

# **NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY**

(AN AUTONOMOUS INSTITUTION)

AFFILIATED TO VISVESVARYA TECHNOLOGICAL UNIVERSITY, BELGAUM

APPROVED BY AICTE & GOVT. OF KARNAKATA

YELAHANKA, BANGALORE – 560064

ACADEMIC YEAR 2020-2021



## **Big Data Assignment on Map Reduce and HiveQL**

LA-2 Programming Assignment

Submitted by

**Navachethan M      ISE      1NT18IS099**

### **COURSE C-ORDINATOR**

**Mrs.Ramya.B.S**

Assistant Professor

Department of Information Science and Engineering  
NMIT

## **What is Map Reduce?:**

MapReduce is a Hadoop framework used for writing applications that can process vast amounts of data on large clusters. It can also be called a programming model in which we can process large datasets across computer clusters. This application allows data to be stored in a distributed form. It simplifies enormous volumes of data and large scale computing.

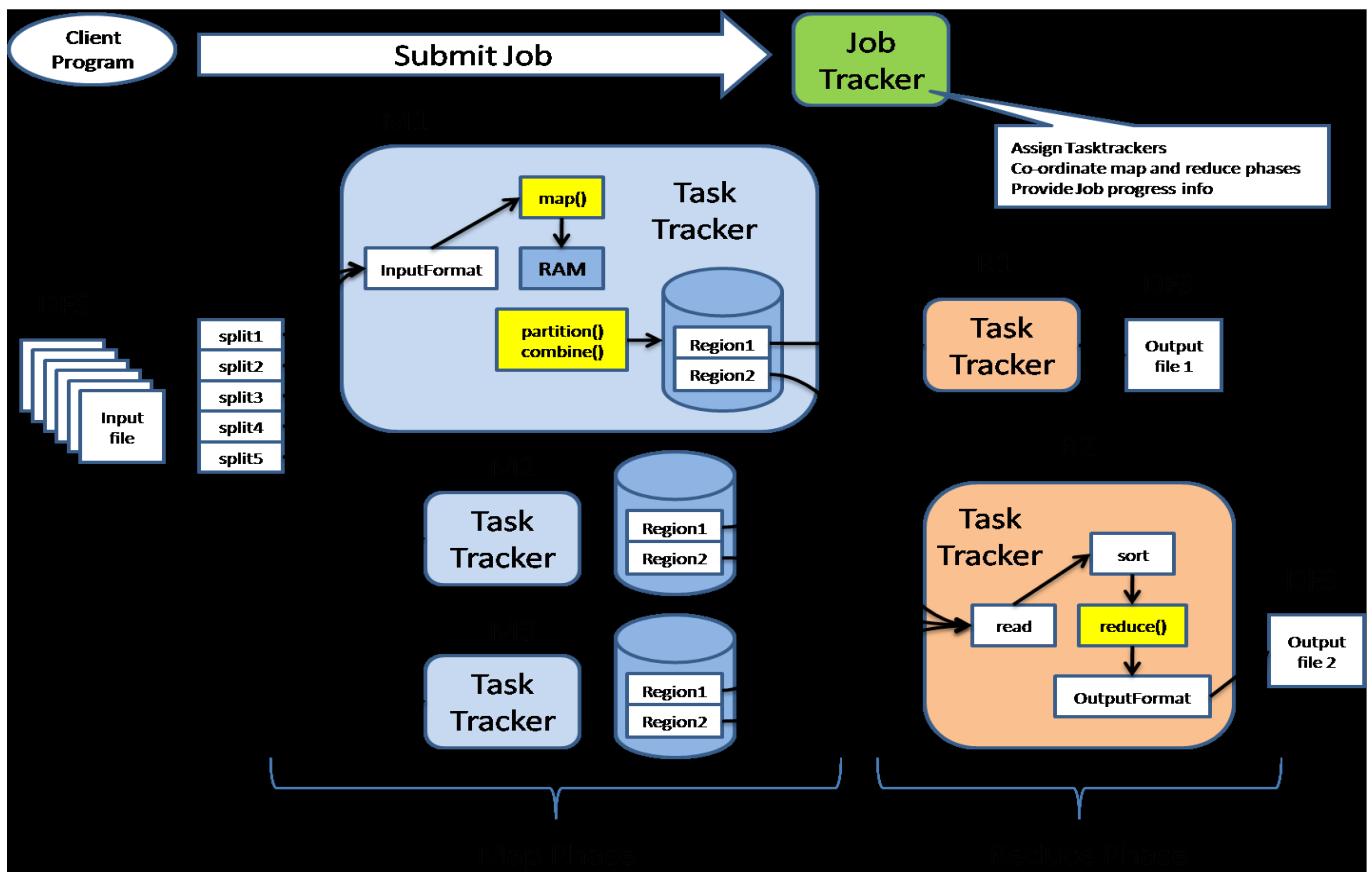
There are two primary tasks in MapReduce: map and reduce. We perform the former task before the latter. In the map job, we split the input dataset into *chunks*. Map task processes these chunks in parallel. The *map* we use outputs as inputs for the *reduce* tasks. Reducers process the intermediate data from the maps into smaller tuples, that reduces the tasks, leading to the final output of the framework.

The MapReduce framework enhances the scheduling and monitoring of tasks. The failed tasks are re-executed by the framework. This framework can be used easily, even by programmers with little expertise in distributed processing. MapReduce can be implemented using various programming languages such as Java, Hive, Pig, Scala, and Python.

# How MapReduce in Hadoop works

An overview of *MapReduce Architecture* and *MapReduce's phases* will help us understand how MapReduce in Hadoop works.

## MapReduce architecture



The following diagram shows a MapReduce architecture.

Image Source: A4Academics

MapReduce architecture consists of various components. A brief description of these components can improve our understanding on how MapReduce works.

- **Job:** This is the actual work that needs to be executed or processed

- **Task:** This is a piece of the actual work that needs to be executed or processed. A MapReduce job comprises many small tasks that need to be executed.
- **Job Tracker:** This tracker plays the role of scheduling jobs and tracking all jobs assigned to the task tracker.
- **Task Tracker:** This tracker plays the role of tracking tasks and reporting the status of tasks to the job tracker.
- **Input data:** This is the data used to process in the mapping phase.
- **Output data:** This is the result of mapping and reducing.
- **Client:** This is a program or Application Programming Interface (API) that submits jobs to the MapReduce. MapReduce can accept jobs from many clients.
- **Hadoop MapReduce Master:** This plays the role of dividing jobs into job-parts.
- **Job-parts:** These are sub-jobs that result from the division of the main job.

In the MapReduce architecture, clients submit jobs to the MapReduce Master. This master will then sub-divide the job into equal sub-parts. The job-parts will be used for the two main tasks in MapReduce: mapping and reducing.

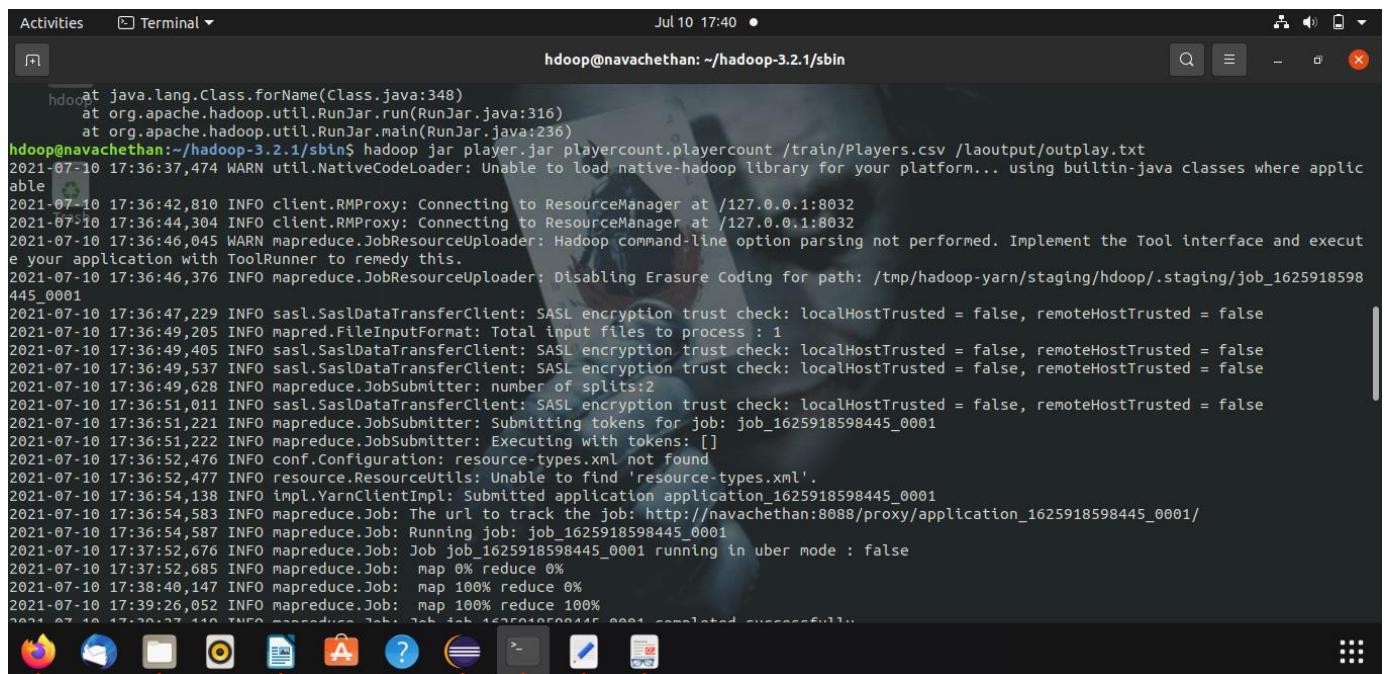
The developer will write logic that satisfies the requirements of the organization or company. The input data will be split and mapped. The intermediate data will then be sorted and merged. The reducer that will generate a final output stored in the HDFS will process the resulting output.

The following diagram shows a simplified flow diagram for the MapReduce program.

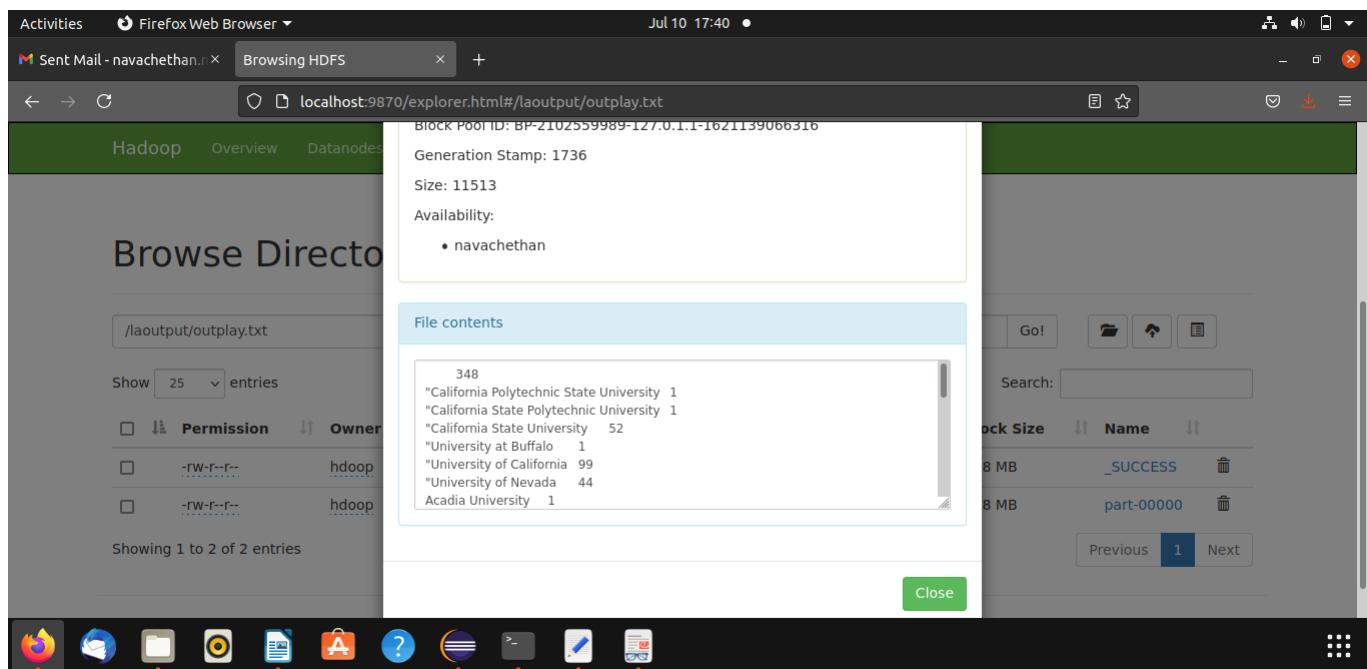


# Map Reduce:

1. To count the number of players from each university:



```
Activities Terminal Jul 10 17:40 • hdoop@navachethan: ~/hadoop-3.2.1/sbin
hdoop@navachethan:~/hadoop-3.2.1/sbin$ hadoop jar player.jar playercount.playercount /train/Players.csv /laoutput/outplay.txt
2021-07-10 17:36:37,474 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-07-10 17:36:42,810 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 17:36:44,304 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 17:36:46,045 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-10 17:36:46,376 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1625918598445_0001
2021-07-10 17:36:47,229 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 17:36:49,205 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-10 17:36:49,405 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 17:36:49,537 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 17:36:49,628 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-10 17:36:51,011 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 17:36:51,221 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625918598445_0001
2021-07-10 17:36:51,222 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-10 17:36:52,476 INFO conf.Configuration: resource-types.xml not found
2021-07-10 17:36:52,477 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-10 17:36:54,138 INFO impl.YarnClientImpl: Submitted application application_1625918598445_0001
2021-07-10 17:36:54,583 INFO mapreduce.Job: The url to track the job: http://navachethan:8088/proxy/application_1625918598445_0001
2021-07-10 17:36:54,587 INFO mapreduce.Job: Running job: job_1625918598445_0001
2021-07-10 17:37:52,676 INFO mapreduce.Job: Job job_1625918598445_0001 running in uber mode : false
2021-07-10 17:37:52,685 INFO mapreduce.Job: map 0% reduce 0%
2021-07-10 17:38:40,147 INFO mapreduce.Job: map 100% reduce 0%
2021-07-10 17:39:26,052 INFO mapreduce.Job: map 100% reduce 100%
2021-07-10 17:39:27,140 INFO mapreduce.Job: Job job_1625918598445_0001 completed successfully.
```



Sent Mail - navachethan.it Browsing HDFS localhost:9870/explorer.html#/laoutput/outplay.txt

Block Pool ID: BP-2102559989-127.0.1.1-1621139066316

Generation Stamp: 1736

Size: 11513

Availability:

- navachethan

File contents

348	"California Polytechnic State University 1
	"California State Polytechnic University 1
	"California State University 52
	"University at Buffalo 1
	"University of California 99
	"University of Nevada 44
	Acadia University 1

Block Size Name SUCCESS part-00000

Go! Search: Previous 1 Next Close

## 2. To find number of players with height 190cms at UNIVERSITY OF CALIFORNIA:

```
Activities Terminal Jul 10 17:56 • hdoop@navachethan: ~/hadoop-3.2.1/sbin
hdoop Bytes Read=285207
File Output Format Counters
Bytes Written=11513
hdoop@navachethan:~/hadoop-3.2.1/sbin$ hadoop jar plyheight.jar playerheight.playerheight /train/Players.csv /laoutput/outheight.txt
2021-07-10 17:52:30,905 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-07-10 17:52:34,866 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 17:52:36,465 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 17:52:37,512 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-10 17:52:37,652 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1625918598445_0002
2021-07-10 17:52:38,371 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 17:52:39,186 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-10 17:52:39,534 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 17:52:39,672 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 17:52:39,730 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-10 17:52:40,543 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 17:52:40,691 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625918598445_0002
2021-07-10 17:52:40,693 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-10 17:52:42,092 INFO conf.Configuration: resource-types.xml not found
2021-07-10 17:52:42,093 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-10 17:52:42,448 INFO impl.YarnClientImpl: Submitted application application_1625918598445_0002
2021-07-10 17:52:42,845 INFO mapreduce.Job: The url to track the job: http://navachethan:8088/proxy/application_1625918598445_0002/
2021-07-10 17:52:42,852 INFO mapreduce.Job: Running job: job_1625918598445_0002
2021-07-10 17:53:12,387 INFO mapreduce.Job: Job job_1625918598445_0002 running in uber mode : false
2021-07-10 17:53:12,521 INFO mapreduce.Job: map 0% reduce 0%
2021-07-10 17:53:54,046 INFO mapreduce.Job: map 100% reduce 0%
2021-07-10 17:55:13,231 INFO mapreduce.Job: map 100% reduce 100%
2021-07-10 17:55:14,004 INFO mapreduce.Job: Job job_1625918598445_0002 completed successfully.
```

Activities Firefox Web Browser Jul 10 17:56 •

Sent Mail - navachethan.rx Browsing HDFS localhost:9870/explorer.html#/laoutput/outheight.txt

Hadoop Overview Datanodes

Browse Director

/laoutput/outheight.txt

Show 25 entries

Permission	Owner
-rW-R--r--	hdoop
-rW-R--r--	hdoop

Showing 1 to 2 of 2 entries

File contents

```
Total no.of players: 5
```

Block Pool ID: BP-2102559989-127.0.1-1621139066316  
Generation Stamp: 1746  
Size: 24  
Availability:  
• navachethan

Go Search: \_SUCCESS part-00000

Block Size Name Previous 1 Next

### 3.To know the number of players who were born in 1928 and represent UNIVERSITY OF CALIFORNIA

```
Activities Terminal • Jul 10 18:04 • hdoop@navachethan: ~/hadoop-3.2.1/sbin
hdoop@navachethan:~/hadoop-3.2.1/sbin$ hadoop jar plybirth.jar playerheight.playerheight /train/Players.csv /laoutput/outbirth.txt
2021-07-10 18:00:21,818 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2021-07-10 18:00:25,656 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 18:00:27,056 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-10 18:00:28,118 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-10 18:00:28,280 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1625918598445_0003
2021-07-10 18:00:28,886 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 18:00:29,768 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-10 18:00:29,971 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 18:00:30,074 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 18:00:30,136 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-10 18:00:30,949 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-10 18:00:31,139 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625918598445_0003
2021-07-10 18:00:31,141 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-10 18:00:32,529 INFO conf.Configuration: resource-types.xml not found
2021-07-10 18:00:32,530 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-10 18:00:33,101 INFO impl.YarnClientImpl: Submitted application application_1625918598445_0003
2021-07-10 18:00:33,467 INFO mapreduce.Job: The url to track the job: http://navachethan:8088/proxy/application_1625918598445_0003/
2021-07-10 18:00:33,476 INFO mapreduce.Job: Running job: job_1625918598445_0003
2021-07-10 18:01:01,860 INFO mapreduce.Job: Job job_1625918598445_0003 running in uber mode : false
2021-07-10 18:01:01,863 INFO mapreduce.Job: map 0% reduce 0%
2021-07-10 18:01:52,251 INFO mapreduce.Job: map 33% reduce 0%
2021-07-10 18:01:53,432 INFO mapreduce.Job: map 83% reduce 0%
2021-07-10 18:01:56,542 INFO mapreduce.Job: map 100% reduce 0%
2021-07-10 18:03:06,140 INFO mapreduce.Job: map 100% reduce 100%
2021-07-10 18:03:06,210 INFO mapreduce.Job: Job job_1625918598445_0003 completed successfully!
```

The screenshot shows a Linux desktop environment with a terminal window and a Firefox browser window.

**Terminal Window:**

- Terminal title: hdoop@navachethan: ~/hadoop-3.2.1/sbin
- Terminal content: Command output for running a Hadoop job to calculate the total number of players born in 1928.

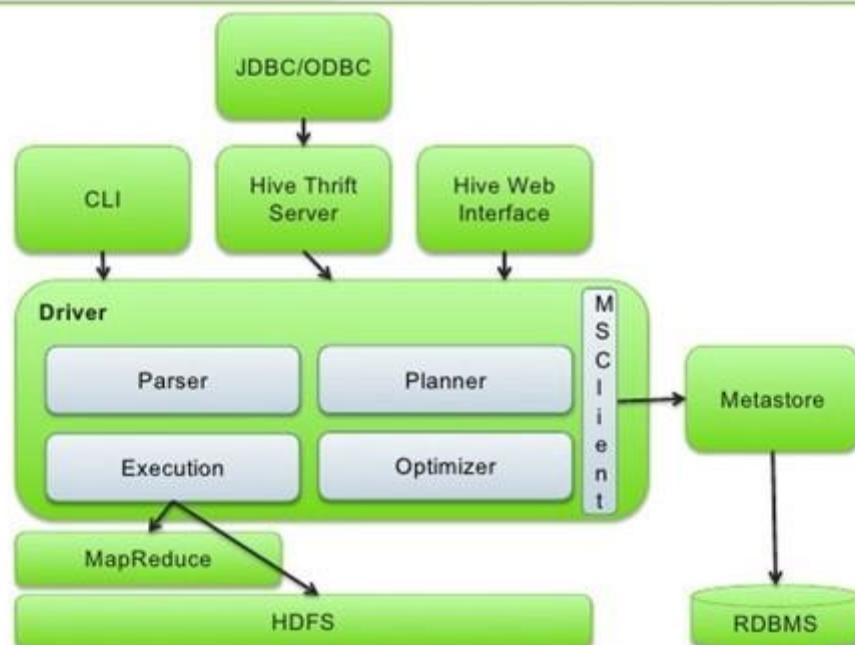
**Firefox Browser Window:**

- Browser title: Browsing HDFS
- URL: localhost:9870/explorer.html#/laoutput/outbirth.txt
- Content:
  - Block Pool ID: BP-2102559989-127.0.1.1-1621139066316
  - Generation Stamp: 1756
  - Size: 24
  - Availability: navachethan
  - File contents:
    - Total no.of players: 5
  - File list:
    - \_SUCCESS (3 MB)
    - part-00000 (3 MB)

# HIVE:

**Apache Hive** is a data warehouse software project built on top of Apache Hadoop for providing data query and analysis.<sup>[3]</sup> Hive gives an SQL-like interface to query data stored in various databases and file systems that integrate with Hadoop. Traditional SQL queries must be implemented in the MapReduce Java API to execute SQL applications and queries over distributed data. Hive provides the necessary SQL abstraction to integrate SQL-like queries (HiveQL) into the underlying Java without the need to implement queries in the low-level Java API. Since most data warehousing applications work with SQL-based querying languages, Hive aids portability of SQL-based applications to Hadoop.

## Apache Hive Architecture



# Hive:

## Hive Queries:

### 0. create database and tables:

```
Activities Terminal Jun 25 07:07 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
hdoop@navachethan:~/apache-hive-3.1.2-bin/bin$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/hdoop/apache-hive-3.1.2-bin/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/hdoop/hadoop-3.2.1/share/hadoop/common/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]
Hive Session ID = bd78e3a1-c40c-439a-9d74-4d1d81681d65
Logging initialized using configuration in jar:file:/home/hdoop/apache-hive-3.1.2-bin/lib/hive-common-3.1.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or
using Hive 1.X releases.
Hive Session ID = 4597a23a-b3e8-4b0e-827f-4fb15eba0e21
hive> show databases;
OK
default
Time taken: 3.294 seconds, Fetched: 1 row(s)
hive> create database if not exists players;
OK
Time taken: 0.956 seconds
hive> use players;
OK
Time taken: 0.111 seconds
hive> create table players(pid int,player string,height int,weight int,college string,born int,b_city string,b_state string) row format delimited file
lds terminated by ',' stored as textfile;
OK
Time taken: 2.802 seconds
hive> desc players;
OK
pid          int
player        string
height        int
weight        int
college       string
born          int
b_city        string
Time taken: 2.802 seconds
hive> desc players;
OK
pid          int
player        string
height        int
weight        int
college       string
born          int
b_city        string
b_state       string
Time taken: 1.419 seconds, Fetched: 8 row(s)
hive> load data local inpath '/home/hdoop/Downloads/Players.csv' into table players;
Loading data to table players.players
OK
Time taken: 3.422 seconds
hive> select * from players;
OK
NULL    Player  NULL    NULL    collage  NULL    birth_city      birth_state
0      Curly Armstrong 180     77    Indiana University   1918
1      Cliff Barker   188     83    University of Kentucky 1921     Yorktown      Indiana
2      Leo Barnhorst  193     86    University of Notre Dame 1924
3      Ed Bartels    196     88    North Carolina State University 1925
4      Ralph Beard    178     79    University of Kentucky 1927     Hardinsburg   Kentucky
5      Gene Berce    180     79    Marquette University 1926
6      Charlie Black  196     90    University of Kansas 1921     Arco        Idaho
```

```
Activities Terminal Jun 25 07:07 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
hdoop@navachethan:~/apache-hive-3.1.2-bin/bin
hive> create database if not exists players;
OK
Time taken: 0.956 seconds
hive> use players;
OK
Time taken: 0.111 seconds
hive> create table players(pid int,player string,height int,weight int,college string,born int,b_city string,b_state string) row format delimited file
lds terminated by ',' stored as textfile;
OK
Time taken: 2.802 seconds
hive> desc players;
OK
pid          int
player        string
height        int
weight        int
college       string
born          int
b_city        string
b_state       string
Time taken: 1.419 seconds, Fetched: 8 row(s)
hive> load data local inpath '/home/hdoop/Downloads/Players.csv' into table players;
Loading data to table players.players
OK
Time taken: 3.422 seconds
hive> select * from players;
OK
NULL    Player  NULL    NULL    collage  NULL    birth_city      birth_state
0      Curly Armstrong 180     77    Indiana University   1918
1      Cliff Barker   188     83    University of Kentucky 1921     Yorktown      Indiana
2      Leo Barnhorst  193     86    University of Notre Dame 1924
3      Ed Bartels    196     88    North Carolina State University 1925
4      Ralph Beard    178     79    University of Kentucky 1927     Hardinsburg   Kentucky
5      Gene Berce    180     79    Marquette University 1926
6      Charlie Black  196     90    University of Kansas 1921     Arco        Idaho
```

```
Activities Terminal Jun 25 07:22 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
3919 Rec Stephen Zimmerman 213 108 "University of Nevada NULL 1996 Hendersonville
3920 Paul Zipser 203 97 1994 Heidelberg Germany
3921 Ivica Zubac 216 120 1997 Mostar Bosnia and Herzegovina
Time taken: 11.08 seconds, Fetched: 3923 row(s)
hive> show databases;
OK
default
players_top
Time taken: 0.182 seconds, Fetched: 2 row(s)
hive> create database if not exists students;
OK
Time taken: 0.18 seconds
hive> show databases;
OK
default
players_top
students
Time taken: 0.195 seconds, Fetched: 3 row(s)
hive> use students;
OK
Time taken: 0.161 seconds
hive> create table student(s_id int,s_name string,s_age int,s_gender string,dept string,hostelite int) row format delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.412 seconds
hive> load data local inpath '/home/hdoop/Downloads/1.txt' into table students;
FAILED: SemanticException [Error 10001]: Line 1:64 Table not found 'students'
hive> desc student;
OK
s_id          int
s_name        string
s_age         int
s_gender      string
dept          string
hostelite    int
Time taken: 0.01 seconds
```

```
Activities Terminal Jun 25 07:23 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
hive> select * from student;
OK
NULL NULL NULL NULL NULL NULL
1 ★ Starr sam 18 M ISE 1
2 raj 19 M CSE 0
3 Hom roy 17 M CSE 0
4 Abhishek 19 M MECH 1
5 Desk Rohini 18 F ISE 0
6 Asha 17 F AERO 0
7 Docu Rishab 17 M CVL 0
8 Kunal 18 M ISE 1
9 Down Megha 19 F MECH 1
10 Shashank 19 M ISE 0
11 Music Nitin 18 M CSE 1
12 Sumanth 18 M AERO 1
13 Nkhil 18 M ECE 0
14 Shilpa 17 F EEE 1
15 Vidya 19 F EEE 0
16 Sagar 18 M ECE 0
17 Vyshnavi 19 F CSE 1
18 Soumya 18 F ISE 0
19 John 18 M ISE 1
20 Priya 18 F CVL 0
21 vikas 17 M ISE 0
22 Aishwarya 18 F AERO 1
23 Varun 18 M CVL 0
24 Shankar 18 M CVL 0
25 Sujay 19 M ECE 1
Time taken: 0.646 seconds, Fetched: 26 row(s)
hive> create table course(s_id int,st_name string,dept_name string,cid string,cname string,faculty string) row format delimited fields terminated by ',' stored as textfile;
OK
Time taken: 0.39 seconds
hive> load data local inpath '/home/hdoop/Downloads/course.txt' into table course;
Loading data to table students.course
Time taken: 0.01 seconds
```

```

Time taken: 0.646 seconds, Fetched: 26 row(s)
hive> create table course(s_id int,st_name string,dept_name string,cid string,cname string,faculty string) row format delimited fields terminated by
',' stored as textfile;
OK
Time taken: 0.39 seconds
hive> load data local inpath '/home/hadoop/Downloads/course.txt' into table course;
Loading data to table students.course
OK
Time taken: 0.776 seconds
hive> select * from course;
OK
NULL NULL NULL NULL NULL
NULL sam ISE 1i cloud mohan
2 Musi raj CSE 2c AI kishore
3 roy CSE 3c iot vivek
4 Abhishek MECH 2m fluid-dynamics mahesh
5 Rohini ISE 2i blockchain keshav
6 Asha AERO 1a propellers sandya
7 Rishab CVL 1cv Environment_engg Raghav
8 Kunal ISE 1i cloud mohan
9 Megha MECH 3m engine_design anand
10 Shashank ISE 3i distributed_systems rohit
11 Nitin CSE 1c HPC Abhishek
12 Sunanth AERO 2a aero_design Inchara
13 Nkhil ECE 1ec VLSI divya
14 Shilpa EEE 1ee S&S anand
15 Vidya EEE 1ee S&S anand
16 Sagar ECE 2ec circuit_design Gaurav
17 Vyshnavi CSE 2c AI kishore
18 Soumya ISE 3i distributed_systems rohit
19 John ISE 1i cloud
20 Priya CVL 2cv structural_engg rakesh
21 vikas ISE 2i blockchain keshav
22 Aishwarya AERO 1a propellers sandya
23 Varun CVL 2cv structural_engg rakesh

```

## 1. Insert 5 records using insert command:

```

23 Rec Varun CVL 2cv structural_engg rakesh
24 Shankar CVL 1cv Environment_engg Raghav
25 Sujay ECE 1ec VLSI divya
Time taken: 0.668 seconds, Fetched: 26 row(s)
hive> insert into student values('26','anil','18','M','CVL','0'),('27','rohan','19','M','EEE','1');
Query ID = hdoop_20210625072741_8f742eba-Sc96-4282-9286-4d09a1353d99
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0001, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0001/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-25 07:28:38,983 Stage-1 map = 0%, reduce = 0%
2021-06-25 07:29:03,842 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.5 sec
2021-06-25 07:29:25,588 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.5 sec
MapReduce Total cumulative CPU time: 11 seconds 430 msec
Ended Job = job_1624584026893_0001
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/students.db/student/.hive-staging_hive_2021-06-25_07-27-41_804_3279798713849659650
-1/-ext-10000
Loading data to table students.student
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 11.43 sec HDFS Read: 20339 HDFS Write: 439 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 430 msec
OK
Time taken: 109.461 seconds
hive> []

```

```
Activities Terminal Jun 25 07:30 •
hadoop@navachethan: ~/apache-hive-3.1.2-bin/bin

Total MapReduce CPU Time Spent: 11 seconds 430 msec
OK
Time taken: 109.461 seconds
hive> select * from student;
OK
26) Hon' anil    18    M      CVL      0
27) rohan     19    M      EEE      1
NULLDeskNULL NULL NULL NULL NULL
1)   sam      18    M      ISE      1
2) Docu rajts  19    M      CSE      0
3)   roy      17    M      CSE      0
4) DowAbhishek 19    M      MECH     1
5) Rohini    18    F      ISE      0
6) Mus Asha   17    F      AERO     0
7) Rishab     17    M      CVL      0
8) Pict Kunal   18    M      ISE      1
9) Megha     19    F      MECH     1
10) Shashank   19    M      ISE      0
11) Nitin     18    M      CSE      1
12) Sumanth   18    M      AERO     1
13) Trash Nkhil 18    M      ECE      0
14) Shilpa    17    F      EEE      1
15) Vidyaions 19    F      EEE      0
16) Sagar     18    M      ECE      0
17) Vyshnavi  19    F      CSE      1
18) Soumya    18    F      ISE      0
19) John      18    M      ISE      1
20) Priya     18    F      CVL      0
21) vikas     17    M      ISE      0
22) Aishwarya  18    F      AERO     1
23) Varun     18    M      CVL      0
24) Shankar   18    M      CVL      0
25) Sujay     19    M      ECE      1
Time taken: 0.542 seconds, Fetched: 28 row(s)
hive>
```



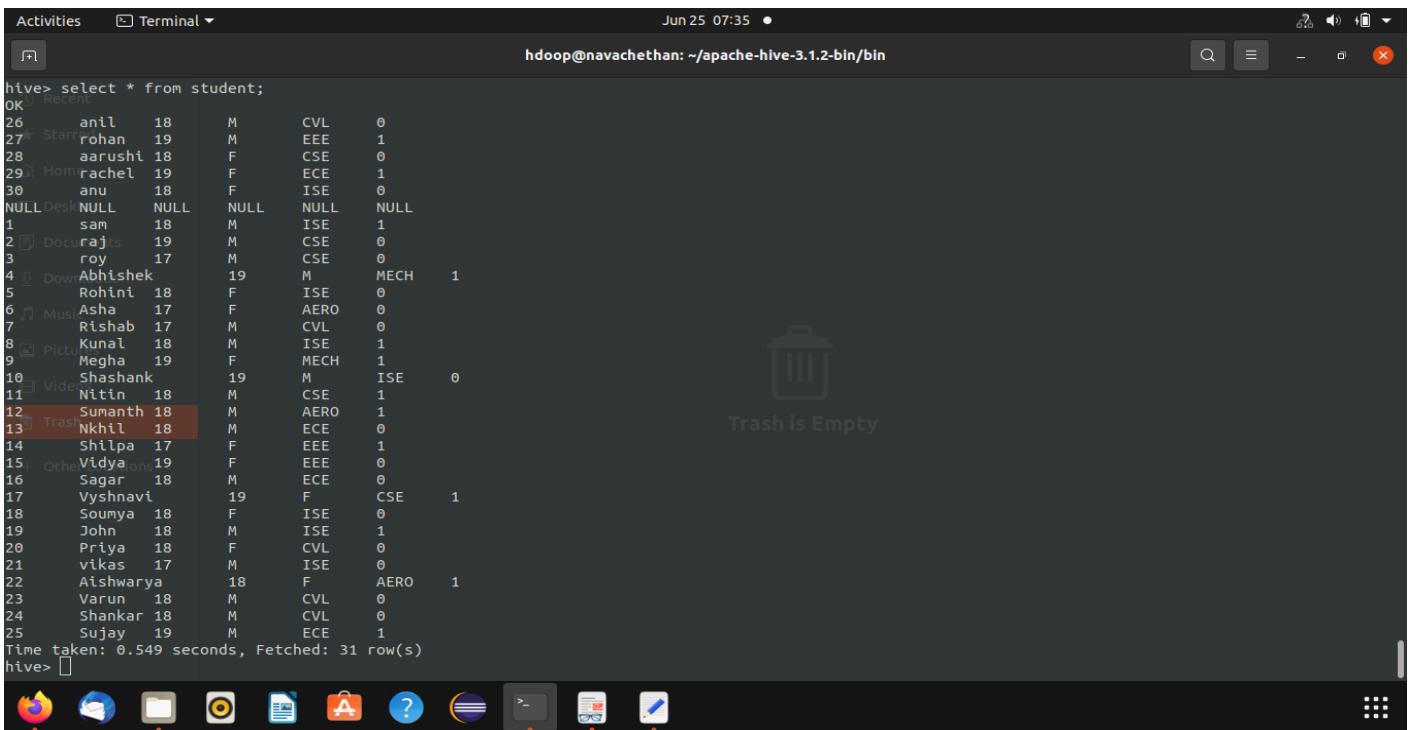
```
Activities Terminal Jun 25 07:34 •
hadoop@navachethan: ~/apache-hive-3.1.2-bin/bin

23) Rec Varun   18    M      CVL      0
24) Shankar  18    M      CVL      0
25) Sujay    19    M      ECE      1
Time taken: 0.542 seconds, Fetched: 28 row(s)
hive> insert into student values('28','aarushi','18','F','CSE','0'),('29','rachel','19','F','ECE','1'),('30','anu','18','F','ISE','0');
Query ID = hdoop_20210625073332_96786514-2f9d-4363-bc77-e1897bc30db4
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0002, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0002/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-25 07:34:05,680 Stage-1 map = 0%, reduce = 0%
2021-06-25 07:34:27,461 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 6.56 sec
2021-06-25 07:34:47,446 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 11.62 sec
MapReduce Total cumulative CPU time: 11 seconds 620 msec
Ended Job = job_1624584026893_0002
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/students.db/student/.hive-staging_hive_2021-06-25_07-33-32_764_401417560279354815-
1/-ext-10000
Loading data to table students.student
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 11.62 sec HDFS Read: 20545 HDFS Write: 492 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 620 msec
OK
Time taken: 77.974 seconds
hive>
```

Activities Terminal Jun 25 07:35 ● hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin

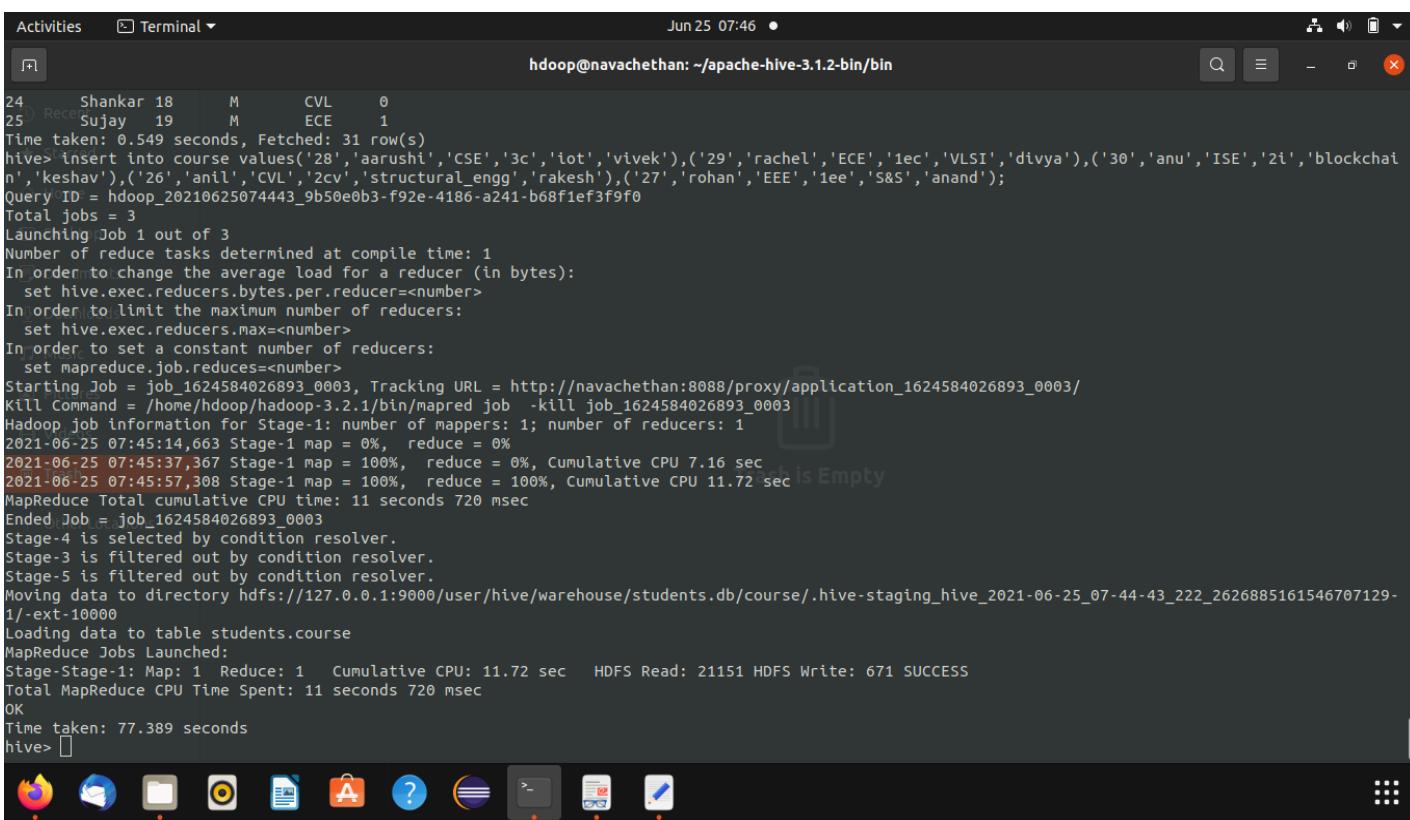
```
hive> select * from student;
OK
26      anil    18     M     CVL    0
27      Starr   rohan  19     M     EEE    1
28      aarushi  18     F     CSE    0
29      Hom     rachel 19     F     ECE    1
30      anu     18     F     ISE    0
NULL NULL  NULL  NULL  NULL
1       sam     18     M     ISE    1
2      Docu    rajts  19     M     CSE    0
3      roy     17     M     CSE    0
4      Dow     Abhishek 19     M     MECH   1
5      Rohini  18     F     ISE    0
6      Mus     Asha   17     F     AERO   0
7      Mus     Rishab  17     M     CVL    0
8      Pictu   Kunal   18     M     ISE    1
9      Megha   19     F     MECH   1
10     Shashank 19     M     ISE    0
11     Vidyansh Nitin   18     M     CSE    1
12     Sumanth  18     M     AERO   1
13     Nkhil   18     M     ECE    0
14     Shilpa   17     F     EEE    1
15     Other    Vidyansh 19     F     EEE    0
16     Sagar   18     M     ECE    0
17     Vyshnavi 19     F     CSE    1
18     Soumya   18     F     ISE    0
19     John    18     M     ISE    1
20     Priya   18     F     CVL    0
21     vikas   17     M     ISE    0
22     Aishwarya 18     F     AERO   1
23     Varun   18     M     CVL    0
24     Shankar  18     M     CVL    0
25     Sujay   19     M     ECE    1
Time taken: 0.549 seconds, Fetched: 31 row(s)
hive>
```

Trash is Empty



Activities Terminal Jun 25 07:46 ● hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin

```
hive> insert into course values('28','aarushi','CSE','3c','iot','vivek'),('29','rachel','ECE','1ec','VLSI','divya'),('30','anu','ISE','2i','blockchain','keshav'),('26','anil','CVL','cv','structural_engg','rakesh'),('27','rohan','EEE','iee','S&S','anand');
Query ID = hdoop_20210625074443_9b50e0b3-f92e-4186-a241-b68f1ef3f9f0
Total jobs = 3
Launching Job 1 out of 3
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0003, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0003/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-25 07:45:14,663 Stage-1 map = 0%, reduce = 0%
2021-06-25 07:45:37,367 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 7.16 sec
2021-06-25 07:45:57,308 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 11.72 sec
MapReduce Total cumulative CPU time: 11 seconds 720 msec
Ended Job = job_1624584026893_0003
Stage-4 is selected by condition resolver.
Stage-3 is filtered out by condition resolver.
Stage-5 is filtered out by condition resolver.
Moving data to directory hdfs://127.0.0.1:9000/user/hive/warehouse/students.db/course/.hive-staging_hive_2021-06-25_07-44-43_222_2626885161546707129-1/-ext-10000
Loading data to table students.course
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 11.72 sec HDFS Read: 21151 HDFS Write: 671 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 720 msec
OK
Time taken: 77.389 seconds
hive>
```



## 2. Load the data(text or CSV) into the table:

Players - Microsoft Excel

The screenshot shows a Microsoft Excel spreadsheet titled "Players". The data is organized into columns: Player (A), height (B), weight (C), collage (D), born (E), birth\_city (F), and birth\_state (G). The data spans from row 1 to 25. The "Formulas" tab is selected in the ribbon. The status bar at the bottom right shows the date as 10-07-2021 and the time as 18:26.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Player	height	weight	collage	born	birth_city	birth_state						
2	0 Curly Armstrong		180	77 Indiana University		1918							
3	1 Cliff Barker		188	83 University of Kentucky		1921 Yorktown	Indiana						
4	2 Leo Barnhorst		193	86 University of Notre Dame		1924							
5	3 Ed Bartels		196	88 North Carolina State University		1925							
6	4 Ralph Beard		178	79 University of Kentucky		1927 Hardinsburg	Kentucky						
7	5 Gene Berce		180	79 Marquette University		1926							
8	6 Charlie Black		196	90 University of Kansas	1921 Arco	Idaho							
9	7 Nelson Bobb		183	77 Temple University	1924 Philadelphia	Pennsylvania							
10	8 Jake Bornheimer		196	90 Muhlenberg College	1927 New Brunswick	New Jersey							
11	9 Vince Boryla		196	95 University of Denver	1927 East Chicago	Indiana							
12	10 Don Boven		193	95 Western Michigan University	1925 Kalamazoo	Michigan							
13	11 Harry Boykoff		208	102 St. John's University	1922 Brooklyn	New York							
14	12 Joe Bradley		190	79 Oklahoma State University	1928 Washington	Oklahoma							
15	13 Bob Brannum		196	97 Michigan State University	1925								
16	14 Carl Braun		196	81 Colgate University	1927 Brooklyn	New York							
17	15 Frankie Brian		185	81 Louisiana State University	1923 Zachary	Louisiana							
18	16 Price Brookfield		193	83 West Texas A&M University	1920 Floydada	Texas							
19	17 Bob Brown		193	92 Miami University	1923 Versailles	Ohio							
20	18 Jim Browne		208	106	1930 Midlothian	Illinois							
21	19 Walt Budko		196	99 Columbia University	1925 Kearney	New Jersey							
22	20 Jack Burmaster		190	86 University of Illinois at Urbana-Champaign	1926								
23	21 Tommy Byrnes		190	79 Seton Hall University	1923 Teaneck	New Jersey							
24	22 Bill Calhoun		190	81 City College of San Francisco	1927 San Francisco	California							
25	23 Don Carlson		183	77 University of Minnesota	1919 Minneapolis	Minnesota							

3. Demonstrate the alter command for the following cases.

a. Rename the table name:

```
Activities Terminal Jun 25 07:48 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
[+]
hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
14 Rec Shilpa EEE 1ee S&S anand
15 Vidya EEE 1ee s&s anand
16 Sagar ECE 2ec circuit_design Gaurav
17 Starr Vyshnavi CSE 2c AI kishore
18 Soumya ISE 3i distributed_systems rohit
19 Hom John ISE 1i cloud
20 Priya CVL 2cv structural_engg rakesh
21 Desk vikas ISE 2l blockchain keshav
22 Aishwarya AERO 1a propellors sandya
23 Docu Varun CVL 2cv structural_engg rakesh
24 Shankar CVL 1cv Environment_engg Raghav
25 Down Sujay ECE 1ec VLSI divya
Time taken: 0.498 seconds, Fetched: 31 row(s)
hive> use players;
OK
Time taken: 0.102 seconds
hive> desc players;
OK
player          int
pid             string
height          int
weight           int
college         string
born            int
b_city          string
b_state         string
Time taken: 0.247 seconds, Fetched: 8 row(s)
hive> alter table players rename to player_details;
OK
Time taken: 0.428 seconds
hive> show tables;
OK
player_details
Time taken: 0.213 seconds, Fetched: 1 row(s)
hive> [ ]
```

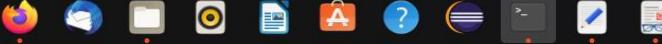


Trash is Empty



b. Rename the column name:

```
Activities Terminal Jun 25 08:24 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
[+]
hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
java.lang.RuntimeException: java.lang.reflect.InvocationTargetException
        at org.apache.hadoop.hive.ql.parse.ParseUtils.parse(ParseUtils.java:67)
        at org.apache.hadoop.hive.ql.Driver.compile(Driver.java:616)
        at org.apache.hadoop.hive.ql.Driver.compileInternal(Driver.java:1826)
        at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:1773)
        at org.apache.hadoop.hive.ql.Driver.compileAndRespond(Driver.java:1768)
        at org.apache.hadoop.hive.ql.reexec.ReExecDriver.compileAndRespond(ReExecDriver.java:126)
        at org.apache.hadoop.hive.ql.reexec.ReExecDriver.run(ReExecDriver.java:214)
        at org.apache.hadoop.hive.cli.CliDriver.processLocalCmd(CliDriver.java:239)
        at org.apache.hadoop.hive.cli.CliDriver.processCmd(CliDriver.java:188)
        at org.apache.hadoop.hive.cli.CliDriver.processLine(CliDriver.java:402)
        at org.apache.hadoop.hive.cli.CliDriver.executeDriver(CliDriver.java:821)
        at org.apache.hadoop.hive.cli.CliDriver.run(CliDriver.java:759)
        at org.apache.hadoop.hive.cli.CliDriver.main(CliDriver.java:683)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at org.apache.hadoop.util.RunJar.run(RunJar.java:323)
        at org.apache.hadoop.util.RunJar.main(RunJar.java:236)
FAILED: ParseException line 1:20 cannot recognize input near 'player' 'p_name' 'string' in alter table statement
hive> alter table players change player p_name string;
OK
Time taken: 0.642 seconds
hive> desc players;
OK
pid          int
p_name        string
height        int
weight        int
college       string
born          int
b_city        string
b_state       string
Time taken: 0.346 seconds, Fetched: 8 row(s)
hive>
```



## 4.AND, OR, IN, NOT IN, SUBSTR, CONCAT, CASE OPERATIONS:

```
Activities Terminal Jun 25 08:47 •
hadoop@navachethan: ~/apache-hive-3.1.2-bin/bin

course
student
Time taken: 0.227 seconds, Fetched: 2 row(s)
hive> desc course;
OK
s_id          int
st_name       string
dept_name    string
cid           string
cname         documents
faculty      string
Time taken: 1.886 seconds, Fetched: 6 row(s)
hive> desc student;
OK
s_id          int
s_name        string
s_age         int
s_gender      string
dept          videos
hostelite    int
Time taken: 0.59 seconds, Fetched: 6 row(s)
hive> select * from student where s_age = '18' and s_gender='M';
OK
+-----+
| id | name | age | gender | dept |
+-----+
| 1  | sam   | 18   | M       | CVL   |
| 2  | Kunal  | 18   | M       | ISE   |
| 3  | Nitin  | 18   | M       | CSE   |
| 4  | Sumanth | 18  | M       | AERO  |
| 5  | Nkhil  | 18   | M       | ECE   |
| 6  | Sagar   | 18  | M       | ECE   |
| 7  | John    | 18   | M       | ISE   |
| 8  | Varun   | 18  | M       | CVL   |
| 9  | Shankar | 18  | M       | CVL   |
+-----+
Time taken: 9.182 seconds, Fetched: 10 row(s)
hive>
```

```
Activities Terminal Jun 25 08:47 •
hadoop@navachethan: ~/apache-hive-3.1.2-bin/bin

12 Rec Sumanth 18 M AERO 1
13 Nkhil 18 M ECE 0
14 Sagar 18 M ECE 0
15 Starr 18 M ISE 1
16 John 18 M ISE 1
17 Varun 18 M CVL 0
18 Shankar 18 M CVL 0
Time taken: 9.182 seconds, Fetched: 10 row(s)
hive> select * from student where s_age = '18' or s_gender='M';
OK
+-----+
| id | name | age | gender | dept |
+-----+
| 1  | sam   | 18   | M       | CVL   | |
| 2  | Kunal  | 18   | M       | ISE   |
| 3  | Nitin  | 18   | M       | CSE   |
| 4  | Sumanth | 18  | M       | AERO  |
| 5  | Nkhil  | 18   | M       | ECE   |
| 6  | Sagar   | 18  | M       | ECE   |
| 7  | John    | 18   | M       | ISE   |
| 8  | Varun   | 18  | M       | CVL   |
| 9  | Shankar | 18  | M       | CVL   |
| 10 | Trashank | 19  | M       | ISE   |
| 11 | Nitin  | 18   | M       | CSE   |
| 12 | Other  | Sumanth | 18  | M       | AERO  |
| 13 | Nkhil  | 18   | M       | ECE   |
| 14 | Sagar   | 18  | M       | ECE   |
| 15 | Soumya  | 18  | F       | ISE   |
| 16 | John    | 18   | M       | ISE   |
| 17 | Priya   | 18   | F       | CVL   |
| 18 | vikas   | 17   | M       | ISE   |
| 19 | Aishwarya | 18  | F       | AERO  |
| 20 | Varun   | 18   | M       | CVL   |
| 21 | Shankar | 18  | M       | CVL   |
| 22 | Sujay   | 19   | M       | ECE   |
+-----+
Time taken: 0.741 seconds, Fetched: 24 row(s)
hive>
```

Activities Terminal Jun 25 08:58 ● hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin

```
hive> select substr('South Carolina State University',7,8);
OK
Carolina
Time taken: 0.579 seconds, Fetched: 1 row(s)
hive> select concat('University of Texas','-', 'Austin');
OK
University of Texas-Austin
Time taken: 0.617 seconds, Fetched: 1 row(s)
hive> show tables;
OK
Documents
player_detail
players
Time taken: 0.19 seconds, Fetched: 2 row(s)
hive> select * from player_detail;
OK
107 Tony Lavelli    190     83      Yale University 1926      Somerville      Massachusetts
1668 Butch Graves   190     90      Yale University 1962      Scarsdale      New York
Time taken: 0.81 seconds, Fetched: 2 row(s)
hive> desc player_detail;
OK
pid          int
p_name        string
height        int
weight        int
college       string
born          int
b_city         string
b_state        string
Time taken: 0.32 seconds, Fetched: 8 row(s)
hive> select upper(p_name) from player_detail;
OK
TONY LAVELLI
BUTCH GRAVES
Time taken: 0.846 seconds, Fetched: 2 row(s)
hive> 
```

Trash is Empty

Activities Terminal Jun 25 08:59 ● hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin

```
player_detail
players
Time taken: 0.19 seconds, Fetched: 2 row(s)
hive> select * from player_detail;
OK
107 Tony Lavelli    190     83      Yale University 1926      Somerville      Massachusetts
1668 Butch Graves   190     90      Yale University 1962      Scarsdale      New York
Time taken: 0.81 seconds, Fetched: 2 row(s)
hive> desc player_detail;
OK
Documents
pid          int
p_name        string
height        int
weight        int
college       string
born          int
b_city         string
b_state        string
Time taken: 0.32 seconds, Fetched: 8 row(s)
hive> select upper(p_name) from player_detail;
OK
TONY LAVELLI
BUTCH GRAVES
Time taken: 0.846 seconds, Fetched: 2 row(s)
hive> select lower(p_name) from player_detail;
OK
tony lavelli
butch graves
Time taken: 0.728 seconds, Fetched: 2 row(s)
hive> select length(p_name) from player_detail;
OK
12
12
Time taken: 0.674 seconds, Fetched: 2 row(s)
hive> 
```

Trash is Empty

Activities Terminal Jun 25 12:11

```
hadoop@navachethan: ~/apache-hive-3.1.2-bin/bin
ain only 1 item in Select List.
hive> select s.s_id,s.s_name from student s where s.s_id in(select s.* from student s);
FAILED: SemanticException org.apache.hadoop.hive.ql.optimizer.calcite.CalciteSubquerySemanticException: Invalid SubQuery expression SubQuery can contain only 1 item in Select List.
hive> select * from student where student.s_id in(select s_id from course where s_id is not null);
Query ID = hdoop_20210625120615_f2240d52-8a74-4d7e-b4b6-d3239acc06a9
Total jobs = 1
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1624584026893_0013, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0013/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0013
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-25 12:07:29,266 Stage-3 map = 0%, reduce = 0%
2021-06-25 12:07:50,855 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 6.49 sec
MapReduce Total cumulative CPU time: 6 seconds 490 msec
Ended Job = job_1624584026893_0013
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 6.49 sec HDFS Read: 10596 HDFS Write: 1027 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 490 msec
OK
26 Other 18 M CVL 0
27 rohan 19 M EEE 1
28 aarushi 18 F CSE 0
29 rachel 19 F ECE 1
30 anu 18 F ISE 0
2 raj 19 M CSE 0
3 roy 17 M CSE 0
4 Abhishek 19 M MECH 1
5 Rohini 18 F ISE 0
6 Asha 17 F AERO 0
7 Rishab 17 M CVL 0
8 Kunal 18 M ISE 1
9 Megha 19 F MECH 1
```

Activities Terminal Jun 25 12:21

```
hadoop@navachethan: ~/apache-hive-3.1.2-bin/bin
Stage-Stage-2: Map: 1 Reduce: 1 Cumulative CPU: 10.55 sec HDFS Read: 13630 HDFS Write: 112 SUCCESS
Stage-Stage-3: Map: 1 Cumulative CPU: 4.26 sec HDFS Read: 6411 HDFS Write: 87 SUCCESS
Total MapReduce CPU Time Spent: 14 seconds 810 msec
OK
Time taken: 165.22 seconds
hive> select * from student where student.s_id not in(select s_id from course where s_id is not null);
Warning: Map Join MAPJOIN[42][bigTable=] in task 'Stage-7:MAPRED' is a cross product
Query ID = hdoop_20210625121514_81474d76-ef34-4a35-8851-b78cbb94b1dc
Total jobs = 5
Launching Job 1 out of 5
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0017, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0017/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0017
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 1
2021-06-25 12:15:46,710 Stage-3 map = 0%, reduce = 0%
2021-06-25 12:16:11,433 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 6.8 sec
2021-06-25 12:16:31,176 Stage-3 map = 100%, reduce = 100%, Cumulative CPU 11.34 sec
MapReduce Total cumulative CPU time: 11 seconds 340 msec
Ended Job = job_1624584026893_0017
Launching Job 2 out of 5
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0018, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0018/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0018
```

## 5. Aggregate functions COUNT,MIN,MAX,SUM,AVG:

```
Activities Terminal Jun 25 08:27 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
[+]
pid    hdoop      int
p_name   string
height   int
weight   int
college  string
born     int
b_city   string
b_state  string
Time taken: 0.346 seconds, Fetched: 8 row(s)
hive> select count(*) from players;
Query ID = hdoop_20210625082551_55962492-c7cf-40cf-841c-4edb84b643be
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0004, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0004/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-25 08:26:33,619 Stage-1 map = 0%, reduce = 0%
2021-06-25 08:26:54,845 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.27 sec
2021-06-25 08:27:14,507 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 8.85 sec
MapReduce Total cumulative CPU time: 8 seconds 850 msec
Ended Job = job_1624584026893_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 8.85 sec HDFS Read: 294400 HDFS Write: 104 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 850 msec
OK
3923
Time taken: 85.667 seconds, Fetched: 1 row(s)
hive> [ ]
```

```
Activities Terminal Jun 25 08:31 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
[+]
Time taken: 85.667 seconds, Fetched: 1 row(s)
hive> select sum(height) from players group by college;
Query ID = hdoop_20210625082931_a85737c8-3a4e-4bbe-ac45-7ec320b6df1a
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0005, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0005/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-25 08:30:02,923 Stage-1 map = 0%, reduce = 0%
2021-06-25 08:30:23,010 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.2 sec
2021-06-25 08:30:43,205 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.35 sec
MapReduce Total cumulative CPU time: 10 seconds 350 msec
Ended Job = job_1624584026893_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 10.35 sec HDFS Read: 295303 HDFS Write: 6929 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 350 msec
OK
71105
193
185
10337
206
19761
8830
188
211
198
829
---
```

```
Activities Terminal Jun 25 08:35 hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin

Time taken: 74.016 seconds, Fetched: 417 row(s)
hive> select college,avg(height) from players group by college;
Query ID = hdoop_20210625083155_f2d94a61-accf-426f-854b-92308cd17b09
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0006, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0006/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-25 08:32:26,690 Stage-1 map = 0%, reduce = 0%
2021-06-25 08:32:46,976 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.7 sec
2021-06-25 08:33:09,177 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 12.66 sec
MapReduce Total cumulative CPU time: 12 seconds 660 msec
Ended Job = job_1624584026893_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 12.66 sec HDFS Read: 297084 HDFS Write: 19798 SUCCESS
Total MapReduce CPU Time Spent: 12 seconds 660 msec
OK
      204.32471264367817
*California Polytechnic State University      193.0
*California State Polytechnic University      185.0
*California State University     198.78846153846155
*University at Buffalo    206.0
*University of California      199.6060606060606
*University of Nevada      200.6818181818182
Acadia University      188.0
Alabama A&M University    211.0
Alabama State University    198.0
Albany State University    207.25

 Trash is Empty
```

Activities Terminal Jun 25 08:37

hadoop@navachethan: ~/apache-hive-3.1.2-bin/bin

```
Time taken: 75.854 seconds, Fetched: 417 row(s)
hive> select college,min(height),max(height) from players group by college;
Query ID = hdoop_20210625083549_134be098-723d-457a-bd0b-7a02a153963f
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
| set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
| set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
| set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0007, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0007/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0007
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-25 08:36:20,121 Stage-1 map = 0%, reduce = 0%
2021-06-25 08:36:42,516 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 5.79 sec
2021-06-25 08:37:03,573 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 11.08 sec
MapReduce Total cumulative CPU time: 11 seconds 80 msec
Ended Job = job_1624584026893_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 11.08 sec HDFS Read: 295952 HDFS Write: 18985 SUCCESS
Total MapReduce CPU Time Spent: 11 seconds 80 msec
OK
      175      231
"California Polytechnic State University"      193      193
"California State Polytechnic University"      185      185
"California State University"      178      211
"University at Buffalo"      206      206
"University of California"      178      224
"University of Nevada"      180      213
Acadia University      188      188
Alabama A&M University      211      211
Alabama State University      198      198
Albany State University      206      211
```

## 6.Create separate view:

```
Activities Terminal Jun 25 08:43 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
Whitworth      2
Wichita State University      11
Wilberforce University      1
William Paterson University      2
Wingate University      1
Winston-Salem State University      5
Wofford College      1
Wright State University      2
Xavier University      16
Xavier University of Louisiana      3
Yale University      2
Youngstown State University      1
college      1
nan      1
Time taken: 72.276 seconds, Fetched: 417 row(s)
hive> desc players;
OK
pid          int
p_name        string
height        int
weight        int
college       string
born          string
b_city        string
b_state       string
Time taken: 0.294 seconds, Fetched: 8 row(s)
hive> create view player_detail as select * from players where college='Yale University';
OK
Time taken: 1.172 seconds
hive> select * from player_detail;
OK
107    Tony Lavelli    190    83    Yale University 1926    Somerville    Massachusetts
1668   Butch Graves    190    90    Yale University 1962    Scarsdale    New York
Time taken: 0.805 seconds, Fetched: 2 row(s)
hive> 
```

## 7.Group by:

```
Activities Terminal Jun 25 08:40 • hdoop@navachethan: ~/apache-hive-3.1.2-bin/bin
FAILED: SemanticException [Error 10001]: Line 1:29 Table not found 'plays'
hive> select college,count(*) from players group by college;
Query ID = hdoop_20210625083847_7c9eef3b-69d3-45c2-b293-9b70be8784aa
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0008, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0008/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-06-25 08:39:17,211 Stage-1 map = 0%, reduce = 0%
2021-06-25 08:39:37,356 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.94 sec
2021-06-25 08:39:57,190 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 10.03 sec
MapReduce Total cumulative CPU time: 10 seconds 30 msec
Ended Job = job_1624584026893_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 10.03 sec HDFS Read: 295061 HDFS Write: 16604 SUCCESS
Total MapReduce CPU Time Spent: 10 seconds 30 msec
OK
348
"California Polytechnic State University      1
"California State Polytechnic University      1
"California State University      52
"University at Buffalo      1
"University of California      99
"University of Nevada      44
Acadia University      1
Alabama A&M University      1
Alabama State University      1
Albany State University      4
```

8. Perform the following joins (Outer, Left outer, Right outer):

```
Activities Terminal Jun 25 10:53 •
hadoop@navachethan:~/apache-hive-3.1.2-bin/bin

faculty          string
Time taken: 0.377 seconds, Fetched: 6 row(s)
hive> select s.s_id,s.s_name,s.dept,c cname from student s course c on(s.s_id = c.s_id);
FAILED: ParseException line 1:53 missing EOF at 'course' near 's'
hive> select s.s_id,s.s_name,s.dept,c cname from student s join course c on(s.s_id = c.s_id);
Query ID = hdoop_20210625105107_0ca93b2b-2085-4ee9-824a-a1eb3596546f
Total jobs = 1
[Desktop]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1624584026893_0009, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0009/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0009
Hadoop job Information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-25 10:52:24,337 Stage-3 map = 0%, reduce = 0%
2021-06-25 10:52:48,019 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 6.43 sec
MapReduce Total cumulative CPU time: 6 seconds 430 msec
Ended Job = job_1624584026893_0009
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 6.43 sec HDFS Read: 10693 HDFS Write: 1111 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 430 msec
OK
28      aarushi  CSE      iot
29      rachel   ECE      VLSI
30      anu      ISE      blockchain
26      anil     CVL      structural_engg
27      rohan   EEE      S&S
2       raj     CSE      AI
3       roy     CSE      iot
4       Abhishek  MECH    fluid-dynamics
5       Rohini  ISE      blockchain
6       Asha    AERO    propellors
7       Rishab   CVL    Environment_engg
[Trash is Empty]
```

```
Activities Terminal Jun 25 10:55 •
hadoop@navachethan: ~/apache-hive-3.1.2-bin/bin

hive> select s.s_id,s.s_name,s.dept,c cname from student s left outer join course c on(s.s_id = c.s_id);
Query ID = hdoop_20210625105338_4a02cf6b-ad44-4410-81a1-1cbfd6be00b2
Total jobs = 1
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1624584026893_0010, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0010/
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0010
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-25 10:54:47,315 Stage-3 map = 0%, reduce = 0%
2021-06-25 10:55:08,881 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 5.7 sec
MapReduce Total cumulative CPU time: 5 seconds 700 msec
Ended Job = job_1624584026893_0010
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 5.7 sec HDFS Read: 10006 HDFS Write: 1160 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 700 msec
OK
26 anil CVL structural_engg
27 rohan EEE S&S
28 aarushi CSE iot
29 rachel ECE VLSI
30 anu ISE blockchain
31 NULL NULL NULL
1 sam ISE NULL
2 raj CSE AI
3 roy CSE iot
4 Abhishek MECH fluid-dynamics
5 Rohini ISE blockchain
6 Asha AERO propellers
7 Rishab CVL Environment_engg
8 Kunal ISE cloud
9 Megha MECH engine_design
10 Shashank ISE distributed_systems
11 Nitin CSE HPC
Trash is Empty
```

Activities Terminal Jun 25 10:59

```
25 Sujay ECE VLSI
Time taken: 93.294 seconds, Fetched: 31 row(s)
hive> select s.s_id,s.s_name,s.dept,c cname from student s right outer join course c on(s.s_id = c.s_id);
Query ID = hdoop_20210625105749_baac51a1-2ff5-43d6-b170-915cfbdd13a0
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.2021-06-25 10:58:23 Uploaded 1 File to: file:/tmp/hadoop/ec2946ff-aeab-4f1b-861d-2a4c60a5379b/hive
_2021-06-25_10-57-50_010_6166543297516995939-1/_local-10004/HashTable-Stage-3/MapJoin-mapfile20--.hashtable (1169 bytes)
Execution completed successfully
MapredLocal task succeeded
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1624584026893_0011, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0011
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0011
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-06-25 10:58:53,628 Stage-3 map = 0%, reduce = 0%
2021-06-25 10:59:16,119 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 5.37 sec
MapReduce Total cumulative CPU time: 5 seconds 370 msec
Ended Job = job_1624584026893_0011
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1 Cumulative CPU: 5.37 sec HDFS Read: 10210 HDFS Write: 1162 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 370 msec
OK
28 - Other aarushi CSE iot
29 rachel ECE VLSI
30 anu ISE blockchain
26 anil CVL structural_engg
27 rohan EEE S&S
NULL NULL NULL NULL
NULL NULL NULL cloud
2 raj CSE AI
3 roy CSE iot
4 Abhishek MECH fluid-dynamics
5 Rohini ISE blockchain
6 Asha AERO propellors
7 Rishab CVL Environment_engg
```

Trash is Empty

Activities Terminal Jun 25 11:02

```
Time taken: 88.539 seconds, Fetched: 31 row(s)
hive> select s.s_id,s.s_name,s.dept,c cname from student s full outer join course c on(s.s_id = c.s_id);
Query ID = hdoop_20210625105952_3edd7497-645f-4d70-a9a3-bb3563d4459b
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks not specified. Estimated from input data size: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set.mapreduce.job.reduces=<number>
Starting Job = job_1624584026893_0012, Tracking URL = http://navachethan:8088/proxy/application_1624584026893_0012
Kill Command = /home/hadoop/hadoop-3.2.1/bin/mapred job -kill job_1624584026893_0012
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2021-06-25 11:00:26,723 Stage-1 map = 0%, reduce = 0%
2021-06-25 11:01:16,019 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 12.14 sec
2021-06-25 11:01:42,456 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 19.06 sec
MapReduce Total cumulative CPU time: 19 seconds 60 msec
Ended Job = job_1624584026893_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1 Cumulative CPU: 19.06 sec HDFS Read: 19087 HDFS Write: 1211 SUCCESS
Total MapReduce CPU Time Spent: 19 seconds 60 msec
OK
NULL NULL NULL NULL
NULL NULL NULL cloud
NULL NULL NULL NULL
1 sam ISE NULL
2 raj CSE AI
3 roy CSE iot
4 Abhishek MECH fluid-dynamics
5 Rohini ISE blockchain
6 Asha AERO propellors
7 Rishab CVL Environment_engg
8 Kunal ISE cloud
```

Trash is Empty

# **Conclusion:**

This report is all about usage of Map-Reduce concept on huge data and retrieves the useful information from them. It is easy to manage the data and learn the data behaviour through Map-Reduce programs. Also another useful tool over Hadoop is the HIVEQL which is used to get the result from the dataset using simple queries just like SQL.

All the above programs and commands are present in the following github link:

[https://github.com/Navachethan-Murugeppa/1NT18IS099\\_navachethan\\_C\\_bdLab](https://github.com/Navachethan-Murugeppa/1NT18IS099_navachethan_C_bdLab)