International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072

BIG DATA ANALYTICS IN HEALTHCARE USING HADOOP

Prof. Manisha Sonawane[1], Bhushan Dhadve[2], Vaibhav Dighe[3], Prachi Chougule[4]

[1]Professor, Dept of Computer Engineering, SSJCOE, Dombivli, India
[2,3,4] Students, Dept of Computer Engineering, SSJCOE, Dombivli, India

---

Abstract - *In the final year project we are planning to work on this aspect of big data. We know hospitals tend to generate insane amounts of image data on daily basis, as the image collection grows we started to get such problem like prevention of data loss, wastes of time and money are those caused by redundant files. So using Hadoop and its Map-Reduce properties we will try to implement a program to efficiently sort and reduce image files. These images can be anything from CT-scans, X-rays, MRI's, 2D Doppler etc. We will realize at the end how Map-Reduce converts binary image data to hexadecimal values and removes duplicates, which indirectly will reduce the amount of stored data.*

***Key Words*: Hadoop, MapReduce, CT-Scan, data, Hexadecimal, image**

1. INTRODUCTION

In Hadoop system, the equal task to all node. Where this technique get fail in heterogeneous environment. Where performance of each and every node consider differently. To avoid this scenario we will consider advance Hadoop Big Data framework. The data explosion i.e. generating large amount of data. And it is very difficult to manage, Retrieve and processing by using traditional base system. This healthcare organization has created by keeping record, and regulatory requirement. This potential will help to improve quality of life. Hadoop consist of basically two Factors Map-Reduce and MD5.

2. LITERATURE REVIEW

BIG DATA IN DATA CENTERS: As classical big data research, the following work reported progress in Big Data networking reveals challenges and opportunities in databases in existence of big data. We see introduction of virtualization planning and cloud computing methods in data center networking. From a platform architecting perspective, their progress for accelerating big data analytics is with current level of technology. As a recent effort, they are contributing on optimizing big data processing efficiency in Hadoop and MapReduce.

BIG DATA IN CLOUD COMPUTING: Remarkable progress of big data networking has also been reported in the area of cloud computing Resource management and allocation in multi-cluster clouds was introduced.

A dataflow-based performance analysis for big data cloud, i.e., Hitune, was presented. Interesting case studies on big data processing in a cloud computing environment were also depicted. A recent online cost minimization approach was depicted specifically for reducing cooling energy cost for big data analytics cloud.

3. PROPOSED SYSTEM

Our project is divided into two sections:

1) Prediction using R language: Basically our project consists of R language is language provides us with a wide variety of statistical and graphical techniques such as linear and non-linear modeling, classical statistical test, clustering, classification, time series analysis, etc.

In this project, we are hiring data from the government of tumor patients. Using R language we are segregating the data according to the ratio of tumor patients of different years.

2) Analysis using big data: This project also consists of big data analytics. Big data concept is used in different healthcare institutions. It is an act of gathering and storing a large amount of data information across the world. It also helps us to analyze data to take a better decision and correct business moves. Map-Reduce is used to segregate image data as well as text data and it also removes duplicate images from the data.

4. RESULT

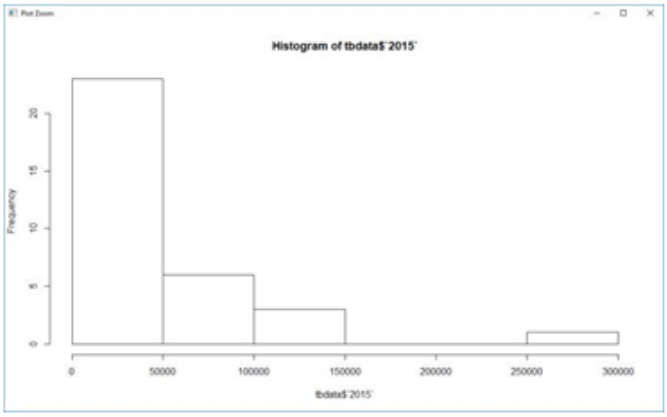

Fig-1: Screenshot 1

We try and represent the data in statistical format. Evaluation of the data is done based on patterns and pictorial representation is done. Gives a better understanding of data as it is represented in graphical form. No manual work required to see various demographics of the given data. Easier to understand and take required

business decisions. With the help of available commands in R, it is easier to perform Correlation and Linear Regression on the available data set. Prediction of values can be highly accurate if the amount of data and value of correlation is high. Helps in foreseeing the future outcome and to take appropriate decisions before it is too late.

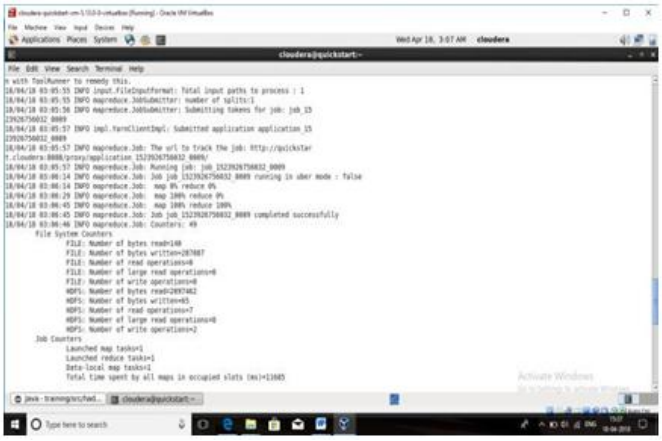

Fig 2: Screenshot 2

Here we have applied map reduce the CT-scan images. The relative output we get after the execution is that all the redundant or duplicate data is removed. Space complexity is reduced. Storage space is compressed. Ambiguity is removed.

5. CONCLUSIONS

In today's world, big data is used in various fields such as network centers, medical field, banks, etc. In healthcare center, big data provide us with clinical data and also helps to make good decision.in nearby future we will see the use of big data analytics across the healthcare industry. As big data helps to overcome issues such as establishing governance and standard, privacy, security and continues improvements in tools and technologies. In these proposed paper it will also improve the performance of the Map-Reduce algorithm and at the same time, it will maintain the security.

6. ACKNOWLEDGEMENT

No project is ever complete without the guidance of those expert how have already traded this past before and hence become master of it and as a result, our leader. So we would like to take this opportunity to take all those individuals how have helped us in visualizing this project. We express our deep gratitude to our project guide Prof. Manisha Sonawane for providing timely assistant to our query and guidance that she gave owing to her experience in this field for past many year. She had indeed been a lighthouse for us in this journey. We extend our sincerity appreciation to all our Professor form SHIVAJIRAO S. JONDHALE COLLEGE OF ENGINEERING for their valuable inside and tip during the designing of the project. Their contributions have been valuable in so many ways that we find it difficult to acknowledge of them individual. We also great full to our HOD Prof. Mr. P.R. Rodge for extending her help directly and indirectly through various channel in our project work.

REFERENCES

1. Raghupathi W. Data Mining in Health Care. In: Kudyba S, editor. Healthcare Informatics: Improving Efficiency and Productivity. 2010. pp. 211–223.

2. Burghard C. Big Data and Analytics Key to Accountable Care Success. 2012.

3. Dembosky A. “Data Prescription for Better Healthcare.” Financial Times, December 12, 2012, p. 19.2012.

4. Feldman B, Martin EM, Skotnes T. “Big Data in Healthcare Hype and Hope.” October 2012. Dr. Bonnie 360.2012.

5. Fernandes L, O’Connor M, Weaver V. J AHIMA. 2012. Big data, bigger outcomes; pp. 38–42. [PubMed]