Review of Big data Challenges in Healthcare Application

Madhura A. Chinchmalatpure, Dr. Mahendra P. Dhore

madhura.naralkar@gmail.com, Department of Computer Science, SSESA's Science College, Congress Nagar, Nagpur

mpdhore@rediffmail.com, Department of Electronics and Computer Science, RTM Nagpur University Campus, Nagpur

Abstract : *Big data is a term defining collection of large datasets. After interaction with patients here we implement electronic health record(EHR) which is highly capable of storing voluminous data in database. and this includes patient's previous medical data, laboratory test values, current treatment given to patient, doctors prescription, diagnostic reports, pharmacy information health insurance related data, medical journals are used to proper investigate and analysis. In this paper, the objective is to discuss the characteristics and challenges of big data in which the process is to extract the information from large sets of big data*
**Keywords:** *Big data analysis, Electronic Health Records(EHR), Hadoop, MapReduce*

I. Introduction

In Digitized world the vast amount of data is generated, to properly analyzed that data big data concept is generated.

All these information collectively forms big data in healthcare. The rising cost of healthcare is one of the world's most important problem. These databases are designed for clinical purposes. Big data has some challenges can be divided into

1.1 Data Challenge: Big data is the term used to describe collection of large and complex datasets having the "4 V" definition: volume(Structured(financial data),unstructured(Social media,photos,videos,voice recording), variety(range of data types and sources), velocity(speed of data in and out) and value (e.g. medical images, electronic Health Record (EHR), biometrics data etc.).
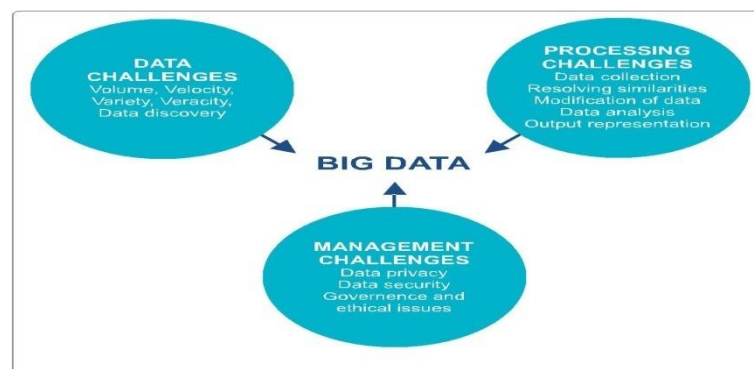
1.2 Processing Challenges: Data Collection, Resolving similarities, Modification Of data, Data Analysis, output representation
1.3 Management Challenges: Data Privacy, Data Security, Governence and ethical issues

Big data analytics has challenging research area, it refers to tools and techniques that are applied to healthcare dataset to obtain the data from database. A well known tools are **Hadoop** and **mapreduce** in which
i) to collect current data and preprocess data
ii)Transform pre processed data into form that can be handled by machine learning method
iii)To extract the related information

The aim in the big data analytics research is to extract the healthcare related information by using Hadoop and mapreduce tools. Large amount of data to be converted into Electronic health reports (EHR) of patient data means data extraction involves obtaining data from one data source and loading it into standard database or data ware house.

All big data analytics tools aim at a better extraction of data to find right solution in the shortest time possible, to find results that are distributed over various server and replicated through several nodes depending on geographical area. By employing proper electronic communication technology to exchange the information of individual patients along healthcare centers will leads to get proper treatment at right time in remote location at low cost.

Big data is very challenging area. Data is too big to process using conventional tools of data processing have to work to design and develop new techniques which effectively handle the processing of big data.

II. Impact of Big Data in Healthcare

Big data have created five pathways to assist them in redefining value and identifying tools that are appropriate for the new era. They focus on the following concepts:

- **Right living.** Patients must be encouraged to play an active role in their own health by making the right choices about diet, exercise, preventive care, and other lifestyle factors.
  - **Right care.** Patients must receive the most timely, appropriate treatment available. In addition to relying heavily on protocols, right care requires a coordinated approach, with all caregivers having access to the same information and working toward the same goal to avoid duplication of effort and suboptimal treatment strategies.
  - **Right provider.** Any professionals who treat patients must have strong performance records and be capable of achieving the best outcomes. They should also be selected based on their skill sets and abilities rather than their job titles. For instance, nurses or physicians' assistants may perform many tasks that do not require a doctor.
  - **Right value.** Providers should continually look for ways to improve value while preserving or improving health-care quality.
  - **Right innovation.** Patients must focus on identifying new therapies and approaches to health-care delivery. They should also try to improve the innovation engines themselves—for instance, by advancing medicine and boosting R&D productivity.

The value pathways evolve as new data become available, fostering a feedback loop. The concept of right care, for instance, could change if new data suggest that the standard protocol for a particular disease does not produce optimal results. And a change in one pathway could spur changes in others, since they are interdependent. An investigation into right value, for example, could reveal that patients are most likely to suffer costly complications after an appendectomy if their physician performs few of these operations

III Methodology

The major challenge in big data analytics is to locate the required information from tables and extract the information contained in database .Defining appropriate tools and techniques for the same in highly desired in order to enrich the healthcare research. Obviously this involves a number of preprocessing and classification steps. For example, a big data analytics can be employed to solve the problem of searching relevant information about a particular individual from the huge database. The big data analytics research focuses on analyzing the data automatically, update the database, extract the most informative data. The information gathered from patients, doctors pharmacy having representation are considered for data analytics. In this proposed research work our approach is to study the existing data and to develop and implement a new system for data analysis which have a better degree of performance.

Platforms & tools for big data analytics in healthcare
Platform/Tool Description
The **Hadoop Distributed File System (HDFS)** : HDFS enables the underlying storage for the Hadoop cluster. It divides the data into smaller parts and distributes it across the various servers/nodes.

MapReduce : MapReduce provides the interface for the distribution of sub-tasks and the gathering of outputs. When tasks are executed, MapReduce tracks the processing of each server/node.

PIG and PIG Latin (Pig and PigLatin) : Pig programming language is configured to assimilate all types of data (structured/unstructured, etc.). It is comprised of two key modules: the language itself, called PigLatin, and the runtime version in which the PigLatin code is executed.

Hive : Hive is a runtime Hadoop support architecture that leverages Structure Query Language (SQL) with the Hadoop platform. It permits SQL programmers to develop Hive Query Language (HQL) statements akin to typical SQL statements.

Jaql : Jaql is a functional, declarative query language designed to process large data sets. To facilitate parallel processing, Jaql converts "'high-level' queries into 'low-level'queries" consisting of MapReduce tasks.

Zookeeper : Zookeeper allows a centralized infrastructure with various services, providing synchronization across a cluster of servers. Big data analytics applications utilize these services to coordinate parallel processing across big clusters.

HBase : HBase is a column-oriented database management system that sits on top of HDFS. It uses a non-SQL approach.

Cassandra : Cassandra is also a distributed database system. It is designated as a top-level project modeled to handle big data distributed across many utility servers. It also provides reliable service with no particular point of failure (http://en.wikipedia.org/wiki/Apache_
Cassandra) and it is a NoSQL system.

Oozie : Oozie, an open source project, streamlines the workflow and coordination among the tasks.

Lucene : The Lucene project is used widely for text analytics/searches and has been incorporated into several open source projects. Its scope includes full text indexing and library search for use within a Java application.

Avro : Avro facilitates data serialization services. Versioning and version control are additional useful features.

Mahout : Mahout is yet another Apache project whose goal is to generate free applications of distributed and scalable machine learning algorithms that support big data analytics on the Hadoop platform.

Outline of big data analytics in healthcare methodology
Step 1 Concept statement
      • Establish need for big data analytics project in healthcare based on the "4Vs".
Step 2 Proposal
      • What is the problem being addressed?
      • Why is it important and interesting?
      • Why big data analytics approach?
      • Background material
Step 3 Methodology
      • Propositions
      • Variable selection
      • Data collection
      •ETL and data transformation
      • Platform/tool selection
      • Conceptual model
      • Analytic techniques -Association, clustering, classification, etc.
      • Results & insight
Step 4 Deployment
      • Evaluation & validation
    • Testing Source:

IV. Challenges in Healthcare application in Big Data
Here we discuss some of the outstanding challenges in this field, with regard to:
(i)privacy and ownership of data
(ii) analysis of user data

Privacy and Ownership of Data*:* Privacy and ownership of data is big issue. There is always conflict between patient and doctor's data, there are many organizations that believe that data should be open and that openness and interoperability provide them with a competitive advantage .

Analysis of User Data*:* The major focus of analysis of user data is in determining user's intent. This is certainly the focus of a lot of the predictive analytics used in online advertising, and the reason that search advertising is far more effective than display advertising.

V. Conclusion

The Big data analytics technique has a wide range of technique. To represent a record, tables must be known in database. The tool we implement contains electronic record which is highly capable of storing voluminous data in database. and this includes patient's previous medical data, laboratory values, current treatment given to patient, doctors prescription, diagnostic reports, pharmacy information health insurance related data, medical journals are used to proper investigate and analysis. In this proposed research work, the objective is to develop the system in which the result should be depends on data we input if the data is correct then it should give consistent result.

References

[1] Wullianallur Raghupathi1* and Viju Raghupathi2," Big data analytics in healthcare: promise and Potential"

[2] Ch.Sai Krishna manohar,"Analytics of Big Data Science using Big Data" Department of Information Technology, Tirumala Engineering College, Narasaraopet, India, IOSR Journal Of Computer Engineering(IOSR-JCE),e-ISSN:2278-0661,p-ISSN:2278 Volume 10,Issue2(Mar-Apr)

[3] Md.Salahuddin,Shaikh Muhammad Allayear,Sung Soon Park,"The Architectural Design Of Healthcare Services with big data storing Mechanism",IOSR Journal Of Computer Engineering(IOSR-JCE),e-ISSN:2278-0661,p-ISSN:2278-8727,Volume 16,Isseue 5,Ver.VIII(Sep-Oct.2014),PP 81-94

[4] Gemson Andrew Ebenezer J.1 and Durga S.2," BIG DATA ANALYTICS IN HEALTHCARE: A SURVEY" ARPN Journal of Engineering and Applied Sciences ©2006-2015 Asian Research Publishing Network (ARPN). All rights reserved. VOL. 10, NO. 8, MAY 2015, ISSN 1819-6608

[5] Manish Kumar Kakhani1, Sweeti Kakhani2 and S.R. Biradar3," Research Issues in Big Data Analytics", International Journal of Application or Innovation in Engineering & Management (IJAIEM), **Volume 2, Issue 8, August 2013, ISSN 2319 – 4847**

[6 **]** Nishchol Mishra1, Dr.Sanjay Silakari2," Predictive Analytics: A Survey, Trends, Applications, Oppurtunities & Challenges", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012, 4434- 4438,ISSN:0975-9646

[7] Elisa Bertino," Big Data - Opportunities and Challenges", 2013 IEEE 37th Annual Computer Software and Applications Conference

[8] Dr Saravana kumar N M, Eswari T , Sampath P & Lavanya S," Predictive Methodology for Diabetic Data Analysis in Big Data", 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

[9] **S. Divya Meena1, M. Revathi2,"** Predictive Analytics on Healthcare: A Survey", **International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064**

[10] Nishchol Mishra1, Dr.Sanjay Silakari2," Predictive Analytics: A Survey, Trends, Applications, Oppurtunities & Challenges", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012, 4434- 4438,ISSN:0975-9646

[11] Ekaterina Olshannikova1*, Aleksandr Ometov1, Yevgeni Koucheryavy1 and Thomas Olsson2," Visualizing Big Data with augmented and virtual reality: challenges and research agenda", Olshannikova et al. Journal of Big Data (2015) 2:22 DOI 10.1186/s40537-015-0031-2