

# GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

## RECENT OPPORTUNITIES AND CHALLENGES ON BIG DATA:A TECHNICAL REVIEW

Jitha Janardhanan<sup>\*1</sup> & Reeja Jose Pavanathil<sup>2</sup>

<sup>\*1&2</sup>Assistant Professor,Acharya Institute of Graduate Studies,India.

### ABSTRACT

Plants are among the earth's most useful and beautiful products of nature. Plants have been crucial to mankind's Data can be defined as a representation of facts or concepts in a formalized manner which should be suitable for communication, interpretation, or processing by human or electronic machine. Big data alters the way the data is managed and used. Big data refers to voluminous data which ranges in Exabyte's and beyond. It is defined as the amount of data just beyond technology's capability to store, manage and process efficiently. Big data, which refers to the data sets that are too big to be handled using the existing database management tools. Big data has developed into a hot topic that attracts extensive attention from academia, industry, and governments around the world. This position of paper, briefly introduce the concept of big data, including its definition and application. Big data has so much of applications in various fields such as agriculture, weather forecasting, health care, ERP etc... The paper highlights some of these applications.

**Keywords-** K-NNC, Fuzzy logic, PNN, ANN, SVM.

### I. INTRODUCTION

Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization.. Although big data doesn't refer to any specific quantity, the term is often used when speaking about petabytes and exabytes of data, much of which cannot be integrated easily. The data it handles will be of either structured or unstructured. Data is increased by 90% in the last 2 to 3 years. Data is generated from various sources :web, social networking, Weather forecasting etc. As the use of internet increasing, the amount of Big Data continues to grow. Big Data is now seen as being of critical important to majority of firms. 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%. Only 1.8% of firms indicated that Big Data was not important to the firm[9].

The 3Vs that define *Big Data* are Variety, Velocity and Volume. The 3v's has been represented in the below fig 1.1  
**Variety:** Data can be stored in multiple format. , it may be in the form of video, SMS, pdf or something we might have not thought about it. It is the need of the organization to arrange it and make it meaningful. It will be easy to do so if we have data in the same format, however it is not the case most of the time. The real world have data in many different formats and that is the challenge we need to overcome with the *Big Data*.

#### Volume

The data can be find in the format of videos, musics and large images on our social media channels. It is very common to have Terabytes and Petabytes of the storage system for enterprises. As the database grows the applications and architecture built to support the data needs to be re-evaluated quite often. Sometimes the same data is re-evaluated with multiple angles and even though the original data is the same the new found intelligence creates explosion of the data. The big volume indeed represents *Big Data*.

#### Velocity

The speed at which the analytics need to be done. The data growth and social media explosion have changed the way people look at data. The news channels and radios have changed their way of communicating data due to the results public are receiving the news updates instantly. Today, people reply on social media to update them with

the latest happening. On social media sometimes a few seconds old messages (a tweet, status updates etc.) is not something interests users. They often discard old messages and pay attention to recent updates. The data movement is now almost real time and the update window has reduced to fractions of the seconds. This high velocity data represent *Big Data*.

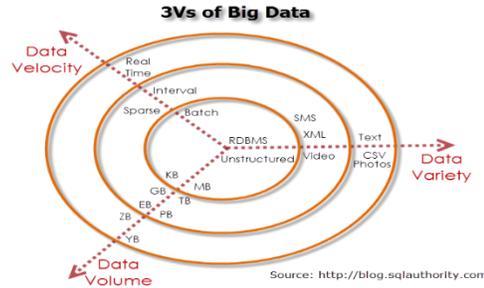


Fig 1.1.3 Vs of Big Data

Table 1.2 Data Vs Big Data

Category	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
Organization	Centralized	Distributed across Multiple Servers

**II. TECHNOLOGIES USED IN BIG DATA**

An enterprise will have a computer to store and process big data. The data will be stored in an RDBMS like Oracle Database, MS SQL Server or DB2 and sophisticated softwares can be written to interact with the database, process the required data and present it to the users for analysis purpose. when it comes to dealing with huge amounts of data, it is really a tedious task to process such data through a traditional database server. However storing and analysing big data effectively is possible using Hadoop

Doug Cutting, Mike Cafarella and team took the solution provided by Google and started an Open Source Project called HADOOP in 2005. Hadoop runs applications using the MapReduce algorithm, where the data is processed in parallel on different CPU nodes. [10]

Hadoop is an open source framework that enables the distributed processing of big quantities of data by using a group of dispersed machines and specific computer programming models. The main components of Hadoop are:

1. Its file system HDFS, that allows access to data scattered over multiple machines without having to cope with the complexity inherent to their dispersed nature;
2. Map Reduce, a programming model designed to implement distributed and parallel algorithms in an efficient way.

### III. BIG DATA USE CASES

#### i. Big Data Is Changing Healthcare

Big data in health-care refers to the patient care data such as physician notes, Lab reports, X-Ray reports, case history, diet regime, list of doctors and nurses in a particular hospital, national health register data, medicine and surgical instruments expiry date identification based on RFID data. Big Data in healthcare is being used to predict epidemics, cure disease, improve quality of life and avoid preventable deaths. Big Data is also helping in the fight against the spread of epidemics. In Africa, mobile phone location data is proving highly valuable in efforts to track population movements, which helps to predict the spread of the Ebola virus.

Another partnership that has just been announced is between Apple and IBM. The two companies are collaborating on a big data health platform that will allow iPhone and Apple Watch users to share data to IBM's Watson Health cloud healthcare analytics service.[4]

#### ii. Big Data in Agriculture

The traditional process used to create successful crop varieties is costly, labour intensive, and can take 10 years or more. Big data speed things up. Production agriculture is complex; where biology, weather, and human actions interact. With the availability of GPS and other technologies, producers can track yields, and control equipment, monitor field conditions, and manage inputs at very precise levels across fields, substantially increasing productivity and profitability. Big data firms can test varieties of genetics, crop inputs, and conditions across hundreds of fields, soils, and climates. They can conduct field-plot trials in real-life conditions over thousands of acres. This gives farmers information to optimize planting down to the individual seed in specific fields with specific soil and climate conditions[5].

#### iii. Big Data on Government

The availability and the skill level of Big Data information management talent will have a direct impact on government's ability to successfully deploy Big Data. The new storage strategy like Cloud-based backup, recovery, and archiving provide different benefits. The Prime minister of India Narendra Modi On July 1, 2015 launched the Digital India Programme to ensure availability of government services to citizens electronically by improving online infrastructure and increasing Internet connectivity. Big Data and analytics will play a predominant role in such transformation by way of cloud, mobile technology and other social technologies that gather data for the purpose of ascertaining and accordingly addressing concerns of people, and increase the level of their engagement to culminate these plans successfully.[4]

#### iv. Big Data in Weather Forecasting

The recent advances in technologies like satellite sensors and other resources are used by weather forecasting system to help general people for accurate prediction of weather. The increase in volume of environmental data put forth the need of Big Data techniques to manage, process and store data. IBM Research scientists are taking the lead in bringing the most sophisticated data analytics to bear on weather forecasting. Their long-term weather analysis project, called Deep Thunder, combines data with sophisticated mathematical algorithms and computing power. Apache Hadoop Map Reduce Framework is used to analyse huge data set of weather forecasting.

#### v. **Big Data in Aligning Corporate System**

Big Data have been a ERP professional for last 14 years. There has been a tremendous changes in the ERP system, implementation methodologies within corporate all across the world from small to large. Each corporate has been investing a lot into upgrading their systems. The main aim is for Capturing the maximum data, processing it faster and making it available to right employee for taking right decision. Data requirements have been huge that meeting it through conventional ERP modules is very difficult. The ERP software companies could have added many extensions ,support packs and newer functionalities have been introduced frequently. Expanding the width and depth of ERP and every other relevant system for Big Data is a way to go[8]

### IV. CHALLENGES WITH BIG DATA

#### i. **Heterogeneity of Data**

The Big Data, data may be either structured or unstructured ,this Heterogeneity is the big challenge in data Analysis and analysts need to cope with it. Consider an example of patient records in Hospital. We will generate each record for each medical test. And we will also generate a record for hospital stay, which may differ according to patient illness Thus resulting in an unstructured design. A good Data Analysis is required for this[1].

#### ii. **Privacy and Security**

Data Security is the practice of protecting data from unauthorized access. The rapid change in technologies results in difficulty in predicting about the usage of information available on cyber space . The inappropriate use of personal data is increasing especially when extracting from multiple sources. For example the location –based services extract data related to location details from subscribers resulting in privacy concerns. The private informations shared through social medias/online services such as Facebook, Twitter etc is also a serious privacy breach[1]

Today many private information are shared through online services like Facebook, Twitter, etc., and until now many people don't understand what it means to share data, how this data can be associated together to come up with more personal details not intended to share location-based services that asks subscribers to share their location resulting in clear privacy concerns. Some people think that hiding their identity alone without hiding their location would not properly address privacy concerns which are not true. A location-based service provider can conclude the identity of the subscriber by tracing subsequent location information. It resembles the user leaving a trail of packer crumbs behind him which could be linked to a certain location or residence and therefore used to detect the user's identity. Also other types of private information such health details(e.g. frequent visit to cancer treatment center) or religious preferences (e.g. presence in a church) can be disclosed by monitoring movement of anonymous users and analyzing movement patterns over time. It is more difficult to hide user location than his identity because to utilize location-based services he/she needs to expose your location. Today many private information are shared through online services like Facebook, Twitter, etc., and until now many people don't understand what it means to share data, how this data can be associated together to come up with more personal details not intended to share

#### iii. **Skills availability**

Big Data is being supported with new tools and is being looked at in different ways. There exist a shortage of people with the skills to bring together the data, analyze it and publish the results or conclusions. [1]

### V. CONCLUSION

Big data has made a strong impact in almost every sector and industry today .This is the era of Big data where almost all application requires data to provide services quickly and efficiently. In this paper, we have briefly reviewed the opportunities and significance of big data, as well as some challenges that big data brings us. Hence Big Data is becoming the new Final Frontier for scientific data research and for business applications. The Securely Management of Big Data with today's threat spectrum is a big challenging issue. Because today's we have an overwhelming growth of data in terms of volume, velocity and variety. So as for future research is needed to build a generic architectural framework towards addressing these 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. XiaolongJin, Benjamin W.Wah, XueqiCheng , YuanzhuoWang “ 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 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. <https://www.tableau.com/resource/top-10-big-data-trends>
10. [https://www.tutorialspoint.com/hadoop/hadoop\\_big\\_data\\_overview.htm](https://www.tutorialspoint.com/hadoop/hadoop_big_data_overview.htm)