

A Survey on Cost Reduction & Health Care Analysis Using Hadoop

Akshay Wadkar¹, Rohan Bedre², Akshay Adagale³, Anantha Subramaniam⁴

U.G. Student, Department of Computer Engineering, JSCOE, PUNE, Maharashtra, India¹

U.G. Student, Department of Computer Engineering, JSCOE, PUNE, Maharashtra, India²

U.G. Student, Department of Computer Engineering, JSCOE, PUNE, Maharashtra, India³

U.G. Student, Department of Computer Engineering, JSCOE, PUNE, Maharashtra, India⁴

ABSTRACT: In today's world there has been a dramatic increase in data, this data is found from various sources which are used by various people from different corners of the world. We know that this data is very important in some manner or the other. But as we know data is found in various format like images, documents, audio, video ,etc. One important thing is that data obtained can be of two types Structured as well as Unstructured. But for processing we need Structured data only. But due to this we cant pretend to loose unstructured data even that is important for us. Our system will solve this problem it basically states that first we accept all data as it comes irrespective of structured/unstructured and then sort it according to needs.

KEYWORDS: Big Data, Medical Analysis, Data Analytics , HIVE, BIRT Reporting, Sqoop.

I. INTRODUCTION

Big data analytic is a growth area with the potential to provide useful insight in health care[1]. While many dimensions of big data still present issues in its use and adoption, such as managing the volume, variety, velocity, veracity and value. Volume is nothing but the quantity of generated and stored data, Variety is the type and nature of the data and Velocity is refers to the speed at which the data is generated and processed to meet the demands and challenges It has provided tools to accumulate, manage, analyse , and assimilate large volumes of disparate, structured, and unstructured data produced by current health care systems[1].

Also the accuracy, integrity, and semantic interpretation are of greater concern in clinical application [2]. However, such challenges have not deterred the use and exploration of big data as an evidence source in health care. This drives the need to investigate health care information to control and reduce the burgeoning cost of health care, as well as to seek evidence to improve patient outcomes. We found that there was a huge and rapid increase in data coming from all sources [1]. But we had our specific research on Medical Analysis and in this we saw that, the data produced here is increasing huge day by day our major intent was not to loose the data.

This study used a systemic review methodology to create a categorization of big data use in health care. The results indicate that the natural classification is not clinical application based, rather it falls into four broad categories: administration and delivery, clinical decision support consumer behavior and support services.

Storing was a problem for this BIG DATA, so then on further survey we found that using Hadoop can solve this problem. Big data analytics provides great potential for health care operations to increase patient safety and quality health care while reducing medical errors and costs. This report gives an insight of how we can uncover additional value from the data generated by health care. Large amount of heterogeneous data is generated by hospitals. But without proper data analytic methods these data became useless. With the development of smart devices more and more public health data can be collected from various sources and can be analyzed in an unprecedented way.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 11, November 2016

Big Data Analytic using Hadoop plays an effective role in performing meaningful real-time analysis on the huge volume of data and able to predict the emergency situations before it happens.

Healthcare is such a field where there is a rapid growth in data day by day [3]. For example even if we consider a normal hospital in any local area we will see that there are at least 45 to 70 patients visiting daily because medical is such a field where people will be affected by a disease in some or the other way. So this was for a single day, if we consider it for months and years there will be a huge data that wont reside on machines which are located at local clinics or hospitals. Data will medicines provided or the treatments given to them, their reports such as pathology, diabetes , blood pressure reports, etc.

II. RELATED WORK

The system covers two different domains one is the Business Intelligence followed by Analysis On Big Data. Starting from Business Intelligence, [1] we saw that there is a rapid increase in data which needs to be managed as we discussed earlier it is not possible to neglect data because the data which we get is very important. So firstly we will understand what is Big Data. A simple example will show you what is Big Data.

Example:

- 1) For a local machine having 1TB hard disk 1001GB will be Big Data.
- 2) For Servers holding 1PB of space 1500TB will be Big Data.

So we saw what is Big Data ,[4] so there should be a solution for preserving this data because we know how important the data is. So our system will solve this problem. The ultimate solution will be dividing or distributing the data into various data storage devices. For example 1TB can be split into 4 machines of each 250GB or more. So this was the first part which was covered. Now we will see how it is related to Business Intelligence. See a normal 1Tb hdd will cost you up to 5000INR so getting a new drive every time is not possible. So our solution of distribution will come in place here.

Now we will discuss about the second domain that is the Analysis On Big Data. Our system majorly focuses on this Domain. The basic idea is collection of huge data sets, then sorting it and firing queries on it which will give us outputs in the form of reports.

III. TECHNIQUE

HDFS: Hadoop File System was developed using distributed file system design. It is run on commodity hardware. Unlike other distributed systems, HDFS is highly fault tolerant and designed using low-cost hardware. HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines. These files are stored in redundant fashion to rescue the system from possible data losses in case of failure. HDFS also makes applications available to parallel processing.

YARN: It is the Architectural Center of Enterprise Hadoop and a part of the core Hadoop project, YARN is the architectural center of Hadoop that allows multiple data processing engines such as interactive SQL, real-time streaming, data science and batch processing to handle data stored in a single platform, unlocking an entirely new approach to analytics.

MapReduce:Hadoop MapReduce is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. A MapReduce *job* usually splits the input data-set into independent chunks which are processed by the *map tasks* in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the *reduce tasks*. Typically both the input and the output of the job are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks. The MapReduce framework consists of a single master Job Tracker and one slave Task Tracker per cluster-node. The master is responsible for scheduling the jobs' component tasks on the slaves, monitoring them and re-executing the failed tasks. The slaves execute the tasks as directed by the master.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 11, November 2016

IV. CONCLUSION

We will be efficiently developing our system accordingly to the aspects stated above. Our system will successfully generate reports which will be as per needs. This implementation will be widely used by all medical fields and also has a wide scope in future for other sectors including public as well as private.

REFERENCES

- [1] Hao Zhang, Gang Chen, Member, IEEE, Beng Chin Ooi, Fellow, IEEE, Kian-Lee Tan, Member, IEEE, and Meihui Zhang, Member, IEEE, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 7, JULY 2015, In-Memory Big Data Management and Processing: A Survey.
- [2] Javier Andreu-Perez, Carmen C. Y. Poon, Robert D. Merrield, Stephen T. C. Wong, and Guang-Zhong Yang, Fellow, IEEE, IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. 19, NO. 4, JULY 2015, Big Data for Health.
- [3] Marco Viceconti, Peter Hunter, and Rod Hose 10.1109/JBHI.2015.2406883, IEEE Journal of Biomedical and Health Informatics, Big data, big knowledge: big data for personalised healthcare.
- [4] TECHNOLOGY AND TRENDS TO HANDLE BIG DATA: SURVEY Sunaina Sharma ME IT, Department of Information technology UIET, Panjab University Chandigarh, India E . .Veenu Mangat Professor, Department of Information technology UIET, Panjab University Chandigarh, India
- [5] Efficient Heart Disease Prediction System using Decision Tree Purushottam Research Scholar (PhD.), R.G.T.U. Bhopal (M.P), India. puru.mit2002@gmail.com Prof. (Dr.) Kanak Saxena, Prof & Head, Department of Computer Application, S.A.T.I. Vidisha (M.P), India. kanak.saxena@gmail.com Richa Sharma Assistant professor Amity University Uttar Pradesh Noida, India richa.sharma@gmail.com.
- [6] Using Data Science & Big Data Analytics to Make Healthcare Green Nina S. Godbole, Green IT Professional, Certified Information Privacy Professional and Healthcare Researcher, Pune, India ninagodbole@yahoo.com John Lamb, PhD, Adjunct Faculty Mathematics Department, Pace University, Pleasantville, NY, USA jlamb@pace.edu.
- [7] Analysis of Multi-diseases using Big Data for improvement in Healthcare Asif Adil Department of IT/CS BFIT Dehradun, India Hushmat Amin Kar Department of IT NIT Srinagar, India Rajendra Jangir Department of IT NIT Srinagar, India Shabir Ahmad Sofi Department of IT NIT Srinagar, India.
- [8] Big data/or Better Health Planning Prof. Jigna Ashish Patel Assistant Professor, CE Department, Institute Of Technology Nirma University, Ahmedabad Gujarat, India jignas.patel@nirmauni.ac.in Dr. Priyanka Sharma Associate Professor , MCA Department Institute of science & technology for Advanced Studies & Research, VV Nagar Gujarat, India pspriyanka@yahoo.com.
- [9] Improving Patient Matching: Single Patient view for Clinical Decision Support using Big Data Analytics Reena Duggal¹, Sunil Kumar Khatri², Balvinder Shukla³ 1-3Amity Institute of Information Technology Amity University Uttar Pradesh, Noida, India reena.duggal@student.amity.edu 2skkhatri@amity.edu, sunilkkhatri@gmail.com 3bshukla@amity.edu .
- [10] An Assessment of Visualization Tools for Patient Monitoring and Medical Decision Making Max Craft, Bev Dobrenz, Erik Dornbush, Molly Hunter, Jordan Morris, Michelle Stone, and Laura E. Barnes University of Virginia, mac5xb, bgd5de, meh6zb, jrm4fg, mts3se, lbarnes@virginia.edu .
- [11] Big Data Driven Healthcare Services and Wearables Vladimir Hahanov¹, Volodymyr Miz² 1. CAD Department, Kharkiv National University of Radio Electronics, UKRAINE, Kharkiv, Lenin avenue 14, E-mail: hahanov@kture.kharkov.ua 2. CAD Department, Kharkiv National University of Radio Electronics, UKRAINE, Kharkiv, Lenin avenue 14, E-mail: mizvladimir@gmail.com.
- [12] Predicting the Risk of Readmission of Diabetic Patients using MapReduce, M Gowsalya, K Krushitha ,C Valliyammai .