

**Institute of Computer Technology**  
**B. Tech. Computer Science and Engineering**  
**Sub: Big Data Analytics**  
**Project Report**

**1. AIM of the project given to you.**

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worldwide. Heart failure is a common event caused by CVDs and this dataset contains 12 features that can be used to predict mortality by heart failure. Most cardiovascular diseases can be prevented by addressing behavioural risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity and harmful use of alcohol using population-wide strategies. People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management.

**2. Name of students along with enrolment numbers who were included in the project- along with contribution of each student in the project.**

Pandey Manthan (19162121016)

Parmar Ankit (19162121017)

Patel Darshil (19162121018)

**3. The tool chosen to execute the project in- Explanation of the tool as well as the conceptualization of using the tool WRT the project.**

**What is hive?**

Hive in Big Data is an easy-to-use software application that lets one analyze large-scale data through the batch processing technique. An efficient program, it uses a familiar software that uses HiveQL, a language that is very similar to SQL- structured query language used for interaction with databases.

Such a software can be operated by both programmers and non-programmers, making it a very accessible and easy-to-use application for converting petabytes of data into useful data strands. This is one of the biggest benefits of Apache Hive that has made it a popular choice for data analytics among large organizations with vast data.

**Why are we using hive?**

Hive was created to allow non-programmers familiar with SQL to work with petabytes of data, using a SQL-like interface called HiveQL. Traditional relational databases are designed for interactive queries on small to medium datasets and do not process huge datasets well. Hive instead uses batch processing so that it works quickly across a very large distributed database.

Hive transforms HiveQL queries into MapReduce or Tez jobs that run on Apache Hadoop's distributed job scheduling framework, Yet Another Resource Negotiator (YARN). It queries data stored in a distributed storage solution, like the Hadoop Distributed File System. Hive stores its database and table metadata in a metastore, which is a database or file backed store that enables easy data abstraction and discovery.

#### 4. Code used for execution of the project.

##### 1. Load csv file:

- cd Desktop
- `hadoop fs -mkdir miniproject`
- `Hadoop fs -put heart_failure_clinical_records_dataset.csv miniproject`
- `Hadoop fs -ls miniproject`
- `hive`
- `create table bda_project`  
(age INT, anaemia SMALLINT, creatinine\_phosphokinase INT, diabetes SMALLINT, ejection\_fraction INT, high\_blood\_pressure SMALLINT, platelets INT, serum\_creatinine FLOAT, serum\_sodium INT, sex BINARY, smoking SMALLINT, time INT, DEATH\_EVENT SMALLINT)
- row format delimited fields terminated by ',';
- `LOAD DATA LOCAL INPATH 'Desktop/ heart_failure_clinical_records_dataset.csv' INTO TABLE bda_project`
- `select *from bda_project;`

##### 2. Execution of project:

- `select count(anaemia)+count(diabetes) from bda_project where Age>60;`
- `select count(sex) from bda_project where sex=1;`
- `select (count(sex)/194)*100 from bda_project where smoking=0 AND sex=1;`
- `select count(sex) from bda_project where age<50 AND DEATH_EVENT=1;`
- `select count(sex) from bda_project where serum_creatinine > 2 AND sex = 1;`
- `select count(sex) from bda_project where sex = 0 and time <= 20;`
- `select count(sex) from bda_project;`
- `select (count(sex)/299)*100 from bda_project where (anaemia=0 OR diabetes=0) AND smoking=0 AND DEATH_EVENT=1;`
- `select count(sex) from bda_project;`
- `select (count(sex)/299)*100 from bda_project where high_blood_pressure=1 AND DEATH_EVENT=0;`
- `select count(sex) from bda_project where platelets < 150000;`
- `select count(sex) from bda_project where high_blood_pressure=1;`

## 5. Problem definition and output:

- How many people above the age of 60 are both anaemic and diabetic?

```

cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select count(Anaemia)+count(diabetes) from bdaproject where Age>60;
Query ID = cloudera_20211025020404_ef4c0513-4a1e-4431-bb5f-a63d53819b56
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_1635149137368_0004, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0004/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0004
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:04:53,473 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:04:59,816 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.5 sec
2021-10-25 02:05:06,187 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 6.15 sec
MapReduce Total cumulative CPU time: 6 seconds 150 msec
Ended Job = job_1635149137368_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 6.15 sec HDFS Read: 22469 H
DFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 150 msec
OK
270

```

- What percentage of women in the dataset do not smoke?

```

cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select count(sex) from bda_project where sex=1;
Query ID = cloudera_20211025020606_d7ebac75-f943-47c2-ae10-fa5aee43f7f0
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_1635149137368_0005, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0005/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0005
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:06:14,660 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:06:19,960 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.34 sec
2021-10-25 02:06:26,331 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.6 sec
MapReduce Total cumulative CPU time: 2 seconds 600 msec
Ended Job = job_1635149137368_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.6 sec HDFS Read: 21058 HD
DFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 600 msec
OK
194
Time taken: 18.483 seconds, Fetched: 1 row(s)

```

```

cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select (count(sex)/194)*100 from bda_project where smoking=0 AND sex=1;
Query ID = cloudera_20211025020808_ba0ca558-777c-4344-8b0e-2f1a65d55c4b
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_1635149137368_0006, Tracking URL = http://quickstart.cloudera:8080/proxy/application_1635149137368_0006/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:08:49,675 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:08:54,963 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.7 sec
2021-10-25 02:09:00,234 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.33 sec
MapReduce Total cumulative CPU time: 4 seconds 330 msec
Ended Job = job_1635149137368_0006
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.33 sec HDFS Read: 22186 HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 330 msec
OK
52.57731958762887
Time taken: 14.759 seconds, Fetched: 1 row(s)

```

- How many men and women below the age of 50 died?

```

cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select count(sex) from bdaproject where age <50 AND DEATH EVENT=1;
Query ID = cloudera_20211025021010_5e485ed9-63bb-47b9-b3c7-5d227d327b5a
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_1635149137368_0008, Tracking URL = http://quickstart.cloudera:8080/proxy/application_1635149137368_0008/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0008
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:10:48,299 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:10:53,567 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.32 sec
2021-10-25 02:10:58,875 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.62 sec
MapReduce Total cumulative CPU time: 2 seconds 620 msec
Ended Job = job_1635149137368_0008
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.62 sec HDFS Read: 21797 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 620 msec
OK
11
Time taken: 16.38 seconds, Fetched: 1 row(s)

```

- How many females have a serum\_creatinine level of above 2?

```

cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select count(sex) from bda_project where serum_creatinine > 2 AND sex = 1;
Query ID = cloudera_20211025021212_1523da18-404b-4d07-9c7f-fac9fc675d42
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_1635149137368_0009, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0009/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0009
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:12:30,925 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:12:35,198 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.34 sec
2021-10-25 02:12:41,540 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.78 sec
MapReduce Total cumulative CPU time: 2 seconds 780 msec
Ended Job = job_1635149137368_0009
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.78 sec HDFS Read: 21521 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 780 msec
OK
21
Time taken: 17.442 seconds, Fetched: 1 row(s)
hive>

```

- How many men have been called for a follow up in 20 days?

```

cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select count(sex) from bda_project where sex = 0 and time <= 20;
Query ID = cloudera_20211025023333_94a0bb12-0367-4da3-b0c1-0e83b53c2a47
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_1635149137368_0010, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0010/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0010
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:33:07,735 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:33:13,006 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.55 sec
2021-10-25 02:33:19,279 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.93 sec
MapReduce Total cumulative CPU time: 2 seconds 930 msec
Ended Job = job_1635149137368_0010
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.93 sec HDFS Read: 21537 HDFS Write: 2 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 930 msec
OK
7
Time taken: 17.406 seconds, Fetched: 1 row(s)
hive>

```

- What percentage of people who weren't anaemic, or diabetic and did not smoke, died?

```

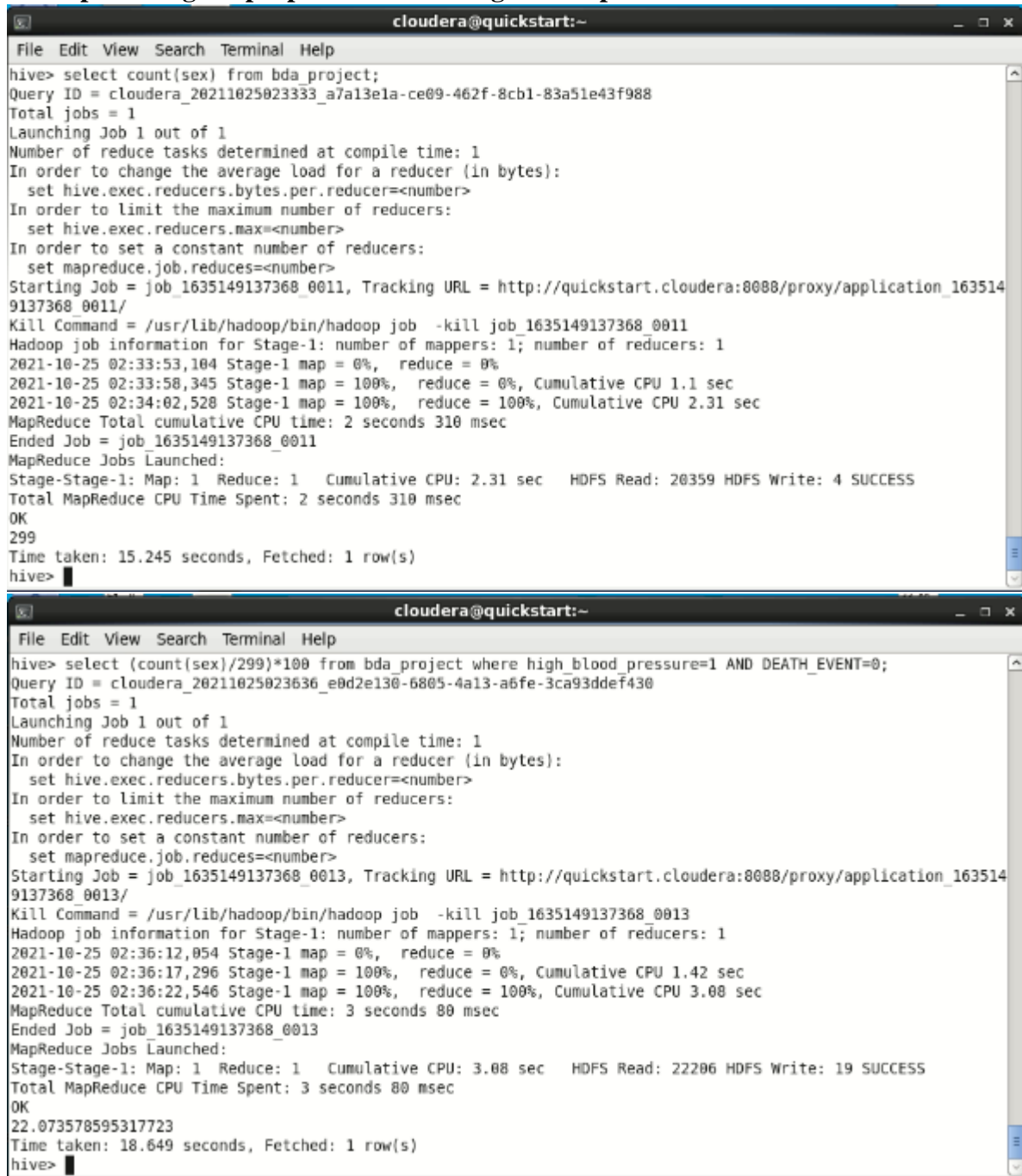
cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select count(sex) from bda_project;
Query ID = cloudera_20211025023333_a7a13e1a-ce09-462f-8cb1-83a51e43f988
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks determined at compile time: 1
In order to change the average load 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_1635149137368_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0011/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0011
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:33:53,104 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:33:58,345 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.1 sec
2021-10-25 02:34:02,528 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.31 sec
MapReduce Total cumulative CPU time: 2 seconds 310 msec
Ended Job = job_1635149137368_0011
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.31 sec HDFS Read: 20359 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 310 msec
OK
299
Time taken: 15.245 seconds, Fetched: 1 row(s)
hive>

cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select (count(sex)/299)*100 from bda_project where (anaemia=0 OR diabetes=0) AND smoking=0 AND DEATH_EVENT=1;
Query ID = cloudera_20211025023434_8495d86e-c41c-404b-a405-435e090f5e47
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_1635149137368_0012, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0012/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0012
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:34:51,184 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:34:56,415 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.4 sec
2021-10-25 02:35:01,714 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.93 sec
MapReduce Total cumulative CPU time: 2 seconds 930 msec
Ended Job = job_1635149137368_0012
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.93 sec HDFS Read: 22596 HDFS Write: 18 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 930 msec
OK
17.05685618729097
Time taken: 16.798 seconds, Fetched: 1 row(s)
hive>

```



- What percentage of people who have a high blood pressure did not die?



```
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> select count(sex) from bda_project;  
Query ID = cloudera_20211025023333_a7a13e1a-ce09-462f-8cb1-83a51e43f988  
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_1635149137368_0011, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0011/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0011  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2021-10-25 02:33:53,104 Stage-1 map = 0%, reduce = 0%  
2021-10-25 02:33:58,345 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.1 sec  
2021-10-25 02:34:02,528 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.31 sec  
MapReduce Total cumulative CPU time: 2 seconds 310 msec  
Ended Job = job_1635149137368_0011  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.31 sec HDFS Read: 20359 HDFS Write: 4 SUCCESS  
Total MapReduce CPU Time Spent: 2 seconds 310 msec  
OK  
299  
Time taken: 15.245 seconds, Fetched: 1 row(s)  
hive>  
  
cloudera@quickstart:~  
File Edit View Search Terminal Help  
hive> select (count(sex)/299)*100 from bda_project where high blood pressure=1 AND DEATH_EVENT=0;  
Query ID = cloudera_20211025023636_e0d2e130-6805-4a13-a6fe-3ca93ddef430  
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_1635149137368_0013, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0013/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0013  
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1  
2021-10-25 02:36:12,054 Stage-1 map = 0%, reduce = 0%  
2021-10-25 02:36:17,296 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.42 sec  
2021-10-25 02:36:22,546 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 3.08 sec  
MapReduce Total cumulative CPU time: 3 seconds 80 msec  
Ended Job = job_1635149137368_0013  
MapReduce Jobs Launched:  
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 3.08 sec HDFS Read: 22206 HDFS Write: 19 SUCCESS  
Total MapReduce CPU Time Spent: 3 seconds 80 msec  
OK  
22.073578595317723  
Time taken: 18.649 seconds, Fetched: 1 row(s)  
hive>
```

- How many people have a platelet count of lesser than 150000?

```

cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select count(sex) from bda_project where platelets < 150000;
Query ID = cloudera_20211025023737_8e781a36-eb23-439e-a6ce-121fba7afc41
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_1635149137368_0014, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0014/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0014
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:37:09,820 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:37:16,121 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 4.17 sec
2021-10-25 02:37:21,404 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 5.47 sec
MapReduce Total cumulative CPU time: 5 seconds 470 msec
Ended Job = job_1635149137368_0014
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 5.47 sec HDFS Read: 21148 HDFS Write: 3 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 470 msec
OK
27
Time taken: 17.806 seconds, Fetched: 1 row(s)
hive>

```

- How many men and women have high blood pressure?

```

cloudera@quickstart:~
File Edit View Search Terminal Help
hive> select count(sex) from bda_project where high blood pressure=1;
Query ID = cloudera_20211025024141_c1716dd4-3804-4e28-8b32-9bab2f296804
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_1635149137368_0016, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1635149137368_0016/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1635149137368_0016
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-10-25 02:41:06,462 Stage-1 map = 0%, reduce = 0%
2021-10-25 02:41:11,733 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 1.35 sec
2021-10-25 02:41:18,000 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 2.69 sec
MapReduce Total cumulative CPU time: 2 seconds 690 msec
Ended Job = job_1635149137368_0016
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 2.69 sec HDFS Read: 21139 HDFS Write: 4 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 690 msec
OK
105
Time taken: 17.365 seconds, Fetched: 1 row(s)
hive>

```



## **6. Conclusion**

Hive is a Data Warehousing package built on top of Hadoop used for data analysis. Hive also uses a language called HiveQL (HQL) which automatically translates SQL-like queries into MapReduce jobs. We have also learned various components of Hive like meta store etc.