International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 05 | May-2016 www.irjet.net p-ISSN: 2395-0072

Behavior Analysis In Health Care System Using BigData Analytics

Perinba Jothi T

Dept.of Information Technology

J. J.College of Engg., & Tech., Trichy, India

Abstract - *The health care industries having a major problem called standardization; this problems are overcome by using big data analytics. In health care industries massive data generated and collected from the different agents and heterogeneous source. The data are medical record, images from patient scan, social media conversation about health informatics from the heterogeneous sources. These data needs to be managing and organizing for an efficient storing and analyzing. The hadoop framework is used to create big data environment to support vast collections of data. The hive component is used to managing and organizing the massive collection of rapid growing health care data. The hive-QL technique is used for analysis. The hive-QL technique generates analysis report in minimum amount of time. The analysis reports are helps to discovery knowledge in health care industries.*

Key Words: Big data; Hadoop; Hive; Hive-QL; Health care data

1. INTRODUCTION

The healthcare industry historically has generated large amounts of data, driven by record keeping, compliance & regulatory requirements, and patient care. While most data is stored in hard copy form, the current trend is toward rapid digitization of these large amounts of data [1]. Driven by mandatory requirements and the potential to improve the quality of healthcare delivery meanwhile reducing the costs, these massive quantities of data known as 'big data' hold the promise of supporting a wide range of medical and healthcare functions, including among others clinical decision support, disease surveillance, and population health management. Reports say data from the U.S. healthcare system [2-5] alone reached, in 2011, 150 exa bytes. At this rate of growth, big data for U.S. healthcare will soon reach the zetta byte (1021 gigabytes) scale and, not long after, the yotta byte (1024 gigabytes). Kaiser Permanente, the California-based health network, which has more than 9 million members are believed to have between 26.5 and 44 petabytes of potentially rich data from EHRs, including images and annotations.

Big data in healthcare is vast not only because of its volume but also because of the diversity of data types and handling massive collection of heterogeneous data in efficient manner. The massive of data has related to patient healthcare and safety make up "big data" in the healthcare industry. It is also includes the clinical data from CPOE and clinical decision support systems, which consists of physician's written notes and prescriptions, medical imaging, laboratory, pharmacy, insurance, and other administrative data.

The patient data in electronic patient records (EPRs); machine generated/sensor data, such as from monitoring vital signs, social media posts, including Twitter feeds (so-called tweets), blogs, status updates on Facebook and other platforms, and web pages, less patient-specific information, including emergency care data, news feeds, and articles in medical journals. Analytics when applied in the context of big data is the process of examining large amounts of data [6], from a variety of data sources and in different formats.

Big data analytical approaches used to recognize inherent patterns, correlations, and anomalies that can discovered because of integrating massive amounts of data from different data sets. The millions of users search the internet for health related information. The big data analytics [7] must be used in the health care system for better performance and also the accurate decision.

This paper provides an overview of big data analytics in healthcare system as it is emerging as a discipline is as follows:

(1) Define and discuss the definition of big data and characteristics of big data analytics in healthcare.

(2) Massive collection of health care data are stored and organized by using of hadoop framework and hive component.

(3) These massive collection of health care data are handled and analyzed by Hive-QL technique After analysis, the final knowledge can be generated with minimum amount of time.

2. RELATED WORK

Traditional enterprise data includes customer information from CRM systems, transactional ERP data, web store transactions, and general ledger data. Machine-generated /sensor data includes Call Detail Records ("CDR"), weblogs, smart meters, manufacturing sensors, equipment logs (often referred to as digital exhaust), and trading systems data. Social data includes customer feedback streams, micro-blogging sites like Twitter, and social media platforms like Facebook. These data's are small so system cannot give the

accurate result. To overcome this problem, big data analytics is used in the health care system.

3. BIG DATA ANALYTICS AND CHARACTERISITICS

In recent, the big data analytics is the emerging one. It has the capabilities to analysis massive collections heterogeneous data at a time.

3.1 Big data analytics

Big data analytics refers to the process of collecting, organizing and analyzing large sets of data ("big data") to discover patterns and other useful information. Not only will big data analytics help us to understand the information contained within the data, but it will also help identify the data that are most important to the business and future business decisions.

Big data is important to perform a successful analytical process. Big data analytics types are predictive analytics, prescriptive analytics, descriptive analytics, and diagnostics analytics. Big data handles Yota /Zeta bytes of information for analysis purpose. A conceptual architecture of big data analytics as shown in Fig 1.

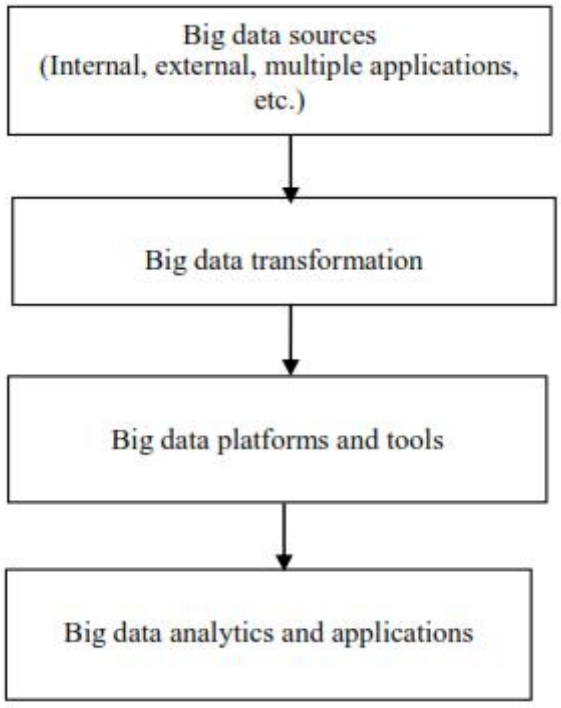

Fig -1: A conceptual architecture of big data analytics

Big Data is data whose has scale of diversity, and complexity. It is required new architecture, techniques, algorithms, and analytics to manage the massive data and extract value, hidden knowledge from massive collection of data. Big data solutions attempt to cost-effectively solve the challenges of large and fast-growing data volumes and realize its potential analytical value. For instance, trend analytics allow you to figure out what happened, while root cause and predictive analytics enable understanding of why it happened and what is likely to happen in the future. The innovative analytics have applied to identify opportunities and improving the future.

3.2 Characteristics of Big data

The big data characteristics are classified as three V's. These three V's are described as follows.

Volume: Big data means there is many data — terabytes or even peta bytes (1,000 terabytes). This is perhaps the immediate challenge of big data, as it requires scalable storage and support for complex, distributed queries across multiple data sources. While many organizations already have the basic capacity to store large volumes of data, the challenge is being able to identify, locate, analyze, and aggregate specific pieces of data in a vast, partially structured data set.

Variety: Big data is an aggregation of many types of data, both structured and unstructured, including multimedia, social media, blogs, Web server logs, financial transactions, GPS and RFID tracking information, audio/video streams and Web content. While standard techniques and technologies exist to deal with large volumes of structured data, it becomes a significant challenge to analyze and process a large amount of highly variable data and turn it into actionable information. However, this is also the potential of big data such as effective analytics allow making better decisions and realizing opportunities that would not otherwise exist.

Velocity: While traditional data warehouse analytics tend based on periodic daily, weekly, or monthly loads and updates of data, big data is processed and analyzed in real- or near-real-time. This is important in healthcare for areas such as clinical decision support, where access to up-to-date information is vital for correct and timely decision-making and elimination of errors. Current data needed to support automated decision-making; after all cannot use five-minute-old data to cross a busy street. Without current data, automated decisions cannot be trusted, forcing expensive and time-consuming manual reviews of each decision.

4 BIG DATA ANALYTICS IN HEALTH CARE SYSTEM

Big data has many implications for patients, providers, researchers, payers and other healthcare constituents. It will affect how these players engage with the healthcare ecosystem, especially when external data, regionalization, globalization, mobility, and social networking are involved.

The healthcare model is undergoing an inversion. In the old model, facilities and other providers were incented to keep patients in treatment — that is, more inpatient days translated to more revenue. The trend with new models, including accountable care organizations (ACO), is to incent and compensate providers to keep patients healthy [8]. At the same time, patients are increasingly demanding information about their healthcare options so that they understand their choices and can participate in decisions about their care. Patients are also an important element in keeping healthcare costs down and improving outcomes. Providing patients with accurate and up-to-date information and guidance rather than just data will help them make better decisions and better adhere to treatment programs. Data is growing and moving faster than healthcare organizations. The 80% of the medical data is unstructured and is clinically relevant. This data resides in multiple places like individual EMRs, lab and imaging systems, physician notes, medical correspondence, claims, CRM [9] systems and finance. Getting access to this valuable data and factoring it into clinical and advanced analytics is critical to improving care and outcomes, incentivizing the right behavior and driving efficiencies.

Healthcare organizations are leveraging big data technology to capture all of the information about a patient to get a more complete view for insight into care coordination and outcomes-based reimbursement models, population health management, and patient engagement and outreach. Existing analytical techniques can be applied to the vast amount of existing (but currently unanalyzed) patient related health and medical data to reach a deeper understanding of outcomes, which then can be applied at the point of care. Ideally, individual and population data would inform each physician and her patient during the decision-making process and help determine the most appropriate treatment option for that particular patient [10]. The millions of users search the internet regarding health information.

5. THE SYSTEM MODEL

The proposed system contains several modules. These modules are listed as follows.

a. System Configuration
b. Data set management
c. Analysis on health care system
d. Report generation and decision making

The big data analytics must be performed in the The whole process of the modules can be explained in the system architecture. The system architecture as shown in the Fig 2.

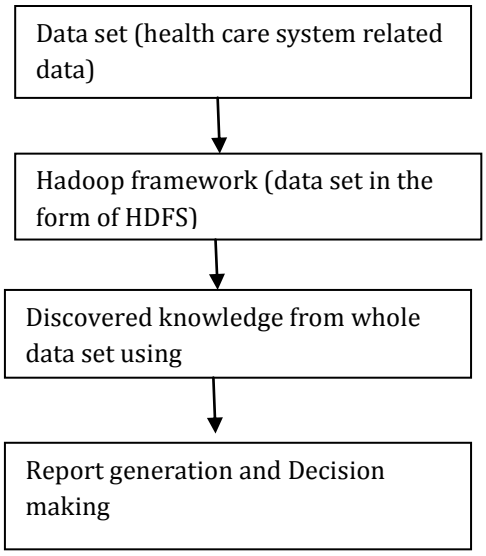

Fig -2: Health care big data analytics system

5.1 System Configuration

The system is configured for big data analytics by installing hadoop framework. The hadoop framework contains the several components. The Hadoop Distributed File System (HDFS) is used for storing the health care system data. The Hive and MapReduce is used for performing analysis. The Hadoop contains the name node, data node, job tracker, task tracker, secondary name node. The name node is the master node of HDFS. It is used to store the Meta data information. The data node is used for storing data blocks in hdfs. The data nodes are run on the same servers as task trackers. The job tracker is used to send job request to the task tracker. The task tracker performs the task that is assigned by the job tracker.

5.2 Data set management

The health care system data can be stored in the form of HDFS. The several millions of data can be used for analysis. If data set is high, can get a more accurate decision. These massive collections of health care data are handled by Sqoop tool. It is an open-source tool that allows users to extract data from a relational database into Hadoop for further processing.

5.3 Analysis on health care system

The analysis can be performed by either MapReduce or Hive_QL technique. The MapReduce contains two functions that is map and reduce. MapReduce performs the job that is specified in the mining algorithm. Finally it discovers the knowledge from the whole data set. The map function is used to mapping the input in the data node. The reduce function is used to performs the mining activity.

In Hive-QL technique, complex query can be used for discovering knowledge from the large data set. Hive-QL performs analysis with minimum amount of time. In hive, data is stored in the form of table but MapReduce takes raw data. Both techniques can give the more accurate decision with minimum amount of time

5.4 Report generation and decision making

The decision making is based on analysis and immediate report generation. The generated report contains details for future action. Report can be generated in the form of graph. From the graph, can take a decision for the action.

5.5 Advantages of proposed system

By digitizing, combining and effectively using big data, healthcare organizations ranging from single-physician offices and multi-provider groups to large hospital networks and accountable care organizations stand to realize significant benefit.

1. Detecting diseases at earlier stages when they can be treated more easily and effectively
2. Each patient record can be maintained efficiently.
3. Numerous questions can be addressed with big data analytics. Certain developments or outcomes may be predicted and/or estimated based on vast amounts of historical data.
4. Accurate result can be generated based on the big data technology
5. Analysis time can be minimized by using Hive-QL technique.

6. CONCLUSION

The proposed system can be handle the large amount of health care system data using big data technology. The Hive-QL technique is used for performing analysis in the large amount of health care system data. The proposed system gives more accurate result. It takes less amount of time to perform analysis and decision making. The patient retrieves information more efficiently and effectively. The system gives correct answers for the patient queries.

ACKNOWLEDGEMENT

I would like to show my gratitude to the (S.Selva ganesan HOD of Information Technology, J. J.College of Engg., & Tech.,) for sharing their pearls of wisdom with us during the course of this research, and I thank 3 “anonymous” reviewers for their so-called insights. I also immensely grateful to E.Subramanian, P.Dinesh Kumar for their comments on an earlier version of the manuscript, although any errors are our own and should not tarnish the reputations of these esteemed persons.

REFERENCES

[1] W. Raghupathi, S. Kudyb, “Data Mining in Health Care Improving Efficiency and Productivity,” in Healthcare Informatics, 2010.

[2] C.Burghard, “Big Data and Analytics Key to Accountable Care Success,” in IDC Health Insights, 2012.

[3] A. Dembosky, “Data Prescription for Better Healthcare,” in Financial Times, 2012, pp. 19-223.

[4] B. Feldman, E. Martin, T. Skotnes, “Big Data in Healthcare Hype and Hope,” 2012, pp. 360-443.

[5] L. Fernandes, M. O’Connor, V. Weaver, “Big data with bigger outcomes,” December 2011, pp. 38–42.

[6] G. Eason, “Transforming Health Care through Big Data Strategies for leveraging big data” in the health care industry, 2013.

[7] A. Frost, S. Sullivan, “Big Data Reducing Information Technology Complexities” in HealthcareOrganizations, vol.A 242, 2011, pp. 529-551.

[8] A. Dembosky, “Data Prescription for Better Healthcare,” in Financial Times, December 12, 2012.

[9] B. Feldman, M. Martin, T. Skotnes, “Big Data in Healthcare Hype and Hope,”2012, pp.220-225.

[10] L. Fernandes, “Transforming Health Care through Big Data Strategies,” in AHIMA 2012, pp. 38-42.

BIOGRAPHIES

T.Prinba Jothi, completed Bachelor degree and master degree in Information Technology under the Anna University.