ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

BIG DATA ANALYTICS IN HEALTHCARE: CHALLENGES, TOOLS AND TECHNIQUES : A SURVEY

M.Ashok Kumar
Ph.D Research Scholar
Dept. of Computer Science
Periyar University
Salem-11
williamashok@gmail.com

Dr. I.Laurence Aroquiaraj
Assistant Professor
Dept. of Computer Science
Periyar University
Salem-11
laurence.raj@gmail.com

P.Surya
Ph.D Research Scholar
Dept. of Computer Science
Periyar University
Salem-11
suriyaa14@gmail.com

Abstract:

In this paper, we introduce about the big data and big data analytics in healthcare . Big data is a popular term, It describes about the exponential growth and availability of data, both structured and unstructured. Bigdata analytics has the possibility of advanced patient care and clinical decision support in healthcare. This paper also explains varies algorithms and platforms for bigdata analytics and discussion on its advantages and challenges.

Keywords--Big data; Map reduce; Hadoop; Healthcare;

I . INTRODUCTION

Digital data have a vital role in the computerized world. Now we are living in data world. Everywhere we are seeing only data. In the 100 % of the data world, 90 % of the data has been created in the last 2 years alone. Every day people create 3 quintillion bytes of data. This data comes from everywhere, It may be a Climate information, Posts in the social network, Digital pictures & videos, Cell phone GPS signals, Purchase records. Currently medical data, weather forecasting data, Government census data and most of the business transactions such as banking, retail marketing, e-commerce, etc., are stored in computer for future process. The needs for processing of these data are growing day-by-day. But presently, data's are stored in numerous format such as text, sensor data, 3D data, video, audio, analog data, images and etc[1].

The storage of data occupies more space, since the data usage is increasing every day. The important thing is 'how the data's are storing' and,' how the data's are processing'. The data in computing are represented as structured format in the olden days. It is stored as tabular format.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

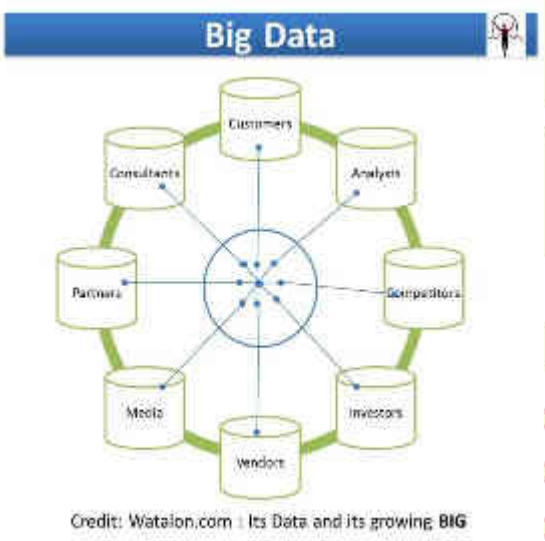
International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

Much data today is not natively in structured format; The problems start right away during data acquisition, when the massive data requires us to make decisions, about what data to retain and what to remove, and how to store what we save reliably with the correct metadata. Dealing with these Big Data (massive sets of data) is a highly challenging issue for the data analysts.

Big Data:

Big data is a buzzword that is used to describe the large amount of data either structured or unstructured data format. Exactly, if the data which is beyond to the storage capacity & which is beyond to the processing power, that data we are calling 'BIG DATA'. Big data is so large and is difficult to process using the old database and software techniques. Having data larger it needs different Approaches, Techniques, Architectures, and Tools.



Credit: Watalon.com : Its Data and its growing **BIG**

Why Big Data?

The purpose of collecting and processing all of the big data is to assist us in making meaningful decisions. The data must be truthful. Big data is "hot" due to the potential value that it brings to us. With high competition for resources becoming more and more intense than ever, organizations – both public and private sectors – have been analysing, searching for methods to separate themselves from their competitors by diving into the wealth of information to improve their effectiveness, competence, cost-effectiveness and more.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

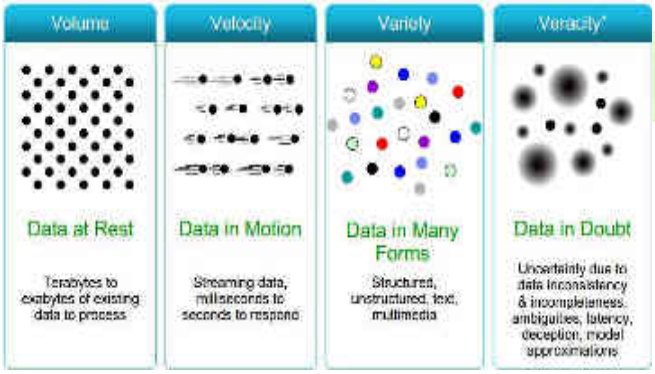
Vol. 2, Special Issue 10, March 2016

Who is generating Big Data?

- Social media and Network.
- Scientific instruments.
- Sensor technology.
- E-commerce.
- Internet.
- Mobile computing.
- Telecom.
- Health care.

II. CHARACTERISTICS

Characteristics of Big Data in various dimensions are, **Volume** (amount of data), **Velocity** (speed of data store and retrieve), and **Variety** (range of data types and sources). Nowadays the researchers include some more dimensions, **Veracity** (uncertainty data) and **Value** (big data is about supporting decisions, need the ability to act on the data and derive value). **Variability** (Growing velocities and varieties of documents, data flows can be highly inconsistent with periodic peaks). **Complexity** (Today's data comes from several sources. And it is still responsibility to link, match, clean and convert data across systems. It is essential to connect and compare relationships). **Validity** (the data correct and accurate for the indexed use), **Volatility** (how long the data is valid and how long should it be stored), **Visualization** (difficult graphs that can contain many variables of data)[8].



ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

Issues and Challenges

The common issues associated with this broad research area are as follows:

- Application management/ implementation of big data analysis.
- Big data use-case and technology assessment.
- Data migration from existing data stores.
- Developing capacity plans for new and existing systems.
- Big data search and Mining.
- Big data security and privacy.
- Data virtualization.
- Visualization.
- Performance tuning.
- Capture .
- Curation.
- Storage.
- Search.
- Sharing .
- Transfer.
- Analysis .
- Integration.

Application area

1. Health care and Bio technology

2. E – governance

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

3. Social network and Social media

4. Weather forecasting, Agriculture

5. Education data

6. Banking, Insurance, Finance

7. Retail, Real estate

8. Crm, Customer

9. Airways, Tranportation

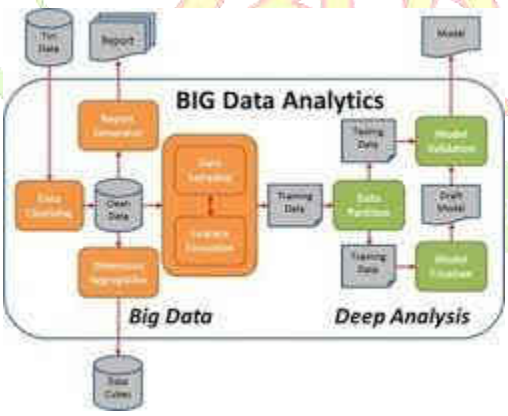
10. Automotive industry

11. Supply chain, Logistics and Industrial engineering

12. Media, Entertainment

Big Data Analytics:

Big data analytics denotes to the process of Collecting, Organizing, Analyzing, Inspecting, Cleaning, Transforming and Modelling, large sets of data (Big data) to discover patterns and other useful information. It basically wants the knowledge that comes from analyzing the data. And also it is the process of examining enormous amounts of data to predict the unknown patterns and correlations, and other useful information from known data.



Big data analytical types

1. DESCRIPTIVE ANALYTICS

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

This purpose is to summarise what has happened.

2. PREDICTIVE ANALYTICS

   It forecasts what might happen in the future.

3. PRESCRIPTIVE ANALYTICS

   It needs to prescribe an action, so the business decision maker can take this information and act.

Data is highly generated by the people from past many years. But available data for analysis is very less.

Why Big Data Analytics possible

- Low cost memory for storing large amount of data.
- High speed process.
- Parallel and distributed systems.
- Cloud based technology.

Challenges in Big Data Analysis

- Heterogeneity and Incompleteness.
- Scalability.
- Timeliness.
- Privacy.
- Human Collaboration.
- System Architecture.

III. BIG DATA TOOLS AND TECHNIQUES

There are some tools and techniques available to analyse the huge datasets.

Hadoop

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

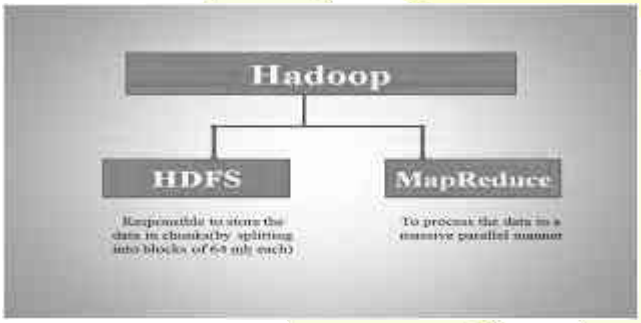
Vol. 2, Special Issue 10, March 2016

Hadoop is a software framework for distributed processing of large datasets across large clusters of computers. It is using for storing and processing huge data sets. The inventor of the Hadoop is "Doug Cutting".

Large data sets → Tera byte and Peta byte of data.

Large clusters → Hundreds (or) Thousands of nodes.

Hadoop is based on a simple programming model called 'MAPREDUCE', and also it based on a simple data model, any data will fit in it. It is an open source project, written in java, optimized to handle, it has a great performance.



Key benefits of Hadoop

- Reliable solution based on an unreliable hardware.
- It designed for large files.
- It loads data first and structure it later.
- It designed for scalability.

Hadoop components

- HDFS
- Mapreduce.

Hadoop distributed file system has been designed to be easily portable from one platform to another. It has been built by using java language. Use blocks to store a file or parts of a file. HDFS is used for storage purpose, Mapreduce is used for processing and distributing purpose.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016

Hadoop architecture

Components	Functions
HDFS	Storage
Map reduce	Distributing
HBASE	Read/write
Pig	Scripting
Hive	SQL
Ooze	Workflow
Zookeeper	Coordination
Kafka	Messaging
Mahout	Machine learning

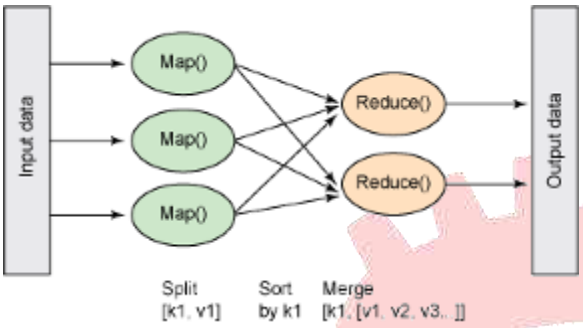
Mapreduce

It is a simple data parallel programming model designed for scalability, A Mapreduce program consists of map and reduce functions. It executing in three stages 1.Map, 2.Shuffle, and 3.Reduce. It is a functional model. Scale free programming model.

ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)

Vol. 2, Special Issue 10, March 2016



IV. HEALTHCARE

Bigdata and its related technologies have improved healthcare enormously from understanding the Origins of diseases, Better diagnoses,Helping patients to moniter their own conditions. Healthcare organizations can improve their quality of service by analyzing the effectiveness of a treatment and also the efficiency of the healthcare delivery process. Due to advance technologies the paper works converted into digital format (Digital Health Records (DHR), or Electronic Health Records (EHR)[2]. Since information is in the digital form, healthcare providers can use some available tools and technologies to analyze that information and generate valuable insights. Full view for every patient is created by electronic health records, scanned documents,medical images,notes from physicians,information about environment.

1. Application of bigdata in healthcare

(i) Modified treatment planning.

Based on the particular patient, diagnosis can be done. It decides the seperate and particular treatement and drugs for that patient. Live and realtime analysis can be done using MapR and Hadoop based on the results after analysis.

(ii) Application review.

To give the best treatement , bigdata analytics of healthcare are needed. It gives meaningful information on a particular health condition, how effective a drug in a particular treatement.

(iii) Helped diagnosis.

Doctors can insulate and treat the patient based on some aspects like, symptoms, medical history, side effects. Hadoop can give information which will be useful to the doctors by using prediction and machine learning.

(iv) Fake detection.

Fake should be avoided in healthcare. By this, cost will be reduced for insurance companies and patients. The previous techniques using the traditional healthcare analytics are not so good to analyzing the fraud. But now, with the help of bigdata analytics using new tools and techniques analyzing the fake can be done easily on a more real time basis.

2. Prediction.

Prediction is telling about what might happen in the future[7].

Conclusion:

Still it is not clear how an optimal architecture of an analytics system should be to deal with healthcare data and with real-time data at the same time. The new architecture for big data analysis will add more values for analyzing the enormous amount of datasets. Applying the data mining techniques or implementing new mining concepts with the proposing data analysis architecture will give an improved result for the large data handlers. The beneficiaries are such as CRM, Climate Predictors, Medical Health Care, Government, Small and Medium Enterprises, etc.

REFERENCES

[1] [WOR, 2014] Worldometers, "Real time world statistics," 2014, http://www.worldometers.info/world-population/

[2] [CHE, 2013] D. Che, M. Saffron, and Z. Peng, "From Big Data to Big Data Mining: challenges, issues, and opportunities," in Database Systems for Advanced Applications, pp. 1–15, Springer, Berlin, Germany, 2013.

[3] [JEA, 2013] Jean Yan,U.S. General Services Administration "Big Data, Bigger Opportunities", April 9 , 2013.

[4] [SER, 2013] Serif SAGIROGLU and Duygu SINANC, Dept of computer science, Gazi University,Ankara, Turkey @2013.

[5] [PAR,2014] ParthChandarana, V.E.S.I.T, ChemburM.Vijayalakshmi, Department of Information Technology, "Big Data Analytics Frameworks", 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA).

[6] [PUN, 2013] Puneet Singh Duggal, Department of Computer Science & Engineering Birla Institute of Technology, Sanchita Paul, Department of Computer Science & Engineering Birla Institute of Technology, "Big Data Analysis: Challenges and Solutions", International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV.

[7] [AMR, 2014] Amrit Pal, PinkiAgrawal, Kunal Jain, Sanjay Agrawal," A Performance Analysis of MapReduce Task with Large Number of Files Dataset in Big Data Using Hadoop", 2014 International conference on communication system and mesh technologies.

[8] [SER,] Serif SAGIROGLU and Duygu SINANC, Gazi University, Ankara, Turkey Department of Computer Engineering, Faculty of Engineering," Big Data: A Review".

[9] [ANI,] Anirban Mukherjee, JoydipDatta, RaghavendraJorapur, Ravi Singhvi, SauravHaloi, WasimAkram,"Shared Disk Big Data Analytics with Apache Hadoop ".

[10] [SAC,] SachchidanandSingh, Business Analytics Division, IBM India Software Lab (ISL),Nirmala Singh Data Warehouse Division, "Big Data Analytics".

[11] [MAD,] MadjidKhalilian, Norwati Mustapha "Data Stream Clustering: Challenges and events".

[12] [BOG,] BogdanBatrinca• Philip C. Treleaven" Social media analytics: a survey of techniques, tools and platforms".

[13] [DEW,] Dewey Sun, Guangyan Zhang, WeiminZheng, and KeqinLi, Department of Computer Scienceand Technology, Tsinghua University, Beijing 100084, China,"Key Technologies for Big Data Stream Computing".