AN INTRODUCTION TO BIGDATA AND ITS APPLICATIONS

K. Banuroopa[1], Dr. M.Mohan Kumar[2]

ABSTRACT

Big data refers to a large amount of data in the range of Exabytes and beyond. Big data refers to the data sets that cannot be handled using the existing traditional database management tools because their capability to store, manage and process huge data efficiently is inadequate. Big data alters the way the data are managed and used, and receive great attention from researchers all over the world. People in academics, industry and in government are equally interested in big data and its technologies. This paper gives an introduction to big data and its application. Big data has so much of applications in various fields such as agriculture, weather forecasting, healthcare, ERP, cloud computing etc. The paper highlights some of these applications.

Keyword : Big Data,Hadoop

I. INTRODUCTION

Big data is high in volume, velocity, and variety. This asset of information requires new ways of processing to get enhanced results, decision-making and process optimization. Although big data doesn't refer to any specific quantity, the term is often used when speaking about Peta bytes and Exabytes of data, much of which cannot be integrated easily. The data it handles will be either structured or unstructured. The volume of the data has increased by 90% in recent years. Data are generated by various sources like web, social networking, weather forecasting etc. As the use of internet is increasing, the amount of Big Data continues to grow. Big Data has become highly important to most of the firms now. According to Big Data Executive Survey, 2016 done by NVP the percentage of firms that see Big Data initiatives as being mission critical has risen from 23.2% in 2014 to 32.1% now. Only 1.8% of firms indicated that Big Data was not important to the firm [11].

The 3Vs that define Big Data are Variety, Velocity and Volume. The 3v's has been represented in fig 1.1 below:

Variety

Data can be stored in multiple formats like audio, video, SMS, documents, etc. It is necessary for an organization to rearrange it and make it meaningful. It will be easy to process the data, if it is in the same format; however, it is not the case most of the time. The data in reality are in many different formats, and to overcome such Big Data is a challenge.

Volume

The data can be found in the format of videos, music files and large images in the social media networks and the storage capacity will be in the range of Terabytes and Petabytes. With the growth of the database the applications and architecture built to support the data

[1,2] Assistant Professor, Department of CS, CA & IT, Karpagam Academy of Higher Education.

have to be re-evaluated. Sometimes the re-evaluation in a different angle produces a new found intelligence, which creates explosion of the data. The big volume indeed represents Big Data.

Velocity

It represents the speed at which the analytics has to be done. The growth of data due to social media explosion has changed the way people look at data. The news channels and radios have changed their way of communicating, and as a result of that, public are receiving news updates instantly. Nowadays people rely on social media to stay updated with the latest happenings. On social media sometimes a message becomes old in a matter of seconds (a tweet, status updates etc.) and users always want to stay updated. The data movement is now almost in real time and the update of data has been reduced to fractions of seconds. This kind of high velocity data represent Big Data.
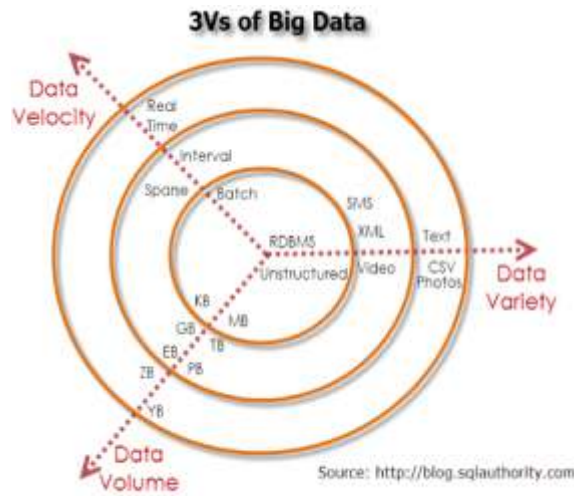

Fig 1.1. 3Vs of Big Data

Table 1: Data Vs Big Data

Category	Traditional Data	Big Data
Sources	Traditional Data sources which includes ERP transaction data Web Transaction Financial Data	Data generated outside enterprise includes Social media, Sensor data, Video, image Device Data
Volume	Gigabytes, Terabytes	Terabytes, Petabytes, Exabytes, Zettabytes
Velocity	Batch or near Real time	Often Real time
Variety	Structured Unstructured	Structured Unstructured Multi -structured
Organiza tion	Centralized	Distributed across Multiple Servers

II TECHNOLOGIES USED IN BIG DATA

Big data can be structured, semi-structured or unstructured. Usually Enterprises store structured data in a traditional RDBMS like Oracle Database, MS SQL Server or DB2, and advanced applications are created to interact, process and analyse data within the database. For a traditional database server, it is really a tedious task to deal with huge amounts of data and process them. However, to store and process big data commendably, we need a new software like Hadoop or MongoDB. Hadoop is used with structured or semi-structured data and MongoDB with unstructured data in the format of word, pdf, text or media logs.

Hadoop was an open source project started by Doug Cutting, Mike Cafarella and team with the help of Google technologies in 2005. MapReduce algorithm is used by Hadoop to run applications in which the data are distributed in various nodes[12]. The open source framework of Hadoop supports the distributed processing of a large amount of data by using a group of distributed nodes. The main components of Hadoop are: 1) Hadoop Distributed File System (HDFS), which allows data to be stored in multiple nodes without having to manage with the complexity of their distributed nature and 2) Map Reduce, which is a programming model aimed to implement distributed and parallel algorithms in a proficient way.

MongoDb is a NOSQL (Not Only SQL) database. It provides features to store both structured and unstructured data. The data in MongoDB is stored in JSON format documents. MongoDB provides the most impressive query language for the availability of data. But the data are not stored in distributed nodes. Its capacity to handle unstructured data and use Map Reduce algorithms makes this the most likely candidate for big data applications. Most of the organizations now tend to use Hadoop and MongoDB together to implement their big data applications.

III BIG DATA USE CASES

A. Big Data is Changing Healthcare

When used in healthcare, the big data can be the patients' data, like notes by doctors, test results from Lab and X-Ray reports, case history and diet regulation. The employees in a particular hospital, expiry date of medicine and instruments used in surgery can also be termed as big data. Big Data can be used in predicting the spread of epidemics, cure diseases or in general, improve the quality of people's lives. It can also be very effective in avoiding preventable deaths of people. The likelihood of the dispersion of Ebola virus in Africa was done using cellular phone's location data, which was used to track the activities of people.

IBM's Watson Health cloud healthcare analytics service gets its data from iPhone and Apple Watch users. Apple and IBM are collaborating with each other in creating a big data health platform. [4]

B. Big Data in Agriculture

The traditional processes, which are used to create successful crop varieties, are costly, time consuming and require lots of labour. Using Big data can speed things up. Agriculture is a complex process which is the combined effort of biology, weather and human effort. The use of GPS and other technologies can help the agriculturist to track produce, control their equipment, observe the environment and manage inputs very accurately across the land. This in turn significantly

increases yield and viability. Big data can be used to test assortments of genetics, yield inputs and conditions across hundreds of fields, soils and climates. The companies using big data can conduct field-plot experiments in real-life situations over thousands of acres of land. This gives them the required facts regarding the kind of seeds to be planted, the kind of soils ideal for it and climatic conditions. [5]

C. Big Data on Government

The availability and the skill level of Big Data information management talent will have a direct impact on the government's ability to successfully deploy Big Data. The new storage strategy like Cloud-based backup, recovery and archiving provide various benefits. On July 1, 2015, the Prime minister of India, Narendra Modi launched the Digital India Movement to guarantee the availability of government services to citizens electronically by enhancing the infrastructure for Internet connectivity. Big Data and analytics will play a major role in such a transformation by gathering data though cloud, mobile technology and other social media, thereby addressing concerns of people and increasing the level of their participation to finish these policies effectively.[4]

D. Big Data in Weather Forecasting

The recent advances in technologies like satellites and other resources, which are used by weather forecasting system, will help people in the precise forecast of weather. The increase in the volume of environmental data put into view the need of Big Data technologies to manage processes and store data. Research scientists at IBM in their long-term weather analysis project Deep Thunder, are heading towards creating the most sophisticated data analytics on weather forecasting.

Apache Hadoop Map Reduce Framework is used to analyse enormous data set of weather forecasting.

E. Big Data in Aligning Corporate System

Big Data has been an ERP profession for the last 14 years. There have been tremendous changes in the ERP system and the implementation methodologies within corporate, both big and small, all across the world. Each corporate has been investing a lot in upgrading their systems. The main aim is for capturing the maximum data, processing it faster and making it available to the right employee for taking the right decision. Data requirements have been so huge that meeting them through conventional ERP modules is very difficult. The ERP software companies have added many extensions and support packs, and newer functionalities have been introduced frequently. In future, incorporating big data in the ERP system will be a big challenge. [8]

F. Big data in cyber security

All major anti-virus and Malware companies, as well as network security vendors, train their systems with a large volume of data that is collected during a security attack. Big data allows enterprises to spot irregular activities in near real-time by amalgamating data from several sources into one large database. Big data technologies help in cyber security by collecting logs and incident data and processing them to determine the suspicious activities very quickly.[9]

G. Cloud computing and Big data

One of the momentous shifts in the recent ICT and enterprise applications which provides a prevailing architecture to perform large-scale computing

applications is cloud computing. Big data is used along with cloud computing to give the users the ability to process distributed queries across multiple datasets and return resultant sets in a timely manner. Distributed storage technology of cloud computing, rather than any local storage device, is used by Big data. Many current cloud-based applications are using various categories of big data. Cloud computing, big data and their applications are the most favourable new cutting edge technology, as Clouds are also being used to effectively store and use big data, which is unstructured data in the organizations. [10]

IV. CHALLENGES WITH BIG DATA

i. Heterogeneity of Data

Heterogeneity in Big Data refers to data which is either structured or unstructured. This is a huge challenge while analysing data for the analysts. It is difficult to integrate data from different sources in various formats to apply a single algorithm on it and draw any noteworthy statistical suggestion [1].

ii. Privacy and Security

Data Security is to protect data from unauthorized access. The fast change in technologies results in trouble in envisaging about the usage of information obtainable on cyber space. The incongruous use of personal data is growing especially when data is extracted from several sources. For example, the location-based services extract data related to location details from subscribers resulting in privacy concerns. The private statistics shared through social medias/online services such as Facebook, Twitter etc. is also a grim privacy breach [1].

Nowadays people tend to share their personal information through social media like Facebook and Twitter. But people don't have an understanding of how this sensitive information, when put together along with location-based services, gives more personal information than intended. Location based services can identify their subscribers even though the subscribers want to hide their identity. By monitoring the location information of people more private information like religious inclinations and health particulars can be gathered.

iii. Skill availability

Big Data is being supported with relatively new technologies and is being considered in diverse methods. Therefore, there subsists a deficiency of people with the ability to bring together the data, scrutinize it and publish the results or suppositions. [1]

V. CONCLUSION

Big data has a solid influence on almost every sector of all industries. This is the period of Big data where practically all applications entail data to provide services swiftly and proficiently. In this paper, we have briefly reviewed the prospects and importance of big data, as well as some the challenges it faces. Hence, Big Data is becoming the new ultimate cutting edge for scientific data research and for business applications. However, managing big data is the greatest task that researchers have to find to do with security. As we have an overwhelming growth of data in terms of volume, velocity and variety, the future of our research lies in building a generic architectural framework towards addressing these issues including security and privacy challenges in a holistic manner.

REFERENCES

[1]   Avita Kata, Mohammad and Wazid R H Goudar, "Big Data: Issues, Challenges, Tools and Good Practices".

[2]   Andrea De Mauro, Marco Greco and Michele Grimaldi, "What is big data? A consensual definition and a review of key research topics" AIP Publishing Vol.97, 2015.

[3]   Xiaolong Jin, Benjamin W.Wah, Xueqi Cheng and Yuanzhuo Wang, "Significance and Challenges of Big Data Research", Elsevier 2015 pp:59-64[

[4]   J.Archenaa and E.A.Mary Anita, "A Survey of Big Data Analytics in Healthcare and Government", Procedia Computer Science, Vol. 50, pp:408 - 413,2015

[5]   Kuchipudi Sravanthi, Tatireddy and Subba Reddy, "Applications of Big data in Various Fields" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5), pp 4629-4632,2015, ISSN:0975-9645

[6]   Swati Harinkhere, Nischol Mishra, Yogendra PS Maravi and Varsha Sharma, "Big Data Techniques & Applications : A Technical Review", CSI Communications Vol. 40,pp 32-35,2016

[7]   S.Rama Sree and K.Devi Priya, "An insight of Big Data Analytics Using Hadoop", CSI Communications Vol. 40, pp 37-39, 2016

[8]   Sanjay Bhatia, "Big Data -Aligning Corporate Systems to support Business Better", CSI Communications, Vol. 40 pp 14-15, 2016

[9]   Cath Everet, "Big data - the future of cyber-security or its latest threat", Computer Fraud & Security, Volume 2015, Issue 9, Pages 14-17, Elsevier.

[10]   Mrs. Premalatha P and Mrs.Marrynal S. Eastaff, "Big Data and Cloud Computing", International Journal of Engineering and Applied Sciences (IJEAS), Volume-2, Issue-11, November 2015.

[11]   https://www.tableau.com/resource/top-10-big-data-trends [12] https://www.tutorialspoint.com/hadoop/hadoop_big_data_overview.htm