Research Journal of Management Sciences _____ ISSN 2319–1171
Vol. **6(9),** 29-32, September **(2017)** *Res. J. Management Sci.*

Review Paper

Big data analytics to empower rural masses of India - a step towards digital India program

Veena Madaan[1,2*] and Royal Madan[3]
[1]Dept. of Management, ICFAI University, Raipur, CG, India
[2]BIT, CSVTU, Bhilai, CG, India
[3]Mechanical Engineering, Rungta College of Engineering and Technology, Bhilai, CG, India
vin_sam2007@yahoo.com

Available online at: www.isca.in, www.isca.me
Received 10th April 2017, revised 20th August 2017, accepted 4th September 2017

Abstract

Data in a Big data can be of any kind small, big or structured, unstructured, but we called it as big data when the data purchased is beyond to the capacity or beyond to the processing power in comparison to the space available. The developing countries like India, faces big problems in the health care and these problems are quite solemn in the rural areas ,where the treatment expenses is high, less equipment's and unavailability of skilled doctors. With the help of electronic media reports, looking at the given doctor's prescription for a particular disease, a doctor can help a patient to cure from the disease in fewer trials or even in the first visit. In this paper, we analyze and reveal the benefits of Big Data Analytics for rural masses, in the applications of Healthcare and agriculture business, where the data flow to and from is in massive volume, by using the method called Hadoop. Also, National Rural Comprehensive Information Service Platform provides agriculture product markets and agriculture technology information services directly to the farmers and solves last minute problems in rural and agriculture information.

Keywords: Big data analytics, hadoop, rural development, Healthcare.

Introduction

A Big Data is a collection of large and complex data sets which are difficult to process using common database management tools like Oracle or traditional data processing applications. Massive amounts of data are collected across social media sites, mobile communications, business environments and institutions at an alarming velocity, volume and variety. In order to efficiently analyze this large quantity of raw data, the concept of Big Data was introduced. This new concept is expected to help education in the near future, by changing the way they the e-learning process, by encouraging the interaction between students and teachers, by allowing the fulfillment of the individual requirements and goals of learners[1]. Designed three-step system architecture for a consortium of universities, based on actual software solutions, having the purpose to analyze, organize and access huge data sets in the Cloud environment To extract meaningful value from big data, you need optimal processing power, analytics capabilities and skills. The major challenges associated with big data are Capturing data, Curation, Storage, Searching Sharing Transfer, Analysis, and Presentation.

The large Indian health care system needs to bind the healthcare's "big data" and examine a complex set of data, which comprised of electronic medical records and sensor data. This makes the clinicians to access and analyze healthcare big data to ascertain quality, determine best practice, and assess numerous treatment strategies and identifying patients at risk.

The applications running on Hadoop clusters are increasing day by day. This is due to the fact that organizations have found a simple and efficient model that works well in distributed environment. This model is developed in such a way that it enables to work efficiently on thousands of machines and massive data sets using commodity hardware. HDFS and Map Reduce is a scalable and fault-tolerant model that conceals all the complexities for Big Data analytics. Since, Hadoop is becoming increasingly popular, understanding technical details become essential. Map Reduce engine uses Job Tracker and Task Tracker that handle monitoring and execution of job. HDFS a distributed file-system which comprise of Name Node, Data Node and Secondary Name Node for efficient handling of distributed storage purpose.

Today, digitization and big data analytics penetrate all areas of life and create innovative method of working, communicating and cooperating. Connecting individuals, enterprises, devices and governments enables easier transactions, collaboration and social interaction and results in enormous accessible data sources. Not only do humans turn into walking data generators[1] but, in addition, the interaction between objects – the so-called Internet of Things with sensors and IP-addresses – adds a multitude of data sources throughout organizations and society.

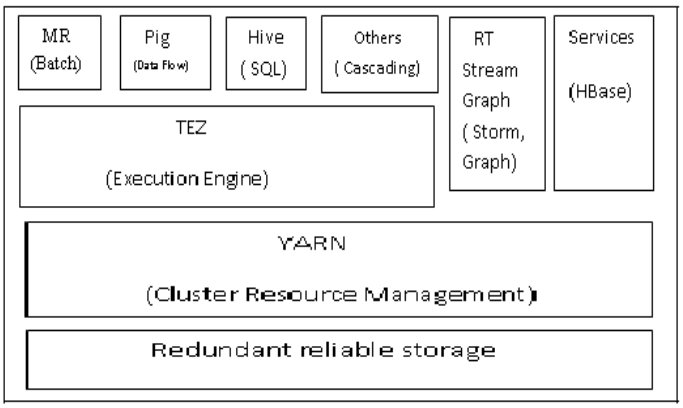
Computers incorporated into products like cars, vacuum cleaners and video consoles create massive amounts of digitized data and processes.

The objective of this paper is to enhance high quality health care without any discrimination on basis of gender, case, social status and economic status and also aims to ensure better health care to rural people and overcome the problems in the health care like expired drugs not administered to patients, fraud management in the health care. The proposed concept enables doctors, patients and staff to have role-based access to information on electronic health records .And also to protect farmers from being misled and impaired by wrong information, and ensure them enable to access information at the economic cost.

Current problems faced by rural people

Problems related to agriculture: Farmers get less money out of their products since the arrival of middlemen's. Fraud is very common between farmers when it comes to agriculture. Tradesperson are taking the benefit of their lack of education and awareness. They are hallucinated by their schemes, and hence get paid less.

When it comes to government schemes which give guidance to farmers in every aspect of their business, then at the last moment they got convinced by middlemen's. They apply for loans considering government schemes but the fact is that they did not get any amount of money, even though they applied for it the reason is that banks are not coopering with them when they come in front.



Problems related to healthcare: To overcome the inequity in healthcare system, the Government of India has launched the National Rural Health Mission (NHRM)[4] in 2005. The aim of mission is to make available effective healthcare to India's rural population. The thrust of this mission is on establishing a fully functional, community owned, decentralized health delivery system with inter-sectoral confluence at all levels, to ensure simultaneous action on a wide range of determinants of health such as water, sanitation, education, nutrition, social and gender equality.

Though it is hard to believe, but the fact is, compared to government health sector the private health sector has highly skilled doctors and the facilities are world-class, they have latest equipment's in their laboratory and innovative technology. Though, the aim of National Rural Health Mission (NRHM) is to provide effective health care to rural population, the hospitals located in remote villages are not equipped with well qualified doctors and sophisticated equipment due to several unknown reasons like the doctors may not be interested to reside in village and provide service to the rural population, lack of funds to procure and supply essential infrastructure to all the remote hospitals. To access the high-end private sector facilities, people spend huge amounts of money on treatment which in-turn affects their livelihoods and slowly poor people are becoming poorer. In such cases, the concept of Telemedicine along with big data analytics is the better alternative to treat the patients in rural areas.

Method available to handle big data

Characteristics of big data: Volume: As the data is coming in from many sources, in current times with the inception of social networking sites the volume of data generated is myriad it increases very rapidly from GB, TB to PB.

Velocity: Sending the data into the data server again and again with such a big volume decreases the processing speed of the system.

Variety: Big data as it comes from a great variety of sources and generally it has three categories: structured, semi-structured and unstructured. Structured data inserts a data warehouse already tagged and easily sorted but unstructured data is random and difficult to analyze .Unstructured data contains Word, PDF, Text, Media etc. Semi structured data does not conform to fixed fields but contains tags to separate data elements it contain data includes log files. Structured data contains relational data.

Veracity: Big Data Veracity refers to the biases, noise and abnormality in data. Is the data that is being stored, and mined meaningful to the problem being analyzed[2]. Veracity in data analysis is the biggest challenge when compared to other characteristics like volume and velocity.

Handloop: Apache software foundation Sponsored apache project named hadoop. It is open source framework written in java that allows distributed processing of large datasets across clusters of commodity hardware using simple programming models. Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

Its process and storage of extremely large data set. Two different tasks that hadoop performs are The Map Task and the reduce task. The Map tasks the first task, which takes input data and converts it into a set of data, where individual elements are broken down into tuples Key/value pairs).

The Reduce Task: This task takes the output from a map taskas input and combines those data tuples into a smaller set of tuples.

MapReduce: MapReduce is programming model or a software framework used in Apache Hadoop. Hadoop Map Reduce is provided for writing applications which process and analyze large data sets simultaneously on large multi node of commodity hardware clusters in a scalable, reliable and fault tolerant manner. Data analysis and processing uses two different steps namely, Map phase and Reduce phase. Map reduce divide and break the data in chunks which are processed by map phase parallel then reduce phase. The output generated by the map phase is known as intermediate result and also used as input for reduce phase[3]. It is used to perform the task of sorting, aggregation and to preserve the efficient storage structure[4]. Data are preferably refined using collaborative filtering, under the prediction mechanism of particular data needed by the user. The proposed method is enhanced by using the techniques such as sentiment analysis through natural language processing for parsing the data into tokens and emoticon based clustering. The process of data clustering is based on user emotions to get the data needed by a specific user.

The Hadoop Distributed File System (HDFS) is based on the Go ogle File System (GFS) designed to run on large clusters (thousands of computers) of small computer machines in a relia ble, fault tolerant manner. HDFS is a self-healing, distributed file system that provides reliable, climbable and fault tolerant data storage on commodity hardware. It works approximately similar to Map Reduce by distributing storage and computation across large clusters by combining storage resources that can scale depending upon requests and queries while remaining inexpensive and in budget. HDFS accepts data in any format like text, images, videos, etc regardless of architecture and automatically optimizes for high bandwidth streaming.

YARN (Yet another Resource Negotiator) the functionalities of yarn is split in two parts one is resource management and job scheduling/ monitoring. This idea generate two segment resource manager and per application master. In which Resource manager role is to distribute the resources among all the application in the system where as node manager behave like an agent who contain, monitor their resource usage and report the same to resource manager. On the other hand the per application master role is to negotiate with resource manager and work with node manager to execute and monitor the task[5].

Big data analytics in health care for rural masses: As the health care data growing drastically day by day, the infrastructural facilities must be enhanced in the health care center's to meet the needs of the rural people and necessary steps must be taken to store the vast amount data from the above said different sources for the future analysis. Thus the need for BDA (Big Data Analytics) arises, which provides clinical decision support through large amounts of data[6], personalized care by early detection and diagnosis before a patient develops disease symptoms, clinical operations with great accuracy, fraud management in the health sector.

Electronic medical records: Medical Experts agree that electronic medical records (EMRs) are a must for the better health care in India. But, at present only limited number of hospitals are maintaining EMR's, mainly because of beyond the budget, privacy issues, and the lack of one compatible, easy-to-use infrastructure. Maintain an electronic report of a problem helps many doctors to get an idea about the problem diagnose previously ad its solutions. Let say, if an X patient having Y symptom is treated with Z medicine, then if another patient with the same symptom, go to another doctor then doctor will do electronic data search for him and the result of that particular medication and then after getting an idea about the disease, a doctor can consult a patient well.

E-Health File: All the patient records which obtained after undergoing many tests like X-Ray report and CT scan etc. that depend upon the type of disease must be put online, so that any doctor can see the already performed tests online and can consult a patient well.

Detecting spreading diseases earlier: On getting all the data online a statistics can be easily performed to identify the particular problems or disease spreading in particular area. For example-after analyzing the data we come to know that at a particular area there are 70% of the people are suffering from Malaria, so, we will take instant corrective action to resolve the problem to stop it, considering the symptoms of the problem and stopping it for further transmission. In rural areas it is seen that, many patients are sufferings from the same problem, these statistical data will helps in vanquishing diseases.

Avoid unnecessary treatments: Doctors should avoid trial and error type of medication. Now a days, as we many problems are not getting resolved in first visit. Also, with the upbringing of new medicines in the market doctors trying it to test it on a patient. The right treatment should be suggested at the first visit only which avoids the disease to become more critical. Most of the issues arise with the incorrect diagnosis and wrong treatment during the early stages.

Tele medicine: The doctor can consult the patient without the patient's presence; with this the disease can be controlled before getting it more critical. The unavailability of doctors in rural areas has become a very big problem for rural people. Doctors are not interested in staying in rural areas, because of less facility, change of life style in comparison to urban areas, less scope and because of family.

Improving the treatment methods: Customized method of treatment to patient monitor the effect of medication frequently and based on the analysis dosages of medications can be changed for faster relief. Monitoring patients is vital signs to

provide proactive care to patients. The data generated by the patients who already suffered from the same symptoms help the doctors to make an analysis and provide effective medicines to new patients[7].

Big data analytics in agriculture business: To integrate information resources and systems in the pilot provinces, "Construction and Application of National Rural Comprehensive Information Service Platform", a national science and technology support program, was approved. This program will build the national rural and agricultural information service platform using cloud technology, and complete the nationwide sharing of information resources. It will also provide agriculture product markets and agriculture technology information services directly to the farmers and solve last minute problems in rural and agriculture information.

Agricultural information services, focused on providing value-added information to farmers, Agro-climatic information services, vocational training, agro-technology information services, and agricultural advice are all included in these information services. NRCISP used cloud computing as the infrastructure with which to integrate data, application, and service resources from provincial level information service platforms. Services are further combined and optimized so as to serve farmers in rural areas in the "one portal site and N platforms" model. NRCISP users also generate a lot of communication data, accessing data, log data, and so on. These data are very important for NRCISP's analysis of users 'behavior so as to provide better services.

Conclusion

Big data provides an opportunity for "big analysis" leading to "big opportunities" to advance the quality of life, or to solve the mysteries of the world. Digitalization of data will give a platform to rural people to grow into their business. Digitalization also; will give chance to the doctors staying in rural areas to interact with researchers, scientists and other doctors and their works.

Hadoop Map Reduce programming paradigm and HDFS are increasingly being used for processing large and unstructured data sets to increases the performance of complexity analysis. With the help of Hadoop the goal of effective citizen care management can be achieved by providing an effective data driven services to citizens by forcasting their needs based on the analysis of survey conducted among different social strata of citizens.

References

1. Loebbecke Claudia and Picot Arnold (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *Journal of Strategic Information Systems,* 24(3), 149-157.
2. Kumar Muni N. and Manjula R. (2014). Role of Big Data Analytics in Rural Health Care - A Step towards Svasth Bharath. *International Journal of Computer Science and Information Technologies*, 5(6), 7172-7178.
3. Ghazi Mohd Rehan and Gangodkar Durgaprasad (2015). Hadoop, MapReduce and HDFS: A Developers Perspective. *Procedia Computer Science,* 48, 45-50.
4. Subramaniyaswamy V., kumar Vijaya V., Logesh R. and Indragandhi V. (2015). Unstructured Data Analysis on Big Data using Map Reduce. *Procedia Computer Science*, 50, 456-465.
5. Archenaa J. and Mary Anita E.A. (2015). A Survey of Big Data Analytics in Healthcare and Government. *Procedia Computer Science*, 50, 408-413.
6. Logica Banica and Magdalena Radulescu (2015). Magdalena Using Big Data in the Academic Environment. *Procedia Economics and Finance*, 33, 277-286.