# A CLINICAL DECISION SUPPORT FRAMEWORK FOR HETEROGENEOUS DATA SOURCES

Dr. S. Vijayaragavan, Professor[1,] Harikrishna kk, Student[2].
Department of computer Science and Engineering,
Muthayammal engineering college,Tamilnadu,India

## ABSTRACT

To keep pace with the updates in obliging informatics, thriving recuperating data is being amassed unendingly. Regardless, inferable from the not too appalling gathering of its classes and sources, therapeutic data has ended up being particularly jumbled in various specialist's work environments that it now needs Clinical Decision Support (CDS) system for its affiliation. To sensibly utilize the party flourishing data, we propose a CDS structure that can interweave mixed thriving data from different sources, for instance, take a gander at office test works out as planned, critical information of patients, and achievement records into a joined depiction of features everything considered. Using the electronic prospering healing data so made, multi-name delineation was used to endorse a layout of afflictions and thusly help experts in diagnosing or treating their patients' therapeutic issues more capably. Once the ace sees the contamination of a patient, the running with organize is to consider the conceivable complexities of that perplexity, which can impel more infections.

**Keywords :** Clinical Decision Support, analysing and file.

## OBJECTIVE:

In this paper we are analyzing clinical data by using hadoop tool along with some hadoop ecosystems like hdfs, mapreduce, sqoop, hive and pig. By using these tools, we can process no limitation of data, no data lost problem, we can get high throughput, maintenance cost also very less and it is an open-source software, it is compatible on all the platforms since it is Java based. In clinical data details, this is based on whenever patient enters their id in hospital database within no time complex their data will present.

## INTRODUCTION

It is broadly acknowledged that wellbeing data devices and machine learning methods can be misused effectively to help specialists in diagnosing and treating their patients all the more productively. Utilizing their experience and information, the doctors order patients and analyze their infections, however in doing as such, it is plausible that they submit a few errors, especially when they need sufficient experience or when their staff of judgment is poor. In such circumstances, Clinical Decision Support (CDS) frameworks, including frameworks that give finding, customized restorative estimation, treatment and important learning, would be useful to the doctors by method for giving them particular learning, patients' data and keen applications, which can enhance the proficiency of their basic leadership forms. Cds frameworks center around separating attributes of patients, in light of which they order patients and give relating clinical recommendations to the doctors. Patients' restorative data is restored from their own medicinal information.

## EXISTING SYSTEM:

Existing concept deals with providing backend by using mysql which contains lot of drawbacks i.e.,

data limitation is that processing time is high when the data is huge and once data is lost, we cannot recover so thus we proposing concept by using Hadoop framework.

## DRAWBACKS IN EXISTING SYSTEM:

- ✓ We can process limitation of data.
- ✓ We get results with take more time and miniatous cost is very high.

## PROPOSED SYSTEM

In this paper we are analyzing clinical data by using hadoop framework along with some hadoop ecosystems like hdfs, mapreduce, sqoop, hive and pig. By using these tools we can process no limitation of data, no data lost problem, we can get high throughput, maintence cost also very less and it is a opensouresoftware,it is compatible on all the platforms since it is Java based

## METHODOLOGIES

Following modules involves

## MODULES

- ➢ Preprocessing clinical database
- ➢ Storage
- ➢ Analyse query
- ➢ Analysis latin script (pig)
- ➢ Processing (mapreduce)

## MODULE DIAGRAMS & MODULE DESCRIPTION

## PREPROCESSING CLINICAL DATABASE:

In this module, analyzing the data with different kinds of fields in Microsoft Excel then it converted into comma delimited format which is said to be csv(comma separator value) file and moved to mysql backup through Database.

## PREPROCESSING:



Clinical.csv          MYSQL DB
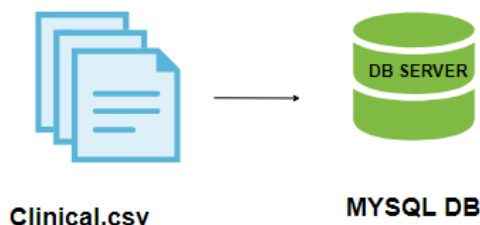
Here by Getting historical data we have to convert those historical batch processing data from (.xlsc) format to (.csv) format and by taking backup of all those data in MYSQL Database to avoid loss of data.

## STORAGE:
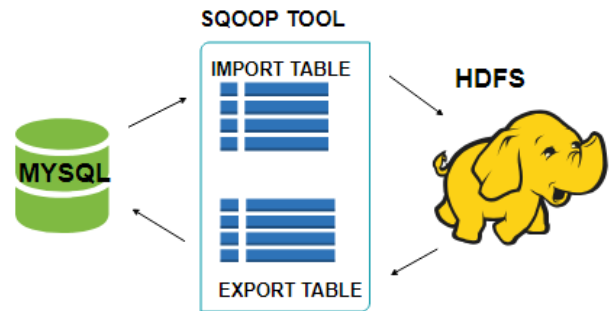


In this module we are getting all those backup data which we have stored in MYSQL and importing all those data by use of sqoop commands to HDFS( Hadoop Distributed File System).now all the data are stored in HDFS were it is ready to get processed by use of hive.

## ANALYSE QUERY:
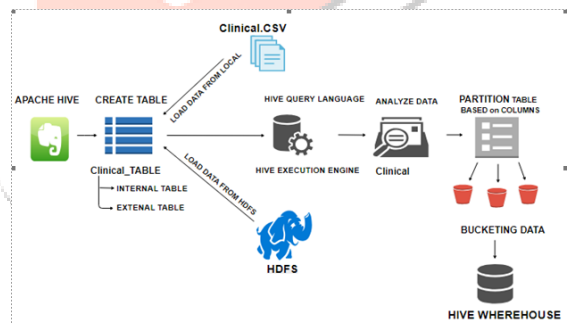


In this module we are getting all those data from HDFS to HIVE by use of sqoop import command .were hive is ready to analyze .here in HIVE we can process only structured data to analyze. by extracting only the meaningful data and neglecting unclenched data we can analyze the data in more effective manner by use of hive.

## ANALYSIS LATIN SCRIPT (PIG):

To analyze Clincal using Pig, programmers need to write scripts using Pig Latin language and execute them in interactive mode using the Grunt shell. All these scripts are internally converted to Map and Reduce tasks.After invoking the Grunt

shell, you can run your Pig scripts in the shell.Except LOAD and STORE, while performing all other operations, Pig Latin statements take a relation as input and produce another relation as output. As soon as you enter a Load statement in the Grunt shell, its semantic checking will be carried out. To see the contents of the schema, you need to use the Dump operator. Only after performing the dump operation, the MapReduce job for loading the data into the file system will be carried out. Pig provides many built-in operators to support data operations like grouping, filters, ordering, etc.
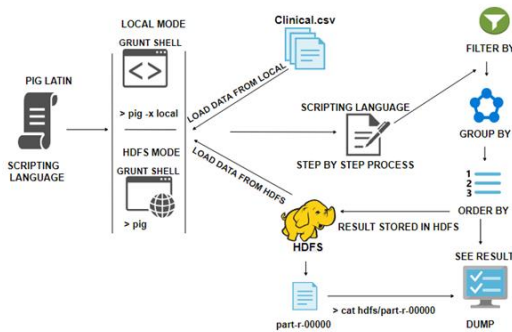


## PROCESSING (MAPREDUCE):

Map Reduce is a framework using which we can write applications to process huge amounts of Clinical, in parallel, on large clusters of commodity hardware in a reliable manner. Map Reduce is a processing technique and a program model for distributed computing based on java. The Map Reduce algorithm contains two important tasks, namely Map and Reduce. Map Reduce program executes in three stages, namely map stage, shuffle stage, and reduce 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. 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.
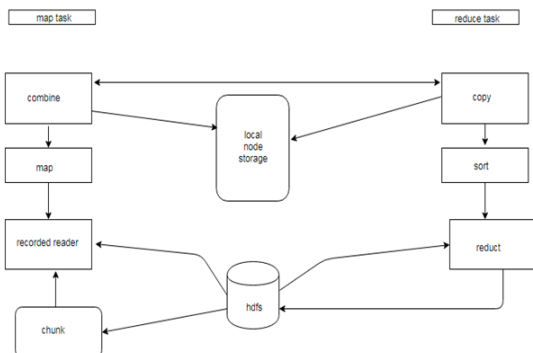


## SYSTEM TECHNIQUES:

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.

## USE CASE DIAGRAM:
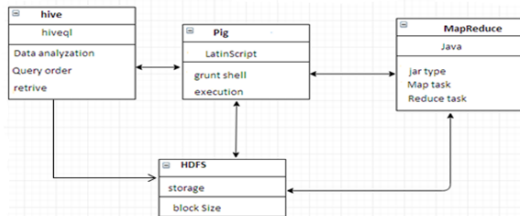


## OBJECT DIAGRAM ANALYSIS STORE
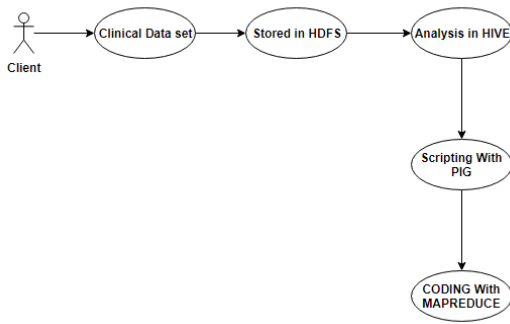


## EXPLANATION:

A use case diagram is a graphic depiction of the interactions among the elements of a system. A use case is a methodology used in system analysis to identify, clarify, and organize system requirements. In our use case diagram we analyze of clinical data, which will be easy to predicate. Let us see the process of use case diagram in which whatever data analyzation done to be stores in hdfs (hadoop distribution file system) by different type of styles like hive,pig and mapreduce.

## EXPLANATION:

An object diagram in the Unified Modeling Language (UML) is a diagram that shows a complete or partial view of the structure of a modeled system at a specific time. If the user is new register the account. In object diagram the client can be analysis the clinical, which will be easy to predicate. Let us see the process of use object diagram in which whatever data analyzation done to be stores in hdfs (hadoop distribution file system) by different type of styles like hive,pig and mapreduce.
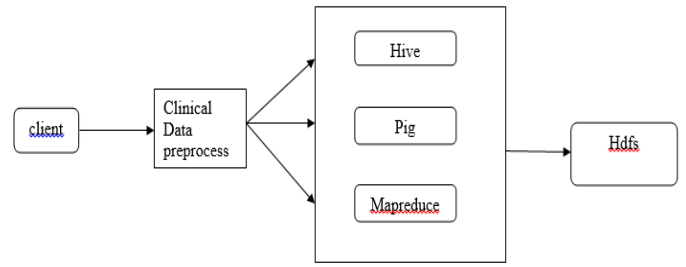
## CLASS DIAGRAM:



## STATE DIAGRAM:



## EXPLANATION:

A class diagram is an illustration of the relationships and source code dependencies among classes in the Unified Modeling Language (UML). In our class diagram the user login to his/her own account. In our class diagram the client can be analysis the clinical, which will be easy to predicate. Let us see the process of use class diagram in which whatever data analyzation done to be stores in hdfs (hadoop distribution file system) by different type of styles like hive,pig and mapreduce.
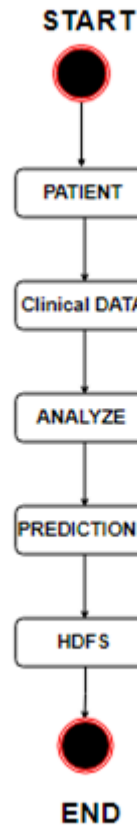
## EXPLANATION:

A state diagram is a type of diagram used in computer science and related fields to describe the behavior of systems. State diagrams require that the system described is composed of a finite number of states; sometimes, this is indeed the case, while at other times this is a reasonable abstraction. In state diagram the client can be analysis the clinical, which will be easy to predicate. Let us see the process of use state diagram in which whatever data analyzation done to be stores in hdfs (hadoop distribution file sys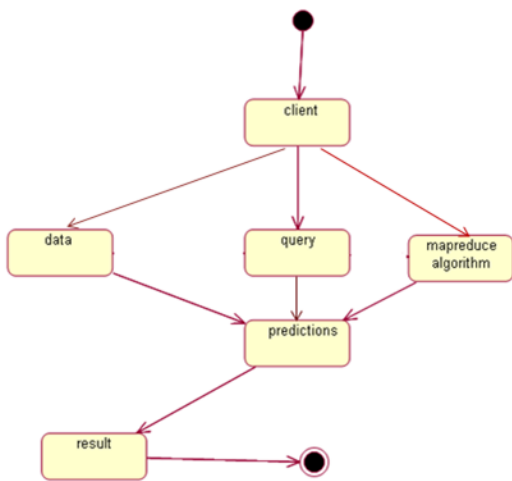tem) by different type of styles like hive,pig and mapreduce and call detail record analysis can be delivered into client.

## ACTIVITY DIAGRAM:



## EXPLANATION:

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams are intended to model both computational and organizational processes (i.e. workflows). We can analysis the clinical, which will be easy to predicate. Let us see the process of use activity diagram in which whatever data analyzation done to be stores in hdfs (hadoop distribution file system) by different type of styles like hive,pig and mapreduce and cdranalysis can be delivered into client.

## SEQUENCE DIAGRAM:
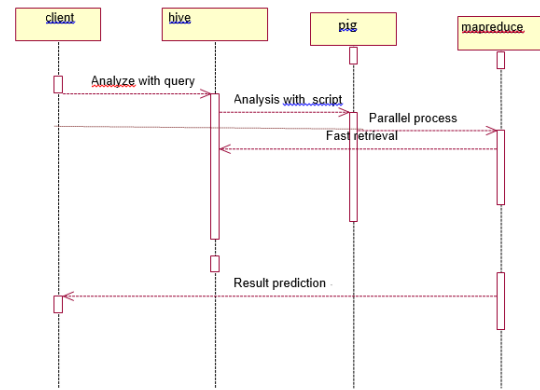


## EXPLANATION:

A Sequence diagram is an interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. A sequence diagram shows object interactions arranged in time sequence. In sequence diagram the client can be analysis theclinical data, which will be easy to predicate. Let us see the process of use sequence diagram in which whatever data analyzation done to be stores in hdfs (hadoop distribution file system) by different type of styles like hive,pig and mapreduce and power plant analysis can be delivered into client.
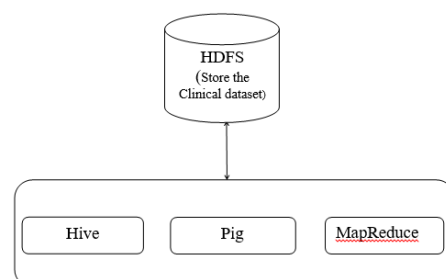
## DATA FLOW DIAGRAM:

## LEVEL-0:
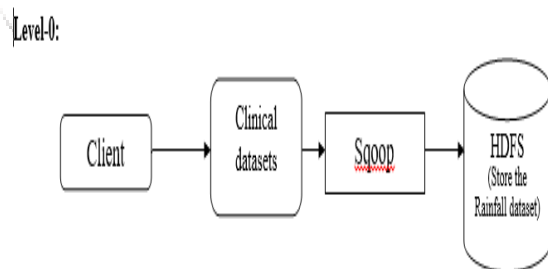
## EXPLANATION:

A Data Flow Diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. A DFD is often used as a preliminary step to create an overview of the system, which can later be elaborated. In data flow diagram the client can be analysis theclinical data, which will be easy to predicate. Let us see the process of use data flow diagram in which whatever data analyzation done to be stores in hdfs (hadoop distribution file system) by different type of styles like hive,pig and mapreduce and  Power plant analysis can be delivered into client.

## COMPONENT DIAGRAM
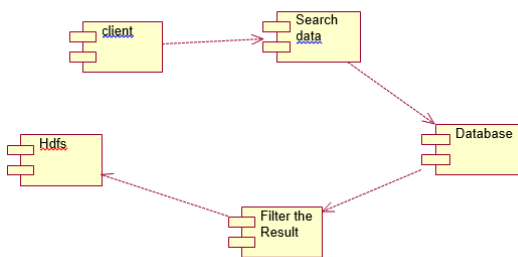


## EXPLANATION:

In the Unified Modeling Language, a component diagram depicts how components are wired together to form larger components and or software systems. They are used to illustrate the structure of arbitrarily complex systems. In component diagram the client can be analysis theclinical data, which will be easy to predicate. Let us see the process of use component diagram in which whatever data analyzation done to be stores in hdfs (hadoop distribution file system) by different type of styles like hive,pig and mapreduce.

## GENERAL

These are the requirements for doing the project. Without using these tools and software's we can't do the project. So we have two requirements to do the project.
 They are

1. Hardware Requirements.
2. Software Requirements.

## 1. HARDWARE REQUIREMENTS

The hardware MapReduce may serve as the basis for a contract for the implementation of the system and should therefore be a complete and consistent specification of the whole system. They are used by software engineers as the starting point for the system design. It shows what the system does and not how it should be implemented.

PROCESSOR :PENTIUM IV 2.6 Hz,Intel Core 2 Duo.
RAM            : 4GB DD RAM
MONITOR    : 15" COLOR
HARD DISK  : 40 GB

## 2 .SOFTWARE REQUIREMENTS

The software requirements document is the specification of the system. It should include both a definition and a specification of requirements. It is a set of what the system should do rather than how it should do it. The software requirements provide a basis for creating the software requirements specification.  It is useful in estimating cost, planning team activities, performing tasks and tracking the teams and tracking the team's progress throughout the development activity.

Framework      : hadoop
Database        :  my sql 5.5
Lanuage        : pig, hive, corejava
Data access tool: sqoop
Operating system: cent os
Ide                    :  eclipse

## IMPLEMENTATION

## GENERAL

 In this we implement the coding part using eclipse. Below are the coding's that are used to generate the domain module for electronic text books. Here the proposed techniques are used in the coding part to generate the e-books.

## EXISTING APPLICATION (MYSQL):

Dataset Creation
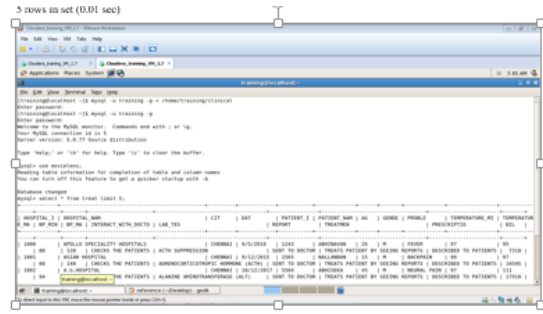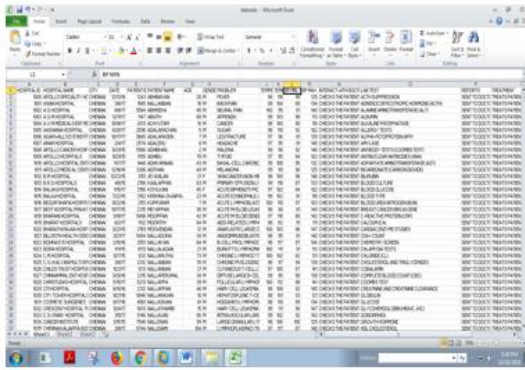
## Restore Backup File into Mysql Database
=============================

[training@localhost ~]$ mysql -u training -p movielens< /home/training/mobility

|---->Database Name
Enter password:training

=============================

## How To Open Mysql Through Terminal
=============================

[training@localhost~]$ [training@localhost ~]$mysql -u training -p
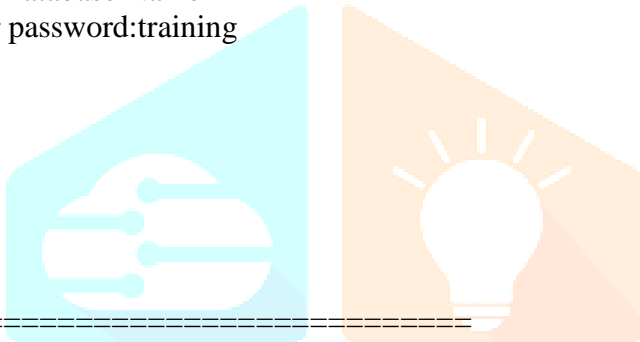
Enter password:

Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 7
Server version: 5.0.77 Source distribution

Type 'help;' or '\h' for help.Type '\c' to clear the buffer.

mysql>
=============================
mysql> show databases;
+--------------------+
| Database          |
+--------------------+
| information_schema |
| movielens         |
| training          |
+--------------------+
3 rows in set (0.03 sec)
=============================
mysql> use movielens;

# SQOOP

=============================**How To Import Table(phonedata)Mysql Database To Hadoop[HDFS]:Using Sqoop Tool**
=============================

[training@localhost ~]$sqoop import --connect jdbc:mysql://localhost/movielens --username training --password training --table treat --m 1 --target-dirclinc;

-------------------------------------------------

18/02/14 22:54:51 INFO mapreduce.ImportJobBase: Beginning import of phonedata
18/02/14 22:54:53 INFO mapred.JobClient: Running job: job_201802142242_0002
18/02/14 22:54:54 INFO mapred.JobClient: map 0% reduce 0%
18/02/14 22:55:02 INFO mapred.JobClient: map 100% reduce 0%
18/02/14 22:55:02 INFO mapred.JobClient: Job complete: job_201802142242_0002
18/02/14 22:55:02 INFO mapred.JobClient: Counters: 12
18/02/14 22:55:02 INFO mapred.JobClient: Job Counters
18/02/14 22:55:02 INFO mapred.JobClient: SLOTS_MILLIS_MAPS=6536
18/02/14 22:55:02 INFO mapred.JobClient: Launched map tasks=1
18/02/14 22:55:02 INFO mapred.JobClient: SLOTS_MILLIS_REDUCES=0
18/02/14 22:55:02 INFO mapred.JobClient: FileSystemCounters
18/02/14 22:55:02 INFO mapred.JobClient: HDFS_BYTES_READ=87
18/02/14 22:55:02 INFO mapred.JobClient: FILE_BYTES_WRITTEN=66226
18/02/14 22:55:02 INFO mapred.JobClient: HDFS_BYTES_WRITTEN=7972205
18/02/14 22:55:02 INFO mapred.JobClient: Map-Reduce Framework

18/02/14 22:55:02 INFO mapred.JobClient: Map input records=40002

18/02/14 22:55:02 INFO mapred.JobClient: Spilled Records=0

18/02/14 22:55:02 INFO mapred.JobClient: Map output records=40002

18/02/14 22:55:02 INFO mapred.JobClient: SPLIT_RAW_BYTES=87

18/02/14 22:55:02 INFO mapreduce.ImportJobBase: Transferred 7.6029 MB in 10.749 seconds (724.2875 KB/sec)

18/09/24 22:02:12 INFO mapreduce.ImportJobBase: Retrieved 49999 records.

==============================

Contents of directory /user/training OUTPUT

==============================

----------------------------------------------------

Name        Type

----------------------------------------------------

CLINCdir

----------------------------------------------------

==============================

Contents of directory /user/training/CLINC/part-m00000

==============================

----------------------------------------------------

Name        Type

----------------------------------------------------

_SUCCESS  file   _logs        dir
part-m-00000  file

----------------------------------------------------

==============================

File: /user/training/CLINC/part-m00000
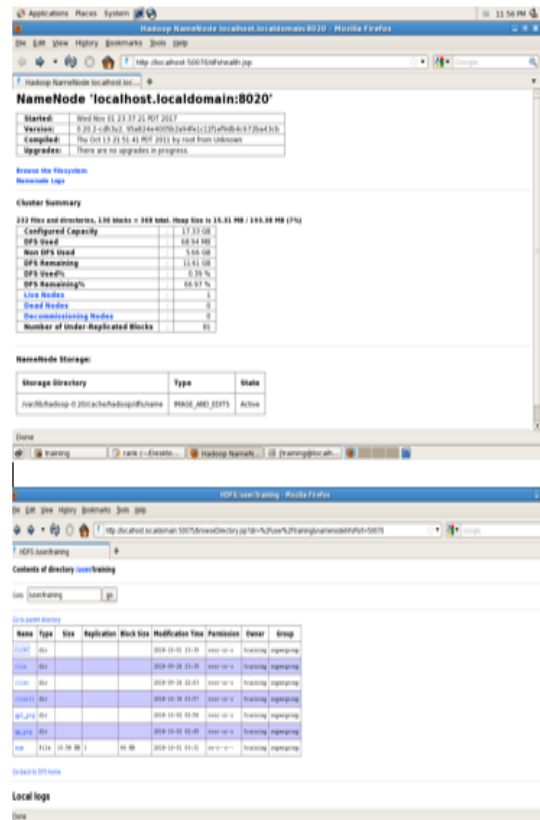




## FUTURE ENHANCEMENTS

Apache Spark is an open source processing engine built around speed, case of use, and analytics. If you have large amounts of data that requires low latency processing that a typical Map Reduce program cannot provide, Spark is the alternative. Spark provides in-memory cluster computing for lightning fast speed and supports Java, Scala, and Python APIs for ease of development.

## CONCLUSION:

In this paper, we presented a study on Clincal is help to doctors by getting idea what treatment he has given to that particular patient. To analysis theClinical data in hadoop ecosystem.Hadoop ecosystem is hive, pig, mapreduce. In future the spark 100 times faster than hadoop, it is easily analysis faster.

**REFERENCES:**

[1] Hermann M, Pentek T, Otto B. Design Principles for Industrie 4.0 Scenarios, pp. 3928-3937, 2016.

[2] Thuemmler C, Bai C. Health 4.0: How Virtualization and Big Data are Revolutionizing Healthcare, Springer International Publishing, 2017.

[3] Cao P, Xian Medical University. "The Impact of Big Data on the Development of Chinese Health Service Industry", Journal of Electronic Test, 2014.

[4] M. Akay, D. I. Fotiadis, K. S. Nikita and R. W. Williams, "Guest Editorial: Biomedical Informatics in Clinical Environments," Biomedical and Health Informatics, IEEE Journal of, vol. 19, no. 1, pp. 149-150, 2015.

[5] Black A D, Car J, Pagliari C, et al. "The impact of eHealth on the quality and safety of health care: a systematic overview", PLoS Medicine, vol. 8, no.1, 2011.

[6] Horsky J, Schiff G D, Johnston D, et al. "Methodological Review: Interface design principles for usable decision support: A targeted review of best practices for clinical prescribing interventions", Journal of Biomedical Informatics, vol. 45, no. 6, pp. 1202-1216, 2012.

[7] Valdez A C, Ziefle M, Verbert K, et al. "Recommender Systems for Health Informatics: State-of-the-Art and Future Perspectives", Machine Learning for Health Informatics, Springer International Publishing, 2016.

[8] Yu W D, Pratiksha C, Swati S, et al. "A Modeling Approach to Big Data Based Recommendation Engine in Modern Health Care Environment", Computer Software and Applications Conference, IEEE Computer Society, pp.75-86, 2015.

[9] Bhatti U A, Huang M, Zhang Y, et al. Research on the Smartphone Based eHealth Systems for Strengthing Healthcare Organization[M]// Smart Health. 2017.

[10] Vancampfort D, Mugisha J, Hallgren M, et al. "The prevalence of diabetes mellitus type 2 in people with alcohol use disorders: a systematic review and large scale meta-analysis", Psychiatry Research, vol. 246, pp. 394-400, 2016.