Survey on Architecture of Healthcare based Big Data

Jayashri Hosur , Dr.Ashok Kumar R

Abstract— Healthcare industry is the one which constantly undergoes the changes with respect to medical dimensions. Healthcare industry involves very huge data which is a raw data so it is very difficult to analyze the raw data and produce proper result. As this task is very critical because these days medical field is using advance technologies like IoT and other computer related tools for finding particular disease of the patient. That generates very huge result so it is very important for patient as well as doctors to find the particular disease from the raw data. In this paper, a survey is made on different architecture of Healthcare based Big data. Also the tools and technologies used in each architecture. The framework of big data analytics for healthcare industry which mainly contains how it gathers the data from different places such as from different branch, from IoT, from electronic devices..

IndexTerms— Big data, Health care Industry, Machine learning, Iot.

I. INTRODUCTION

Big data is the technology which describes any voluminous amount of unstructured, semi structured and structured data. The important challenge for Big data is to store the data and process the data. Big data contains the 7 V's as below,

- Volume which indicates the quantity of the data i.e. what amount of data we have. These days the data generation is growing rapidly due to the use of IoT, computerized tools
- Velocity which indicates the speed in which data is accessible
- Veracity which indicates the accuracy i.e. the data analyzed is how accurate
- Variety which indicates the kinds of data such as structured, semi structured and unstructured
- Variability which indicates how the same kind of data behaves in different places
- Visualization which indicates visualizing the data in the form of graph, pie chart which makes the understanding simpler and clear
- Value is the last one which is the end product that can be obtained by addressing all other 6 V's. [7]

Healthcare industry is one of the notable area for experimental and combinative applications of different technologies over

several years. The healthcare industry is huge an industry which contains many cultivation that gives universal healthcare results on the basis of products and services. Few of the cultivation depends on collection of the data, data storage and communication technologies.

Every day we are seeing very huge data is generated by healthcare industry, and also by weather forecasting and weather prediction. Many such application so it is very difficult to get the useful information from these data as this data is raw data i.e. it is an unstructured form of the data. It is impossible for human to read the data and check for useful information. So we have to go for analytics. That can be achieved in systematic way using various analytics tools provided by Big data.

After comparing the big data analysis with business sectors in terms of the healthcare and also the health sector which is still in its early stages due to various reasons. The crucial key challenge is to accommodate very high volume, velocity and variety of data produced by healthcare industry and also finding the data from different sources such as from different branch, from different electronic system, financial system, administrative systems and genomic data for the analysis. After gathering the data next crucial part is how to get the useful information from these data [6]

Big data in healthcare is nothing but the electronic health data groups which are huge and very complex and complex to handle with software, hardware and data management tools. Big data in healthcare system is huge not only in terms of its volume but also related to the different type of the data. The entire data corresponding to patient healthcare is captured by big data. [5]

Big data computing is divided into two paradigms such as batch oriented computing and stream computing. The stream computing is also called as real time oriented computing. In batch computing it will first collect the data and then store this data and finally process the data into batches to produce the result. The example for batch processing is Apache Hadoop. In Hadoop the time of output generation depends on the volume of the input taken but in stream computing it continuously takes the input and generates the output so it depends on the velocity of the data. The example for stream computing is Apache storm [3]

By analyzing the big data we can group the patient with high risk and high cost in one group and low risk and low cost in another group. It also help to predict some of the diseases. Following are few of the pros those are obtained by Big data analytics in the Healthcare industry.

- Clinical operations: important researches are going on to detect the particular disease and treat the disease with less cost
- Public Health: analyze the disease pattern and try to improve the public health by using the analyzed data

International Journal of Research in Advent Technology, Vol.7, No.6S, June 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

- Remote monitoring: analyzing with the large volume of real time data to monitor and predict the event
- Evidence based medicine: gather heterogeneous data and then combine this and then analyze for the future forecasting

Research and development: predictive models in the field of research to analyze the data and also to make the faster decision.[8].

II. Existing Architecture For Analysis of Healthcare Based Big Data

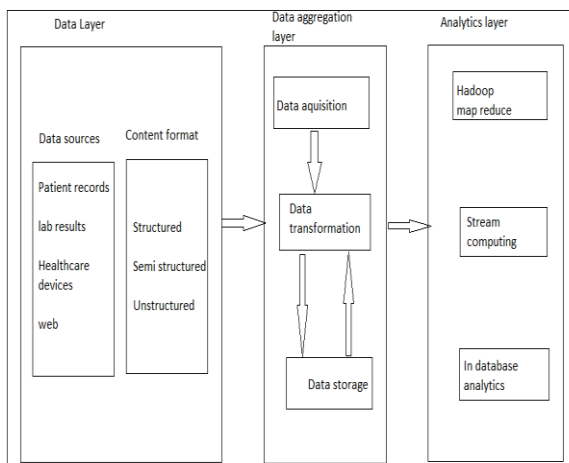The general Big data analytics architecture as shown in Figure 1,

Figure 1: General architecture of Big data analytics

It mainly consists of three layer such as data layer, data aggregation layer and analytics layer.

- Data layer: it has data sources like patient records, lab results, Healthcare devices, and web and from social media. These sources may be in the form of structured, semi structured and unstructured.
- Data aggregation layer: It mainly concentrate on data acquisition, data transformation and data storage
- Analytics layer: it uses either Hadoop map reduce technology to analyze the data or it may also use stream computing such as Apache storm for analyzing the data.

The authors Van-Dai Ta, Chuan-Ming Liu [3] have used the architecture as shown in Figure 2 for the analysis of Healthcare data,
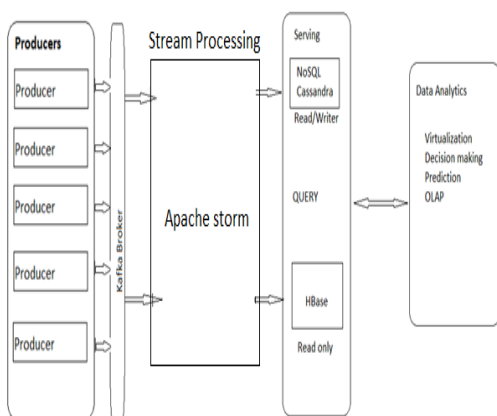

Figure 2: Healthcare analytics system architecture

- Apache Kafka is publish-subscribe messaging system which is formulated for rapid, expandable, and long lasting. A single Kafka broker can manage hundred megabytes of write and read per second from many other clients
- Apache storm is like a Hadoop, in Hadoop data processing is depends on job but in Apache storm data processing is depends on topologies.
- Examples of tuples
  e.g. <tweeter_name, tweet_on>
  <"XYZ1", "High fever and cough"> <"XYZ2", "Sore throat, and headache">
  e.g. <patient_name, patient_id, Date, Time, test_carried, Report>
  <ABC, ABC001, 9112018, 1:30:20, H1Nl, positive>
  < XYZ, XYZ001, 28012016, 12: 33: 50, H1Nl, negative>

The author Sunil Kumar and Maninder Singh [1] have used the architecture as shown in Figure 3 for the analysis of Healthcare data
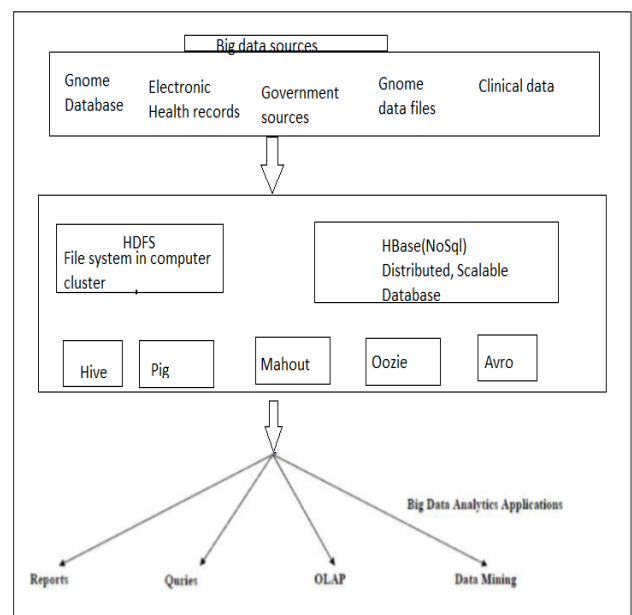

Figure 3: Conceptual architecture of big data analytics for health informatics

The data sources can be from EAR, Gnome database, government sources, clinical data and is stored into Hadoop distributed file system then using the Map Reduce algorithm to process the data.

They have proposed the conceptual architecture for solving healthcare related problems in the big data using the Hadoop technology. By combining the big data and healthcare analytics we can get the treatment which are effective to particular patient.

The authors Prashant Johri, Tanya Singh [7] have used the architecture as shown in Figure 4 for the analysis of Healthcare data

International Journal of Research in Advent Technology, Vol.7, No.6S, June 2019
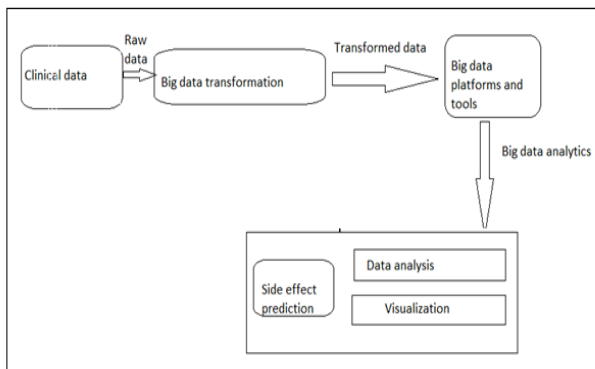*E-ISSN: 2321-9637*
Available online at www.ijrat.org

Figure 4: Conceptual architecture of big data analytics

This architecture also using Map Reduce algorithm to process the data and using the Hive database and HDFS for storing purpose. They have taken the dataset from the site http://tanlab.ucdenver.edu/DSigDB. First they are loading the dataset into HDFS.Using Map Reduce algorithm they are processing the data using eclipse tool.

Hive provides the query similar to SQL which are called as HiveQL, to create the table in Hive it uses the CREATE TABLE statement and then input data is loaded using LOAD DATA.

Below is one of the query to find the side effects in patient database.

Select State_of_patient, Side_effects, Count (Sideffects) from HealthCareDatabase where Side_effects like "%Moderate%" group by State_of_patient, Side_effects

III. COMPARISON OF EXISTING ARCHITECTURE

Table 1 shows the comparison of some of the features when architecture uses Hadoop and Storm.

Table 1: Hadoop and Storm Comparison

Features	Hadoop	Storm
Data processing	Batch oriented	Real time oriented
Data handling	Jobs	Topologies
Performance	High latency	Low latency
Database compatibility	SQL, No SQL	No SQL

Table 1 shows the comparison between features when architecture uses Hadoop and Storm.Hadoop is a batch oriented data processing and Storm uses real time oriented

data processing.Data handling is done by uisng jobs in Hadoop where as is it achieved by topologies in Storm.The perfomance is good in Hadoop compared to Storm.

IV. CONCLUSION

With the rapid growing of healthcare data, it is very important to analyze the data to provide proper treatment to the patient and also to detect the particular disease. We have existing architecture to analyze the healthcare data using Hadoop and Storm, among these two Hadoop is one of the best architecture by referring the various paper. Performance is good when we use Hadoop by comparing these two architecture..

REFERENCES

[1] Sunil Kumar and Maninder Singh, "Big Data Analytics for Healthcare Industry: Impact, Applications and Tools" in BIG DATA MINING AND ANALYTICS March 2019, 2(1): 48-57
[2] Ramesh Mande and G.JayaLakshmi, "Leveraging Distributed Data over Big Data Analytics Platform for Healthcare Services" Proceedings of the 2nd
[3] International Conference on Trends in Electronics and Informatics (ICOEI 2018) IEEE Conference Record: # 42666; IEEE Xplore ISBN: 978- 153863570-4
[4] Van-Dai Ta, Chuan-Ming Liu "Big Data Computing in Healthcare Real-Time Analytics" in 2016 IEEE International Conference on Cloud Computing and Big Data Analysis
[5] [Minerva Panda and Syed Mohd Ali," Big Data in Health Care: A Mobile Based Solution" 2017 International Conference on Big Data Analytics andComputational Intelligence (ICBDACI)
[6] A. Gandomi and M. Haider, Beyond the hype: Big data concepts, methods and analytics, Internatio Journal of Information Management, vol. 35, pp. 137–144, 2015.
[7] [A. Gandomi, and M. Haider, "Beyond the data concepts, methods, and analytics", International Journal of Information Management, vol.35, no.2,
[8] pp.137-144, 2015.
[9] Prashant Johri, Tanya Singh, "Vitality of Big Data Analytics in Healthcare Department" in 2017 International Conference on Infocom Technologies
[10] and Unmanned Systems (ICTUS'2017)
[11] Hardik B. Patel and Snehal Gandhi, "A Review on Big Data Analytics in Healthcare using Machin Learning Approaches" in Proceedings of the 2nd International Conference on Trends in and Informatics (ICOEI 2018) IEEE Conference Record: # 42666; IEEE Xplore ISBN:978-1-5386- 3570-4