

## Cloud based framework to handle and analyze diabetes data C

*K. Vidhya<sup>1</sup> and R. Shanmugalakshmi<sup>2</sup>*

<sup>1</sup>Department of Computer Science and Engineering, KPR institute of Engineering and Technology, Coimbatore, India

<sup>2</sup>Department of Computer Science and Engineering, Government College of Technology, Coimbatore, India

Copyright © 2016 ISSR Journals. This is an open access article distributed under the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ABSTRACT:** Diabetes Mellitus is one of the growing vitally fatal diseases world-wide. Diabetes causes serious health issues invariantly to all. It will not bother about age, ethnic and also the racial group of a people. Diabetes is a Chronic Disease which is increasing rapidly due to the lack of awareness, change in urban culture, unhealthy foods, lack of physical activity and also due to hereditary. As per the statistical reports the impact of Type2 diabetes is very high comparing to Type 1 diabetes. Especially diabetes on children, adults, pregnant women and also people those who are suffering by other serious diseases needs to be monitored closely and their risks should be addressed specially through various researches and studies. Creating awareness and imparting knowledge about managing diabetes is very essential to safeguard our future world. Due to huge population, the volume, velocity and also the varieties of diabetic data increases tremendously. So for storing such large volume of data we need an elastically scalable environment such as Cloud. Cloud is a fast deployable and scalable platform very much suitable for accommodating huge amount of dynamic data. While handing those large size of data we will meet the problems of data synchronization, concurrency, job scheduling and fault tolerance. By applying the Hadoop-programming model the fore mentioned problems can be handled in a simple and efficient way. The dynamic and large sized health care data can be effectively stored and processed by using the proposed architecture where the risk factor (in terms of %) for type 2 diabetes is extracted based on the ratio of BMI (Body Mass Index) and age. This Cloud based framework helps the patients and physicians to access their data globally anywhere at any time.

**KEYWORDS:** Diabetes, Dynamic data, Data Analysis, BMI, Cloud Storage.

### 1 INTRODUCTION

As per the study conducted to asses awareness and knowledge about diabetes in four geographical regions of India such as Chandigarh, Tamilnadu, Jharkhand and Maharastra only 43.2 % of total study population had heard about diabetes. The urban residents had 58.5 % of awareness rate compared to rural residents of 36.8 %. About 46.7% of males and 39.6% of females are aware of diabetes. The 41.5% of people among the general population are aware of diabetes and among them 80.7% of people knew that prevalence of diabetes was increasing. Among the general and diabetic population 56.3% and 63.4% respectively were aware that diabetes could be prevented. Similarly 51.5% of general population and 72.7 % of diabetic population knew that the diabetes could affects other organs of the body [1][2]. Based on the assessment and knowledge score among general population, Tamilnadu had the highest score (31.7%) and Jharkhand has the lowest score (16.3%). The Indian Council of Medical Research India Diabetes study (ICMR-INDIAB study )says that India had 62.4 million people with diabetes in 2011 and by 2030 it will be increased to 101.2 million [1]. Based on the above study the knowledge and awareness about diabetes and its complication in human life span time is very poor. So the process of conducting various awareness and education programs related to diabetes is essential to the society.

The powerful weapon to fight against Diabetes Mellitus is awareness or knowledge. The knowledge and awareness help the people to realize the impact of diabetes and helps them to identify correct health care providers and their respective treatment. Due to the increased complication of diabetes over the various organs of the body, it is essential to assess the health state of the patients periodically and counseling them to understand the need for treatment at right time. Due to

increased population the size of the medical data also increases dynamically. Storing and analyzing such a big data is very much challengeable process nowadays. Researchers demands supportive mechanism to efficiently process large amount of data and hospitals The main challenges in processing large medical dataset is sufficient infrastructure to accommodate the data. Usually such a huge infrastructure requires large investment as capital. But the Cloud computing provides many on-demand virtual infrastructure in pay-per-use model.

In this model we can avail the dynamic infrastructure for data processing as per the increased or decreased size of the data sets. There are various cloud service providers offering services for large data processing as part of the cloud services. Map Reduce as a service is an main advantage in the field of bigdata analytics. But it has some limitations. Such as those services offered in the form of Platform-as-a Service, but there is no transparent information about the implementation and workflow details of MapReduce service. Similarly the user does not have control over the Map Reduce Software stack and its configuration. This causes the optimization, performance and compatibility problems. And also these services are always vendor-specific and are customized corresponding to their own infrastructure and also it locks the clients from preferring multiple service providers[R2]. So a development environment which has the vendor free exploration platform is essential for managing medical records of patients themselves and hospitals. The main objective here is to provide an innovative and cost effective solution to handle large and dynamic diabetes data and analyze them further for future prediction, spreading awareness among the community and preventing the loss of young minds due to diabetes.

## **2 DIABETES AND TYPES OF DIABETES**

### **2.1 UNDERSTANDING DIABETES**

As per the statement of WHO(World Health Organization),the biggest health concern faced by people nowadays is Diabetes. WHO defines diabetes as “a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces”. Insulin is a hormone that regulates blood sugar.A common effect of diabetes is Hyperglycemia or increased blood sugar[2]. There are very serious health issues due to diabetes including blindness, kidney failure, stroke and heart diseases. Type 1 diabetes is the nature when the body produces insufficient quantities of insulin. It is usually detected more in children. Type 2 diabetes occurs when the body does not effectively use the insulin produced. This is very frequently due to lack of physical activity, obesity or incorrect dietary habits. Gestational diabetes occurs among pregnant women. In about 90 percent of cases, it is Type 2 diabetes that people are suffering from. The occurrence of Type 2 diabetes or Diabetes Mellitus may be prevented or delayed by adopting a healthy lifestyle[8][9].

### **2.2 DIABETES IN INDIA**

The statistics report from the International Diabetes Federation (IDF), India has more diabetics than any other nation of the world. Currently there are 62 million people suffering by diabetes daily.This is the 10 million increase comparing to the statistics of 2011.The researchers says that by 2030 over 100 million people in India may have diabetes[11].

### **2.3 PREVENTING DIABETES MELLITUS**

The causes of Type 2 Diabetes can be delayed or prevented by balanced diet and exercise and by maintaining ideal Body Mass Index(BMI) is an important factor in diabetes prevention. Based on the age also ,the possibilities for diabetes attack can be predicted and can be prevented. Middle aged and Older adults are still having highest risk for developing type 2 diabetes. As per the statistics of CDC(Centers for Disease Control and Prevention ),adults aged 45 to 64 were the most diagnosed age group for diabetes. In the study conducted among 12,782 persons who underwent diabetes screening in the past year by a diagnostic lab ,2,713(21%) of people who were aged 20-29 suffering from high sugar levels[15][16].From this it is clearly understood that Type 2 diabetes has genetic causes and is affecting the youth. So the prevention of diabetes is essential one and can be achieved by technology assisted awareness methodologies which should reach the society in easiest way and also should help the health service providers to mange and manipulate data in effective ways[14][17].

### 3 CLOUD BASED FRAMEWORK FOR HANDLING DIABETES DATA

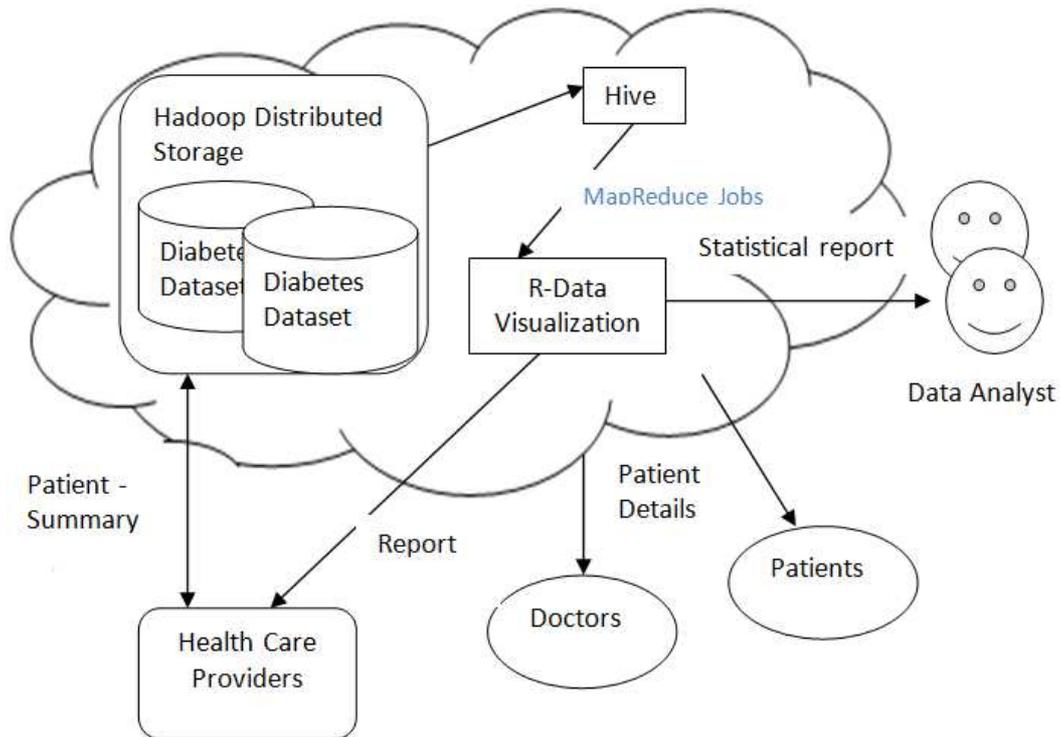


Figure 1: Cloud Architecture for handling diabetes data

The Hybrid Cloud architecture for handling large sized diabetes data is as shown in Figure1.

#### 3.1 ARCHITECTURE OVERVIEW

The proposed Cloud framework of efficient methods and technologies will make possible the collaborative investigation of diabetic risk factor and other diabetic related diseases(as part of future work) with health care providers of hospitals, doctors and patients. The main here is to avoid or minimize the risk of diabetes we have done a preliminary try by considering basic information of a person such as age, height ,weight .Based on the values the BMI(Body Mass Index) level of a person is calculated. Actual BMI is compared with the standard BMI and based on that the possible percentage of risk is calculated individually for both male and female.

The architecture mainly depicts four major part of handling medical data. Such as Distributed Data storage using Hadoop(HDFS) in Cloud by health care providers or hospitals ,SQL like interface for manipulation of data stored in Hadoop, of potentially manipulated Map Reduce Job to R data visualization tool. The statistical report gives the percentage of possibilities based on persons BMI level and age .In the Distributed Data storage module the diabetic datasets collected and maintained by healthcare providers are transferred to the cloud .In cloud ,the data sets are stored in HDFS. Here the Hadoop is dynamically deployed in virtual infrastructure provided by the public or private cloud providers. This architecture facilitates the transparent exposure of MapReduce workflow and its customization.

The Hive-SQL like interface receives the distributed chunks from HDFS and then it explores, structures and manipulates the data for calculating the risk factor based on the criteria posted. Here, for testing purpose 2500 persons age, weight, height are considered as an input and BMI ratio is calculated based on the standards. From the BMI value the level of risk is predicted for each person. The male and female candidates are segregated in two different clusters and are the possibilities for diabetes attack is analyzed individually. Finally the analysis reports were generated .Now the data analyst can perform various complex modeling on subsets of data. Similarly the doctors can view the history of the patients from cloud storage and can treat accordingly irrelevant to the mobility of the patients. At the same time the patients can view and retrieve their treatment history from cloud storage anywhere at any time and they can understand their level of risk for diabetes .

3.2 FLOW OF ARCHITECTURE

The below Figure: 2 shows the flow of data stored and manipulated in the proposed framework.

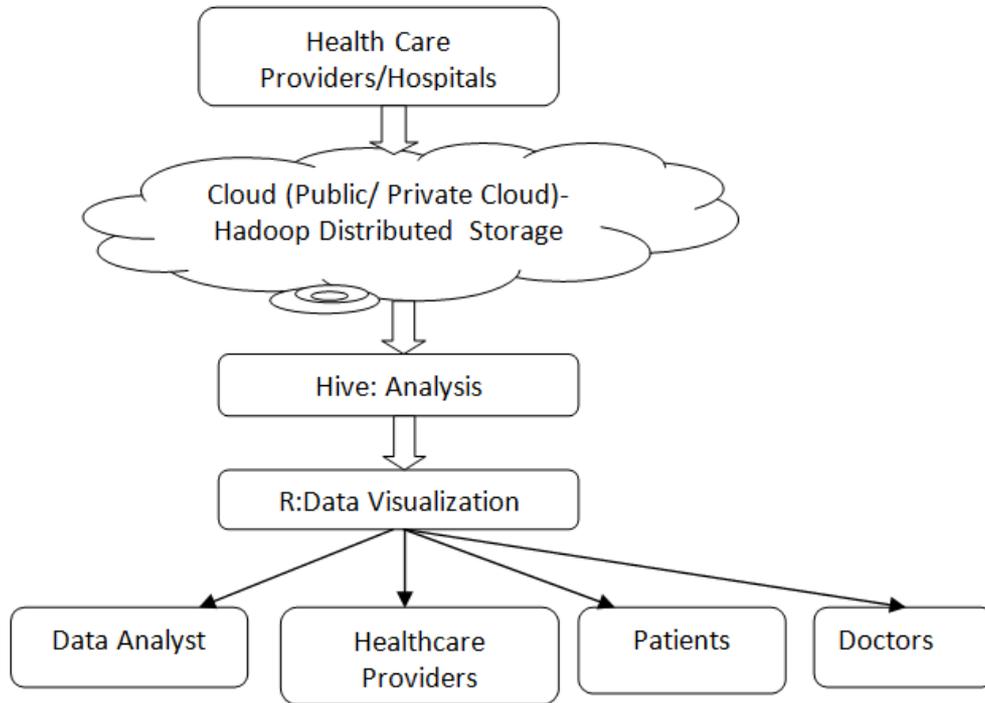


Figure 2:Flow of architecture

3.3 CLOUD STORAGE

Cloud storage enables users to remotely store their data and utilize the on demand high quality cloud applications without the burden of hardware and software management. Migrating data from our traditional storage to cloud storage offers convenient environment for storing and managing huge amount of data free from hardware complexities. Amazon EC2 and S3(Simple Storage Services) are well known public storage services. This internet based service provides large size of storage space and customizable computing resources [18].

3.4 HADOOP

When the data size exceeds the storage capacity of our available storage system, we have to distribute the data to multiple computers. The single framework which has the capacity of thousands of server is Hadoop. The HDFS (Hadoop Distributed File System) has a cluster system which stores the data chunks in multiple clusters and also it achieves reliable replication of data, fast fault detection and automatic recovery . Hadoop-Map Reduce programming model was proposed in 2004 by the Google, which is used in processing and generating large data sets implementation [19]. This model has the Mapper and Reducer interfaces to manipulate the large sized data.

3.5 HIVE-DATA ANALYSIS

The input from HDFS is fed to Hive to explore, structure and analyze the data. By querying the Pima dataset based on the criteria related to age, weight, height and BMI. Based on the query related the male and female data sets are segregated and analyzed as shown in sample Table 1 and Table 2(which is converted to an understandable simple form) . This results the risk level of both male and female individually.

Table 1: Female table

Female			
Age	STD BMI	ACT BMI	Ris Factor
20	63.77	68	28.2
25	59.25	52	50
30	52.25	68	100
35	48.95	65	100
40	59.25	52	50
45	52.25	58.5	40
50	62.9	54.6	50
55	66.65	58.6	50
60	59.25	68	60
65	48.95	58	60
70	70.45	78	50
75	66.65	57	60
80	66.65	51	100

Table 2 : Male Table

Male			
Age	STD BMI	ACT BMI	Risk Factor
20	59	52	40
25	48.95	58	60
30	52.25	68	100
35	48.95	65	100
40	48.95	55	40
45	66.65	58	60
50	70.45	63	50
55	66.65	60	40
60	59.25	52	50
65	62.9	54	60
70	52.25	55	20
75	59.25	52	50
80	55.7	61.5	50

### 3.6 R VISUALIZATION

The analyzed results from Hive are imported as a data frame and using ggvispackage (visualization package) for visualizing the analyzed result as graphs. Normally the level of glucose segregation and the usage of generated glucose become lower when the Body Mass Index (BMI) of a diabetic patient increases. The graph generated using R on the basis of age group and the corresponding graph has age, BMI and risk factor in x-axis and in y-axis. Statistical report as shown in Figure 3 and in Figure 4. The result can be viewed by health care providers or hospitals, patients, doctors and also by data analyst for further review and research.

4 RESULT

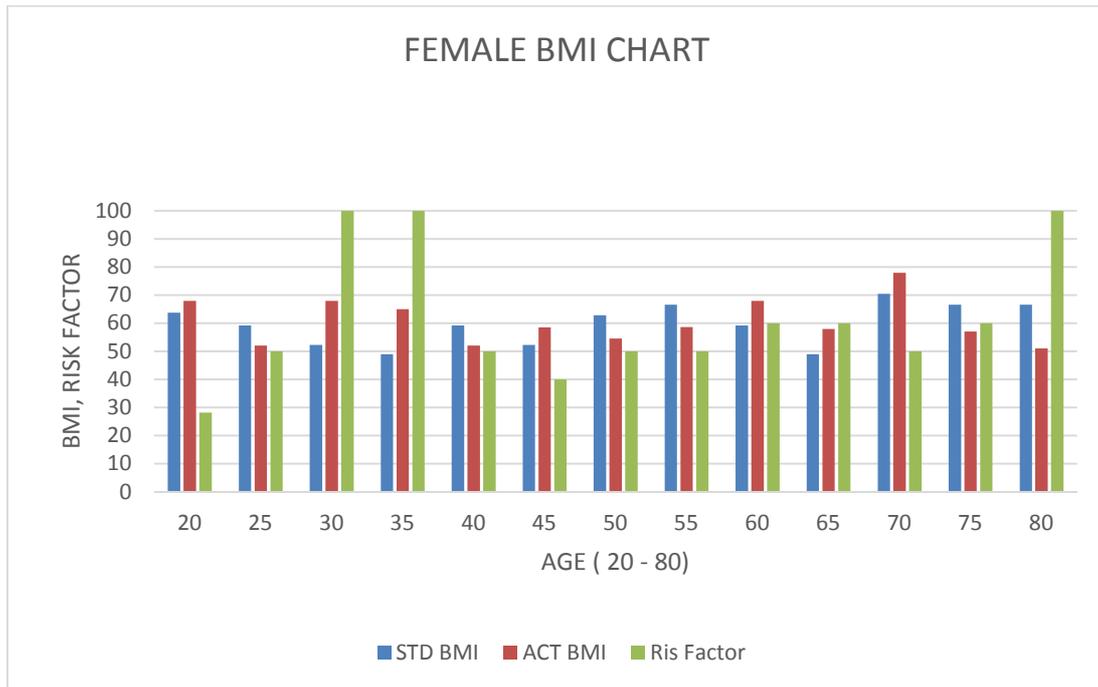


Figure 4: Male BMI and risk factor result

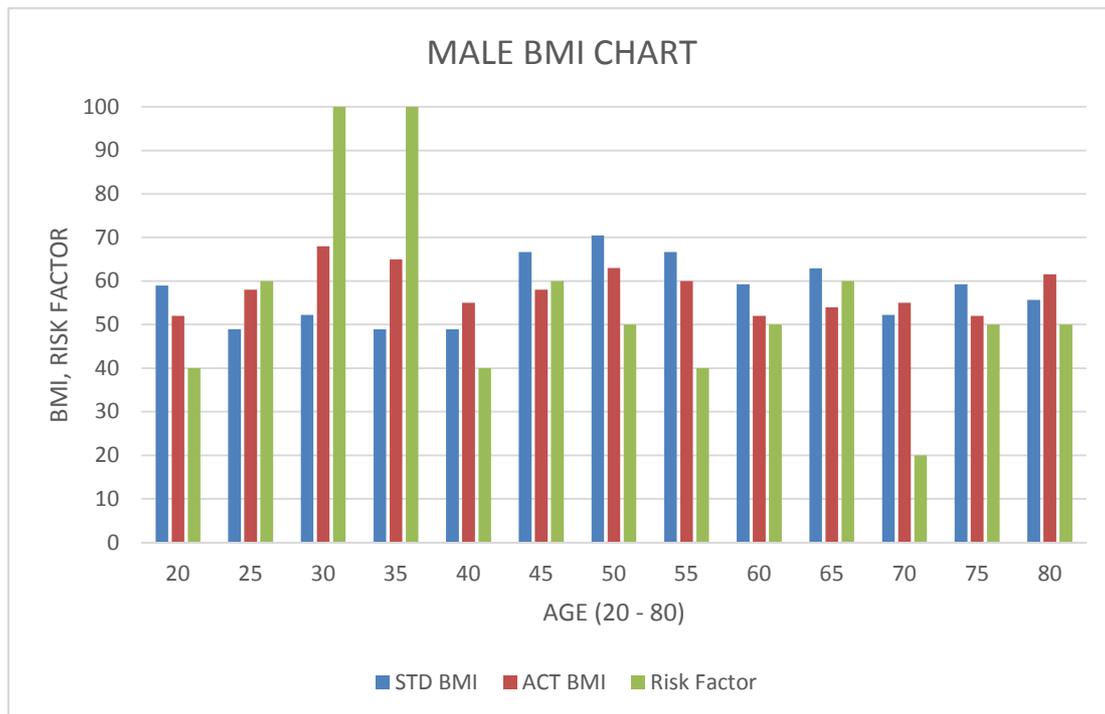


Figure 4: Male BMI and risk factor result

## 5 CONCLUSION AND FUTURE WORK

Currently, chronic disease like diabetes is growing constantly in society. Factors and reasons for the growth of this diabetes are various which includes improper knowledge of diabetic disease among the patients mainly living in rural areas, food habits. Most of the patients are unaware about the available health care centers for treatment. So if there is a global storage and retrieval infrastructure, people can get guidance at where ever they are. In order to effectively manage the huge dynamic diabetes data the proposed cloud based framework is very much suitable. It basically eliminates the pain of creating and maintaining heavy higher end infrastructure with more capital investment. Based on this framework a person who have BMI level greater than the threshold value of standard chart based on the age and their height, they are then categorized as male and female candidates. Based on the statistical report we obtained, male candidates have more possibility to be attacked by diabetes comparing to female candidates who have the same BMI level. Developing an secured cloud based pool with the improved result of above prediction for creating an awareness about diabetes among patients living in rural areas and also importing knowledge about available treatment in health care centers for diabetes will be the part of our future work.

## REFERENCES

- [1] Sujatha, "Prevalence of Diabetes in India" ,[Http://MapsofIndia.com](http://MapsofIndia.com)-June 30,2015
- [2] Advance and Emerging Opportunities in Diabetes Reaserch: A Strategic Planning Report of the DMICC
- [3] Anjana RM, Pradeepa R, Deepa M, Datta M, Sudha V, Unnikrishnan R, et al. Prevalence of diabetes and prediabetes (impaired fasting glucose and/orimpaired glucose tolerance) in urban and rural India: Phase I results of the Indian Council of Medical Research-India Diabetes(ICMR-INDIAB) study.*Diabetologia*. 2011; 54:3022–7.
- [4] [Http://www.diabetes-translation.org/](http://www.diabetes-translation.org/) Clinical Research to practice: Translational Research chapter in Advances and Emerging Opportunities in Diabetes Research: A Strategic Planning Report of the Diabetes Mellitus Interagency Coordinating Committee (DMICC).
- [5] Mohan D, Raj D, Shanthirani CS, Datta M, Unwin NC, Kapur A, et al. Awareness and knowledge of diabetes in Chennai – The Chennai Urban RuralEpidemiology Study [CURES-9] *J Assoc Physicians India*. 2005; 53:283–7.
- [6] Monani M, Nardini C, Mannucci E. Efficacy and safety of sodium glucose co-transport2inhibitor in type of diabetes: a meta-analysis of randomized clinical trials. *Diabetes Obes Metab* 2014; 30:1759-1768.
- [7] Murugesan N, Snehalatha C, Shobhana R, Roglic G, Ramachandran A. Awareness about diabetes and its complications in the general and diabetic population in a city in southern India. *Diabetes Res Clin Pract*. 2007; 77:433–7.
- [8] Patel V, Chatterji S, Chisholm D, Ebrahim S, Gopalakrishna G, Mathers C, Mohan V, Prabhakaran D, Ravindran RD, Reddy KS.Chronic diseases and injuries in India. *Lancet* 2011; 377:413-28.
- [9] [www.niddk.nih.gov/...Advances/Resource](http://www.niddk.nih.gov/...Advances/Resource) and infrastructure needs for diabetes research.pdf
- [10] [Http://emc.ornl.gov/publications/PDF/Population\\_Special\\_Needs.pdf](http://emc.ornl.gov/publications/PDF/Population_Special_Needs.pdf)
- [11] Wullianallur Raghupathi, and Viju Raghupathi, "Big data analytics in healthcare: promise and potential", *Health Information Science and Systems*, vol.2(3) pp. 2-10, 2014.
- [12] [Http://www.mapsofindia.com/my-india/india/prevalence-of-diabetes-in-india](http://www.mapsofindia.com/my-india/india/prevalence-of-diabetes-in-india)
- [13] Chanchal Yadav, Shuliang Wang , Manoj kumar , " Algorithm and approaches to handle large Data-A Survey, *International Journal of Computer Science and Network*, Vol 2, Issue 3, 2013
- [14] Ping ZHOU , Jingsheng LEI, Wenjun YE, " Large-Scale Data Sets Clustering Based on MapReduce and Hadoop", *Journal of Computational Information Systems* 7: 16 (2011)
- [15] Shreelola Hegde,Sonal T.G,Swetha M.S,Muneshwara M.S,Anil G.N, "Efficient Way of Managing the Data for Cloud Storage Applications Depending on the Internet Speed in Cloud Environment", *International Journal of Engineering Innovation & Research* Volume 3, Issue 1.
- [16] [Http://Timesofindia.indiatimes.com/city/bengaluru/Diabetes-now-afflicts-those-in-their-20s/articleshow](http://Timesofindia.indiatimes.com/city/bengaluru/Diabetes-now-afflicts-those-in-their-20s/articleshow)
- [17] Marco Viceconti, Peter Hunter, and Rod Hose, " Big Data, Big Knowledge: Big Data for Personalized Healthcare", *IEEE journal of Biomedical and Health Informatics*, vol. 19, no. 4, July 2015
- [18] Cong Wang,Qian Wang,Kui Ren,Ning Cao,Wenjing Lou,"Toward Secure and Dependable Storage Services in Cloud Computing",*IEEE Transactions on services computing*,vol 5 ,No 2,April,June 2012.
- [19] Ping ZHOU , Jingsheng LEI, Wenjun YE, " Large-Scale Data Sets Clustering Based on MapReduce and Hadoop",*Journal of Computational Information Systems* 7: 16 (2011).