Special Issue - 2018

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICONNECT - 2k18 Conference Proceedings

Smart Health Prediction Using Hadoop

#1 A .Sivaranjani , #2 S .Priyadharshini, #3. A.Porkodi, 4#. A .Vijayalakshmi, #5.S.Suseela

IV CSE Students, Department of Computer Science and Engineering , Periyar Maniammai University, Vallam , Thanjavur, India.

Assistant Professor, Department of Computer Science and Engineering , Periyar Maniammai University, Vallam , Thanjavur, India.

Abstract: In today's modern world, healthcare needs to be modernized .It means that the healthcare data should be properly analyzed so that we can categorize it into groups of gender, disease, city, symptoms and treatments. The gigantic size of analytics will need large computation which can be done with the help of distributed processing HADOOP. The frameworks use will provide multipurpose beneficial outputs which includes getting the healthcare data analysis into various forms. Big data is used to predict epidemics, cure disease, improve quality of life and avoid preventable deaths. with the increasing population of the world, and everyone living longer, and many of the decision behind those changes are being driven by data. The proposed system will group together the disease and their symptoms data and analyze it to provide cumulative information. After the analysis, algorithm could be applied to the resultant and grouping can be made to show a clear result. The group made by the system would be symptoms wise, disease wise etc. as the system will display the data group wise it would be helpful to get a clear idea about the disease and their rate of spreading , so that appropriate treatment could be given proper time.

Keywords: Big data, NoSql, Medical Record, Map reduce.

I. INTRODUCTION

In traditional Hadoop system, the master assign equal task to all node. Where this technique get fail in heterogeneous environment. Where performance of each and every node consider different. to avoid this scenario we will consider advance Hadoop big data frame work . 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
- ✓ HDFS(Hadoop distributed file system)

Hadoop is platform which are in distributed manner and deployed in clustering format. And cluster should be homogeneous. This gigantic size of analytics will need large computation which can be does with help of distributed processing, Hadoop. Map Reduce, a popular computing para diagram for large –scale data processing in cloud computing .disease and their possible symptoms are group together and send it as input to system which generate cumulative information. after analysis done ,if we provide symptoms the system will generate name of disease.it will create clear picture of output in graphical format. Age, gender, disease, doctor, payment type, insurance are some grouping categories based on which analysis and grouping can be done. This will be achieved with the help of Hadoop framework with the help of which we can do a very fast analysis for big data.

This framework consist of two function namely map() and reduce(),each having different parameters. Map function contain two parameters i.e. key and value by default this framework assigns value 1 to all keys. Hadoop uses a specialized scheduling mechanism for allocating task to every node. Hadoop which ensures fair task allocation and load balancing.in heterogeneous clusters the performance of every node differs from all other nodes. The performance of such clusters and for better resource utilization, the task scheduling should be adaptive.

In Hadoop data will not store on single cluster but it will save on number of clusters.so data will be proceed in parallel manner to achieve performance .Hadoop is trying to keep backup of data. Numbers of times data will get vanished, to avoid this group of clusters will be generated.

II. LITERATURE SUREVEY

To enhance the processing of conventional healthcare system, we have a proposed a series of Big Data health Care System by using Hadoop. There are many techniques proposed in order to efficiently process large volume of medical record which has explained below:

1] Aditi Bansal and Priyanka   Ghare proposed "Healthcare Data Analysis using Dynamic Slot Allocation in Hadoop ". In this paper HealthCare System is analysis using Hadoop using Dynamic Hadoop Slot Allocation (DHSA) method. This paper proposed a frame work which focus on improving the performance of MapReduce workloads and maintain the system. DHSA will focuses on the maximum utilization of slots by allocating map (or reduce) slots to map and reduce tasks dynamically.

2] Wullianallur Raghupathi and Viji Raghupathi has proposed "Big data analytics in healthcare: promise and Potential " In this paper author proposed the potential and promise of big data analytics in healthcare. The paper provides a broad overview of big data analytics for healthcare researchers and practitioners. Big data analytics in healthcare is evolving into a promising field for providing insight from very large data sets and improving outcomes while reducing costs. Its potential is great; however there remain challenges to overcome.

III.PROPOSED SYSTEM

Proposed concept deals with providing database by using Hadoop tool we can analyze no limitation of data and simple add number of machines to the cluster and we get results with less time, high throughput and maintenance cost is very less and we are using joins, partiations and bucketing techniques in Hadoop

Advantages:

- ✓ No data loss problem
- ✓ Efficient data processing

Special Issue - 2018

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICONNECT - 2k18 Conference Proceedings

IV. TOOLS AND TECHNOLOGIES

Hadoop is a tool that is perfectly handles big data which is used to work with large amount of data's that has different kind of dataset. For installing Hadoop first choose the open source operating. Hadoop works with the implementation of two main factors such that HDFS and Map reduce.

V. METHODOLOGIES

MODULES:
  - ➢ Data Prepossessing Module
  - ➢ Data Ingestion Module With Sqoop
  - ➢ Data Analytic Module With Hive
  - ➢ Data Analytic Module With Pig
  - ➢ Data Analytic Module With MapReduce
  - ➢ Data Analytic Module With R

Data Preprocessing Module

In this module we have to create Data set for health it contain set of table such that Patient details, disease details, doctor details, billing details and payment details for last four years .and this data first provide in MySQL database with help of this dataset we analysis this project.

Data Migration Module with Sqoop

In this module we have to transfer the dataset into Hadoop(HDFS), that will be happen in this module. Sqoop is a command-line interface application for transferring data between relational databases and Hadoop.

In this module we fetch the dataset into Hadoop (HDFS) using Sqoop Tool. Using Sqoop we have to perform lot of the function, such that if we want to fetch the particular column or if we want to fetch the dataset with specific condition that will be support by Sqoop Tool and data will be stored in Hadoop (HDFS).

Data Analytic Module with Hive

Hive is a data ware house system for Hadoop. It runs SQL like queries called HQL (Hive query language) which gets internally converted to map reduce jobs. Hive was developed by Facebook. Hive supports Data definition Language (DDL), Data Manipulation Language (DML) and user defined functions.

In this module we have to analysis the dataset using HIVE tool which will be stored in Hadoop (HDFS).For analysis dataset HIVE using HQL Language. Using hive we perform Tables creations, joins, Partition, Bucketing concept. Hive analysis the only Structure Language.

Data Analytic Module with Pig

Apache Pig is a high level data flow platform for execution Map Reduce programs of Hadoop. The language for Pig is pig Latin. Pig handles both structure and unstructured language. It is also top of the map reduce process running background.

In this module also used for analyzing the Data set through Pig using Latin Script data flow language.in this also we are doing all operators, functions and joins applying on the data see the result.

Data Analytic Module with MapReduce

MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. In this module also used for analyzing the data set using MAP REDUCE. Map Reduce Run by Java Program.

VI.SYSTEM TECHNIQUES

MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job.

The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes. Under the MapReduce model, the data processing primitives are called mappers and reducers. Decomposing a data processing application into mappers and reducers is sometimes nontrivial. But, once we write an application in the MapReduce form, scaling the application to run over hundreds, thousands, or even tens of thousands of machines in a cluster is merely a configuration change. This simple scalability is what has attracted many programmers to use the MapReduce model.

The Algorithm

  - Generally MapReduce paradigm is based on sending the computer to where the data resides.

  - MapReduce program executes in three stages, namely map stage, shuffle stage, and reduce stage.

Map stage : The map or mapper's job is to process the input data. Generally the input data is in the form of file or directory and is stored in the Hadoop file system (HDFS). The input file is passed to the mapper function line by line. The mapper processes the data and creates several small chunks of data.

Reduce stage: This stage is the combination of the Shuffle stage and the Reduce stage. The Reducer's job is to process the data that comes from the mapper. After processing, it produces a new set of output, which will be stored in the HDFS.

Special Issue - 2018

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICONNECT - 2k18 Conference Proceedings

VII.SYSTEM ARCHITECTURE:

The proposed method performs well in the general population as well as in sub-populations. Results indicate that the proposed model significantly improves predictions over established baseline methods analyzing electricity consumption. The goal of this study was to analyze how much of units consumed in last four years and how much amount they paid previous four year as the forecast for the following year.
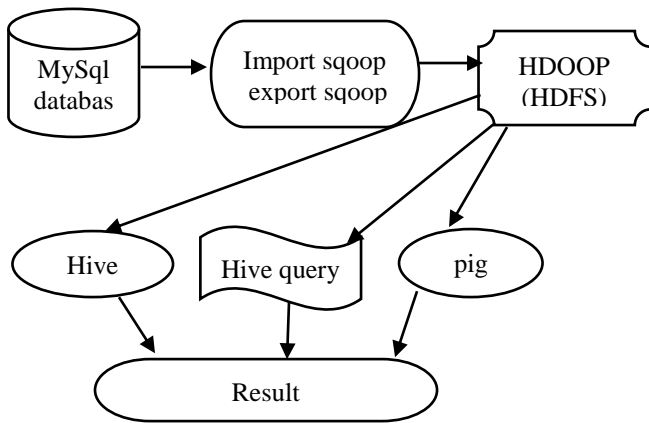

Fig1 System architecture

VIII . RESULT AND DISCUSSION

From the disease section of the health prediction, sample health dataset is taken using the invoice copies or disease or bills of the health prediction. 9620X302 sample Binary dataset is manipulated with Hadoop and Sqoop tool and the results. Hadoop and Sqoop tool are compared based on the frequent diseases and association rules generated. Hadoop and Sqoop tool provides output only for very low support values. Very low support values are meaningless because it shows nothing about the patinets' behavior.

IX. CONCLUSION

This paper illustrated the smart health prediction using Hadoop. The capability of big data will transform the way today's healthcare providers operate the sophisticated technologies to get knowledge from clinical records and make good decisions. In the nearby future we will see implementation of big data analytics in health care industry. Big data provides security and privacy. This paper proposes a framework which is aiming that it will improve the performance of MapReduce workloads and at the same time will maintain the fairness.

ACKNOWLEDGMENT

REFERENCES:

[1] "Big data analytics in healthcare : promise and potential" published by Viju Raghupathi and Wullianallur Raghupathi in 2014

[2] " Big data for Better Health Planning" published by Jigna Ashish Patel and Priyanka Sharma

[3] " Map Reduce Algorithms for Big Data Analysis" published by Kyuseok Shim

[4] " Hadoop Based Analytics on Next Generation Medicare System" publishedby Gopal Tathe , Pratik Patil , Sangram Parle

[5] Aditi Bansal, Ankita Deshpande, Priyanka Ghare, Seema Dhikale, Balaji Bodkhe "Healthcare Data Analysis using Dynamic Slot Allocation in Hadoop" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-3 Issue-5, November 2014 (IJRTE)

[6] Shubham Borikar, Mohan Bhagchandani, Raunak Kochar, Ketansing Pardeshi, Manisha Gahirwal, A Survey on Applications of Big Data Analytics in Healthcare International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-5 Issue-5, November 2015.

[7] R. Sathiyavathi A Survey: Big Data Analytics on Healthcare System, Contemporary Engineering Sciences, Vol. 8, 2015, no. 3, 121 - 125 HIKARI Ltd, www.mhikari.com http://dx.doi.org/10.12988/ces.2015.412255 .

[8] Priya nka K B.V.B.C.ET Hubli, Prof Nagarathna Kulennavar B.V.B C.E.T Hubli.A Survey On Big Data Analytics In Health Care, Priyanka Ketal, / (IJCSIT) International Journal of Computer Science

[9] Divyakant Agrawal, UC Santa Barbara, Philip Bernstein, Microsoft Elisa Bertino, Purdue Univ. "Big Data White pdf", from Nov 2011 to Feb-2012.

[10] Durham, E.-E.A.; Rosen, A.; Harrison, R.W., "Optimization of relational database usage involving Big Data a model architecture for Big Data applications," in Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on , vol., no., pp.454-462, 9-12 Dec.2014.

[11] Shanjiang Tang, Bu-Sung Lee, Bingsheng He, "DynamicMR: A Dynamic Slot Allocation Optimization Framework for MapReduce Clusters ", IEEE Transactions,Vol 2 No.3 Sep 2013,pp.333-345.

[12] Wullianallur Raghupathi and Viju Raghupathi, "Big data analytics in healthcare: promise and potential", Health Information Science and Systems ,pp.1-10, 2014.

[13] Aditi Bansal, Balaji Bodkhe, Priyanka Ghare, Seema Dhikale, Ankita Deshpande, "Healthcare Data Analysis using Dynamic Slot Allocation in Hadoop" International Journal of Recent Technology and Engineering , Vol-3 Issue-5, November 2014, pp. 1518.

[14] Divyakant Agrawal, UC Santa Barbara, Philip Bernstein, Microsoft Elisa Bertino, Purdue Univ. "Big Data White pdf", from Nov 2011 to Feb-2012.