. One of the important one is multiple-layer architecture to address the issues that big data presents. Parallel sort of executions conditions can drastically enhance information handling

speeds. This kind of designs embeds information into a parallel DBMS, which actualizes the utilization of MapReduce and Hadoop Frameworks.
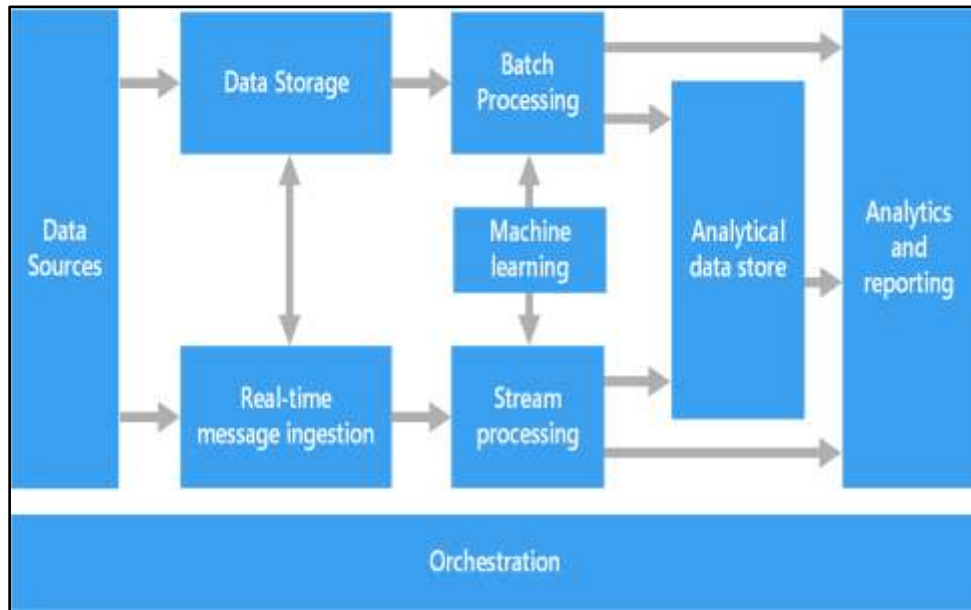
"Figure 3. Architecture"

IV. MANUFACTURING OF BIG DATA

Big data analytics holds importance because of its ability to provide actionable insights and enhance decision-making for an organization. Big data use cases highlight the growing importance of technology across various processes involved in manufacturing. Predictive manufacturing is one of the relevant methodology toward close to zero downtime and straightforwardness requires tremendous measure of data and forecast prediction tools for an orderly procedure of information into helpful data. predictive . starts with information obtaining where distinctive kind of tactile information is accessible to secure, for example, vibration, weight, current, voltage and controller information. Enormous measure of tangible information and recorded information build the huge information in assembling.
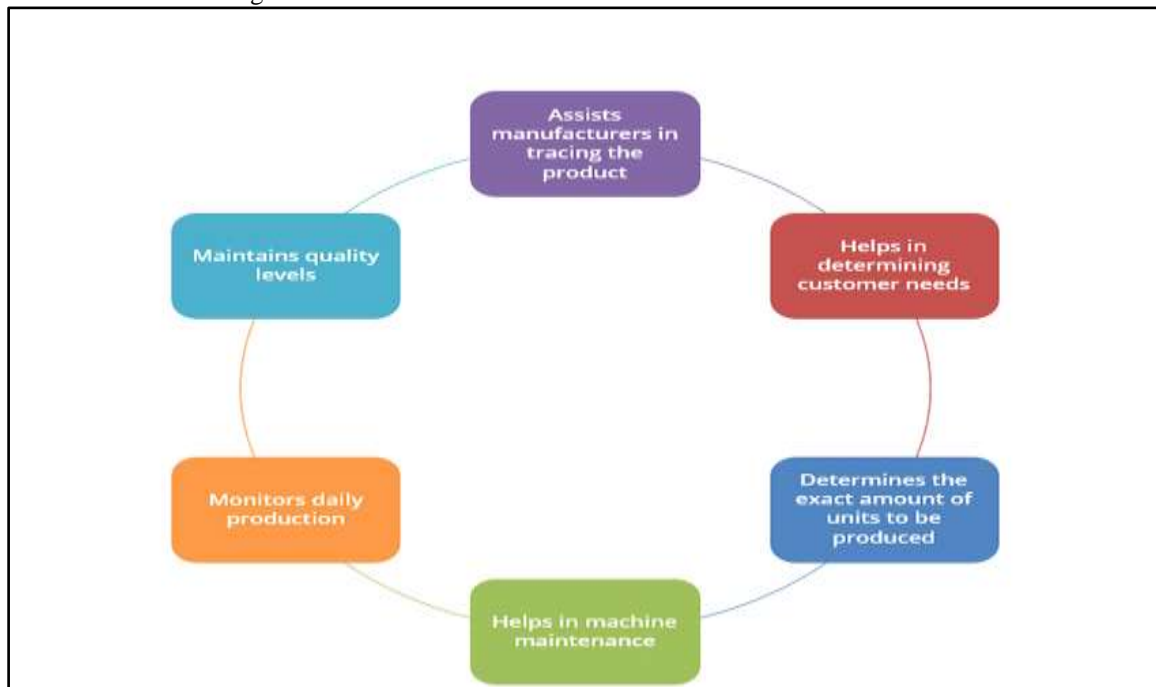
" Figure 4..Manufacturing Cycle"

V. APPLICATIONS

Big data has enlarged the demand of information management specialists in that Oracle Corporation, IBM, Microsoft, SAP, EMC, HP and Dell have spent more than $15 billion on software firms specializing in data management and analytics. In 2010, this industry was worth more than $100 billion and was growing at almost 10 percent a year which is twice as fast as the software business as a whole. Big Data is taking the world by storm. With the vast amounts of data getting from various digital sources and the importance of analytics has immense grown making the companies to tap the dark data that was considered useless all these years. big data has extended across the industries at a swift pace.
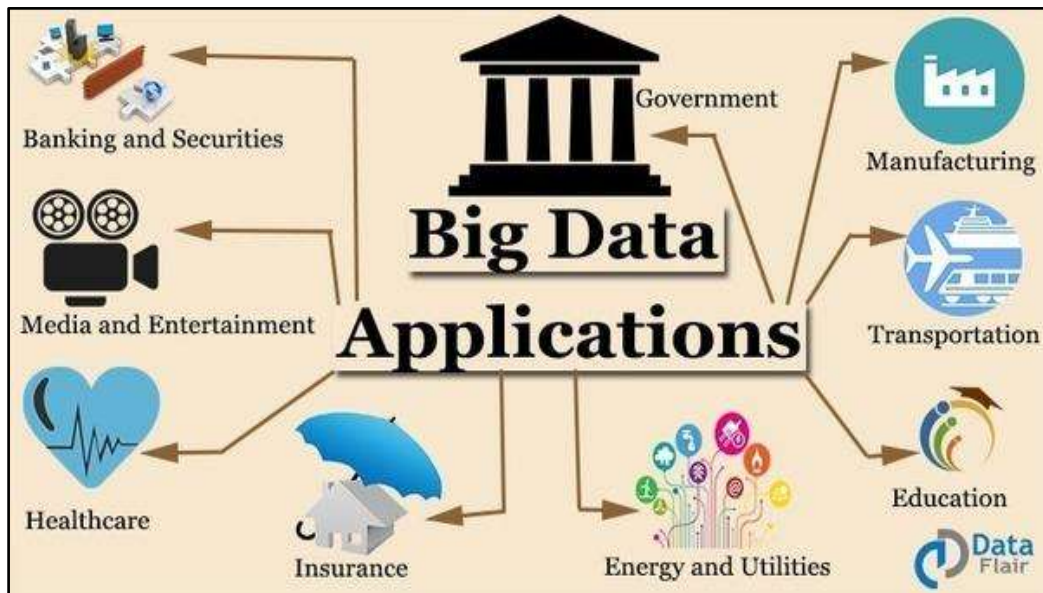

" **Figure 6.**Applications".

VI. CHALLENGES

Big data comes with a lot of opportunity to deal inhealth,education, earth, and businesses But to deal with the data having large volume using traditional models becomes very difficult. So we had big data challenges and to design some computing models for efficient analysis of data.

6.1 Heterogeneity and Incompleteness:
If we want to analyze the data,it should be structured but when we deal with the Big Data, data may be structured or unstructured as well. Heterogeneity is the big challenge in data Analysis And analysts need to cope with it.
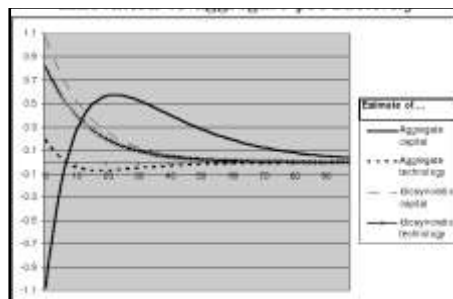
" **Figure 7.** Heterogeneity"

6.2 Scale:
As the name says Big Data is having large size of data sets. Managing with large data sets is a big problem from decades. Earlier, this problem was solved by the processors getting faster but now data volumes are becoming huge and processors are static.
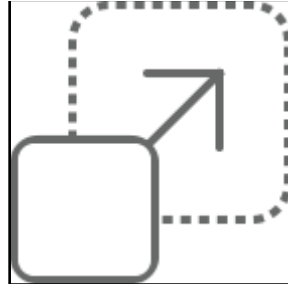
" **Figure 8.**Scale"

6.3 Timeliness:

Another challenge with size is speed. If the data sets are large in size, longer the time it will take to analyze it. Any system which handles effectively with the size performs well in term of speed.
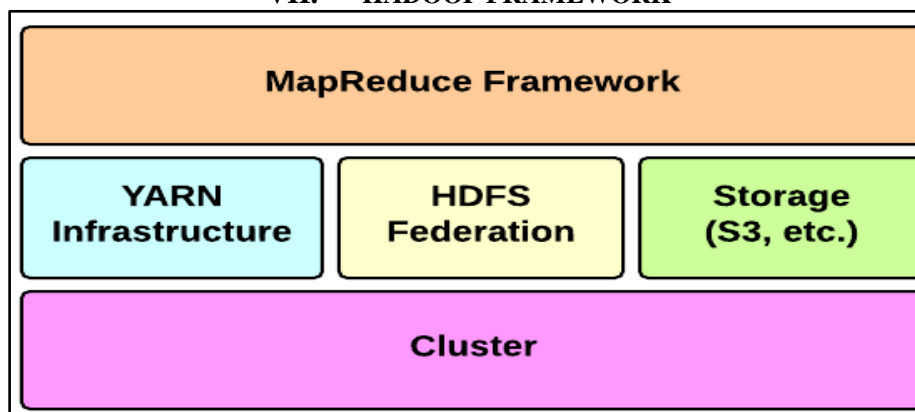

" **Figure 9.** Timeliness".

6.4 Privacy:

Data Privacy is another big problem with big data. In some countries there are strict challenges regarding the data privacy.Some of them are User authentication,Restricting access based on a user's need, Proper use encryption.

" **Figure 10.**Privacy".

VII.   HADOOP FRAMEWORK

" **Figure 11**.Hadoop",

Hadoop is open source software used to process the Big Data. It is very popular used by researchers to analyze the Big Data. Hadoop Is influenced by Google's architecture, Google File System And MapReduce. Hadoop processes the large data sets in distributed computing environment.

Hadoop consists of two main components:
7.1 Storage:
7.1.1 The Hadoop Distributed File System(HDFS):
It is a distributed file system which provides fault tolerance and designed to run on commodity hardware. HDFS provides a very high throughput access to application data and is suitable for applications that have large data sets. HDFS can store the data across thousands of servers.

7.2 Processing:
7.2.1 MapReduce:
It is a programming model introduced by Google in 2004 for easily writing applications which process large amount of data in parallel on large clusters of hardware in fault tolerant manner.

Two functions in MapReduce are as following:
7.2.1.1 Map
The Map function always runs firstly to filter, transform, or parse the data.
The output from Map becomes the input to Reduce.
7.2.1.2 Reduce
The Reduce function is optional normally used to summarize data from the Map Function

VIII. CONCLUSION

In this paper, an overview is provided on Big Data, Hadoop and applications in Data Mining. An overview to big data challenges is given and various applications of big data has been discussed. This paper specifies the Hadoop Framework and its components HDFS and Mapreduce. The Hadoop Distributed File System (HDFS) is a distributed file system structured to run on commodity hardware. Hadoop plays an important role in Big Data. This paper also focuses on The availability of Big Data, low-cost commodity hardware, and new information management and analytic software have produced a unique moment in the history of data analysis. The approach of these trends means that we have the capabilities required to analyze huge data sets quickly and cost-effectively for the first time in history. These are neither theoretical nor trivial.

REFERENCES

[1].   https://www.seminarsonly.com/computer%20science/big-data-seminar-report-ppt-pdf.php
[2].   https://www.quora.com/How-are-big-data-and-machine-learning-related
[3].   https://www.researchgate.net/publication/281403776_Big_Data_And_Hadoo p_A_Review_Paper
[4].   https://dzone.com/articles/the-relationship-between-big-data-and-iot
[5].   https://intellipaat.com/blog/7-big-data-examples-application-of-big-data-in-real-life/
[6].   https://www.guru99.com/what-is-big-data.html#3
[7].   http://www.agiratech.com/introduction-to-big-data-analytics/
[8].   https://www.ugc.edu.hk/doc/eng/rgc/theme/hall/abs1.pdf
[9].   https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/
[10].   https://www.allerin.com/blog/5-big-data-use-cases-in-the-manufacturing-industry
[11].   https://en.wikipedia.org/wiki/Big_data#Manufacturing
[12].   https://www.qubole.com/resources/big-data-challenges