ANALYSIS AND PREDICTION OF CHRONIC KIDNEY DISEASE USING GENETIC AND BIG DATA TECHNIQUES

J. VIJAYALAKSHMI
Ph.D Research Scholar
Department of Computer Science
Alagappa University, Karaikudi.
Tamilnadu, India

Dr. E. RAMARAJ
Professor & HOD
Department of Computer Science
Alagappa University, Karaikudi
Tamilnadu, India

Abstract - A huge quantity of information is obtainable in hospitals and clinics. Big data could be a technology used to manage an oversized volume of structured and unstructured data. It's tough to process ancient database and software techniques Today it's effectively utilized in all technologies to bring distinctive resolution. Big data has the power to assist health care organization to enhance the prediction of kidney diseases and create quicker and a lot of intelligent decisions. Hadoop is used in analysis of big data. It's ascendable, cost-efficient, Flexible, Fast, and Resilient to Failures as compared to different methodology. It's important to search out a prediction system with more accuracy. Genetic rule is used for attribute selection or reduction. Hive is used for querying of data and analysis. It generates reports and graphical charts. By using this analysis doctors suggest medications supported to the patients. The prognostic analysis is useful to treat patients using specific medications, based on variety of factors like life style habits, medical history, obesity, smoking and health status like blood pressure and blood glucose. In this paper map reduce is used to convert unstructured data into structured data and help to find the kidney dataset. The hive tool is used to analyze the different types of kidney diseases. This analysis compares with the patient kidney dataset to predict kidney disease. It supports the medical practitioner to treat the kidney diseases.

Keyword - **Big data, Genetic Algorithm, Map Reduce, HIVE, Kidney Disease**

I. INTRODUCTION

The paper deals primarily with Big data in health care to predict illnesses, identify symptoms, enhance treatment, provide the right medicine for patients to recover from kidney disease, and improve quality of care, reduce costs, improve life expectancy and reduce the impact of death in advance. There are many companies and institutions that have come together to solve the problem. Such concerns include the benefits of Big Data, the Medicare programs and incentives and health care.

Most people around the world are affected by kidney-related diseases today. Big data plays an important role in saving the health of patients and reducing the mortality of patients with the kidney.

Data collected from different hospitals for kidney patients, data collected from experts in the treatment of kidney disease, clinical information, paper format records are entered in the digital format. The huge volume of data is stored, processed and analysed via big data. This data obtained by Big Data Analytics is useful for doctors to treat the patient easily

The diagnosis of Kidney disease relies on the clinical data. Prediction of kidney disease can enable the medical experts to predict kidney disease within the clinical data of the patient. The healthcare industry is gathering big data from healthcare organizations that really need to be mined to discover all the spied information needed to make reliable and enhanced decisions. Kidney is an integral part of the human body which helps remove the body's impurity. The symptoms of acute kidney failure may consists of reduced urine output, although the output of urine remains normal sometimes, Retention of blood, causing the arms, ankles and feet to swell. It leads to Shortness of breath. Smoking, obesity, cholesterol, blood pressure, diabetes, unhealthy diet and physical activity are the major risk factors that cause kidney disease.

II. LITERATURE REVIEW

Anbarasi et al (2010) in their work Genetic Algorithm is used to determine the attributes which contribute more towards the diagnosis of Heart diseases.

Shruti Ratnakar et al (2013) in their work Genetic Algorithm is used to predict the risk level of heart disease with optimal reduced set of attributes using Genetic Search method.

Tripti Mehta, et al (2016) in their work For Map Reduce a code is very lengthy and more development efforts are required but the code efficiency is high as compared to hive. However hive is like SQL query language and having higher level of abstraction.

Subramaniyaswamy et al (2015) in their work, the unstructured data is converted to structured by using Map

Reduce technique. To process huge data Map Reduce is the useful method.

Using a wrapper method from machine learning we identify a set of 12 attributes (down from 24 attributes) which detect CKD with high accuracy.

III.   MAP REDUCE

Map Reduce is a software framework for distributed processing of large data sets on compute clusters of commodity hardware. It is a sub-project of the Apache Hadoop project. The framework takes care of scheduling tasks, monitoring them and re-executing any failed tasks. Map Reduce framework has three operations. They are Map stage, Shuffle stage and Reduce stage. In Map stage mapper job is applied to input data, the mapper processes the data and creates a output keys. In Shuffle stage, based on the output key worker node redistribute the data so that on worker node has the data belongs to one key. In Reduce Stage worker node process each group of output data and produce new set of output. This output is stored in HDFS. In MapReduce, the guide errands and deficient lessen undertakings will be executed rather than the whole guide and decrease assignments if there should be an occurrence of a solitary hub disappointment. We can accomplish the base execution time. A few structures in view of MapReduce are proposed which are equipped for understanding information semantics, rearranging the written work of examination applications and conceivably enhancing execution by lessening MapReduce stages.

IV.   GENETIC ALGORITHM

In   the area of artificial intelligence,   a   genetic algorithm (GA) is a search heuristic that imitates the system of natural evolution.
This heuristic is widely used to produce useful solutions for problems of optimization and search.
Genetic algorithms belong to the larger class of evolutionary algorithms (EA) that produce optimized solutions based on natural evolutionary techniques such as descent, mutation, selection and crossover. It needs a standard genetic algorithm:
1. A Solution Domain Genetic Representation,
2. A fitness method for evaluating the scope of the solution.
A generic solution representation is as a bits set. The genetic representation determines the fitness function. The fitness function is always problem dependent. Initially many individual solutions are (usually) randomly generated to form an initial population. The next successive generation population are selected from the existing. Fitness based process is used to select the individual solutions. Then genetic operator will select the next generation population.

V.   HIVE ARCHITECTURE

Hive is developed over the software base of Hadoop, which is similar to a distributed file system.

Using HiveQL to search information in Hive. HiveQL is submitted using the thrift, ODBC or JDBC interfaces via the Command Line Interface (CLI), the Web UI or an external client such as iReport. Using the metadata stored in Metastore, the driver transfers the request to the compiler where it goes through standard parse, semantic analysis phases and type check. The compiler generates a logical plan, which is optimized by a simple rule-based optimizer. Finally, an optimized plan is generated in the form of tasks for map-reduction and HDFS tasks. Instead, using Hadoop, the execution engine executes these tasks in the order of their dependency. The main components of Hive and its relations with Hadoop are shown in Figure 2. Hive's major components are:

A. External Interface

Hive provides CLI and web user interfaces as well as application programming interfaces (API) such as JDBC and ODBC.
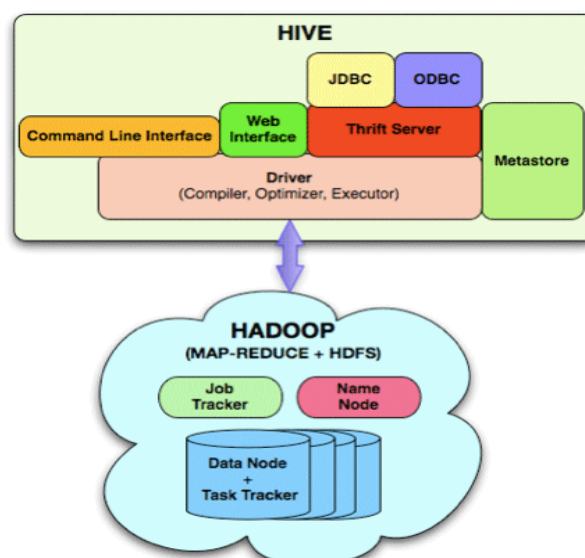

Fig. 1. Hive Architecture

B. Thrift Server:

Hive Thrift Server can execute HiveQL statements with a very simple client APU. Thrift [II] is a cross-language software system where a database written in one language (such as Java) can support clients in other languages as well. The Thrift Hive clients developed in various languages are used to create common drivers such as JDBC (Java), ODBC(C++). PHP, Perl, Python, etc. are used for writing scripting drivers.

C. Driver

The life cycle of a HiveQL statement during optimization, execution, and compilation are managed by driver. HiveQL statement received from the thrift server or other interfaces are used to handle session and it will be

used to keep track of statistics like execution time, number of output rows, etc.

VI. PROPOSED METHODOLOGY

The proposed system detects the chronic kidney disease early using big data. Structured and unstructured data are collected from different servers.
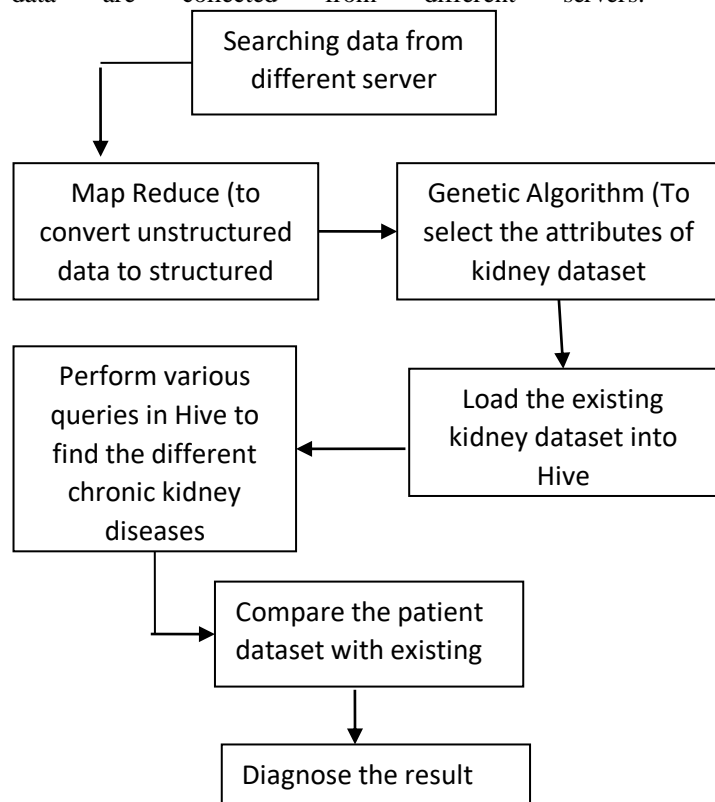

Fig. 2. Proposed Methodology Flow Diagram

Hadoop Map Reduce provides an ideal environment to perform these tasks in when processing massive dataset. After Map Reduce we get the structured data. Genetic algorithm is used to get the attributes of kidney from structured data set. Load the structured kidney dataset into hive tool. Execute different query on hive to predict the different kidney diseases. Diagnosis the result and it helps the doctors to treat the patient.

The attribute which is used for the research are given below

Attribute	Representation	Data type	Description
Age	Age	Numeric	Years
Blood Pressure	BP	Numeric	Mm/Hg
Sugar	Su	Nominal	0,1,2,3,4,5
Albumin	Al	Nominal	1.005,1.010,1.015,1.020,1.025
Blood Urea	Bu	Numerical	Mgs/dl
Serum Creatinin	Sc	Numerical	Mgs/dl

Red Blood Cells	Rbc	Nominal	0,1,2,3,4,5
Potassium	Pot	Numerical	mEq/L
Diabetes Mellitus	Dm	Nominal	Yes,No
Sodium	Sod	Numerical	mEq/L
Haemoglobin	Hemo	Numerical	Gms

Fig. 3. Attribute Table

VII. CONCLUSIONS

Our proposed methodology briefly discussed about how to convert the unstructured data into structure data using Hadoop map reduce method to get useful information. Already knows structured and unstructured data can be analysed through big data analysis technology with tools. Our methodology explained to predict chronic kidney disease using related medical data set from huge volume of dataset it may be structured or unstructured. Genetic algorithm is used to get the attributes of kidney diseases. Hive tool is used to find the different types of kidney diseases. Compare this with the patient dataset and find the status of the kidney. It will help the doctors to treat the patients in an easy way.

VIII. ACKNOWLEDGEMENT

IX. REFERENCES

[1] Mathias Kettner., [online] available on http://www.linuxproblem.org/art_9.html, "SSH login without password".

[2] Dustin Kirkland, [online] available on http://manpages.ubuntu.com/manpages/raring/man1/nvi.1.html, "Ubuntu manuals".

[3] http://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf

[4] Başar.M.D, Sarı.P, Kılıç.N and Akan.A, (2016). Detection of chronic kidney disease by using Adaboost ensemble learning approach, in 24th Signal Processing and Communication Application Conference (SIU), Zonguldak, pp. 773-776.

[5] Basma Boukenze, Hajar Mousannif and Abdelkrim Haqiq (2016). Predictive Analytics In Healthcare System Using Data Mining Techniques in Computer Science & Information Technology pp. 01–09

[6] Anbarasi.M., and Anupriya.E, (2010). Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm, in [pp.5370 – 5376], vol.2 no.10.

[7] Shruti Ratnakar., Rajeswari.K and Rose Jacob, (2013). Prediction of Heart Disease Using Genetic Algorithm For Selection Of Optimal Reduced Set Of Attributes, in International Journal of Advanced Computational Engineering and Networking, on ISSN: 2320-2106, Volume-1, Issue-2, April-2013

[8] Tripti Mehta., and Neha Mangla. (2016). A Survey Paper on Big Data Analytics using Map Reduce and Hive on Hadoop Framework, in Special Issue on International Journal of Recent Advances in Engineering & Technology on (IJRAET) V-4I-2.

[9] https://en.wikipedia.org/wiki/MapReduce

[10] https://www.tutorialspoint.com/hadoop/hadoop_ mapreduce.htm

[11] https://acadpubl.eu/hub/2018-119-18/2/116.pdf

[12] https://www.semanticscholar.org/paper/PREDIC TION-OF-HEART-DISEASE-USING-GENETIC-ALGORITHM-Ratnakar-Rajeswari/8d1aa8d3fe9d4e3bd006841ac10d24bc e23293e0

[13] https://www.researchgate.net/publication/318338 588_HIVE-Processing_Structured_Data_in_Hadoop