NITTE EDUCATION TRUST | NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

DEPARTMENT OF INFORMATION SCIENCE AND ENGINEERING

NITTE MEENAKSHI INSTITUTE OF TECHNOLOGY

(An autonomous institution with A+ Grade by NAAC /UGC, Affiliated to Visvesvaraya Technological University, Belgaum, Approved by UGC/AICTE/Govt. of Karnataka)

Yelahanka, Bengaluru-560064

ACADEMIC YEAR 2020-2021

BIGDATA LABORATORY

Report on,

Learning Activity II-Programming Assignment

Submitted by,

Aditya N Bhatt (1NT18IS015)

Submitted

Disha D N,
Assistant Professor,
Department of Information Science and Engineering,
Nitte Meenakshi Institute of Technology,
Bangalore-064

Table of Contents

1. LIST OF FIGURES

2. BRIEF ABOUT HADOOP AND MAP REDUCE

3. HADOOP MAP-REDUCE PROBLEM STATEMENT

- **I. DATASET DESCRIPTION**
- **II. SOURCE CODE**
- **III. RESULTS AND SNAPSHOTS**

4. BRIEF ABOUT HIVE

5. HIVE PROBLEM STATEMENT

- **I. DATASET DESCRIPTION**
- **II. QUERIES, RESULTS AND SNAPSHOTS**

6. REFERENCES

7. Terminal execution

Brief note on Hadoop and Map Reduce

Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. It is highly fault-tolerant and is designed to be deployed on low-cost hardware. It provides high throughput access to application data and is suitable for applications having large datasets. MapReduce is a parallel programming model for writing distributed applications devised at Google for efficient processing of large amounts of data (multi-terabyte data-sets), on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. The MapReduce program runs on Hadoop which is an Apache open-source framework. It is quite expensive to build bigger servers with heavy configurations that handle large scale processing, but as an alternative, you can tie together many commodity computers with single-CPU, as a single functional distributed system and practically, the clustered machines can read the dataset in parallel and provide a much higher throughput.

Hadoop Map-reduce Problem statement

Create a dataset in excel as .csv file and it should contain the following fields with at least 20 sample datasets in it.

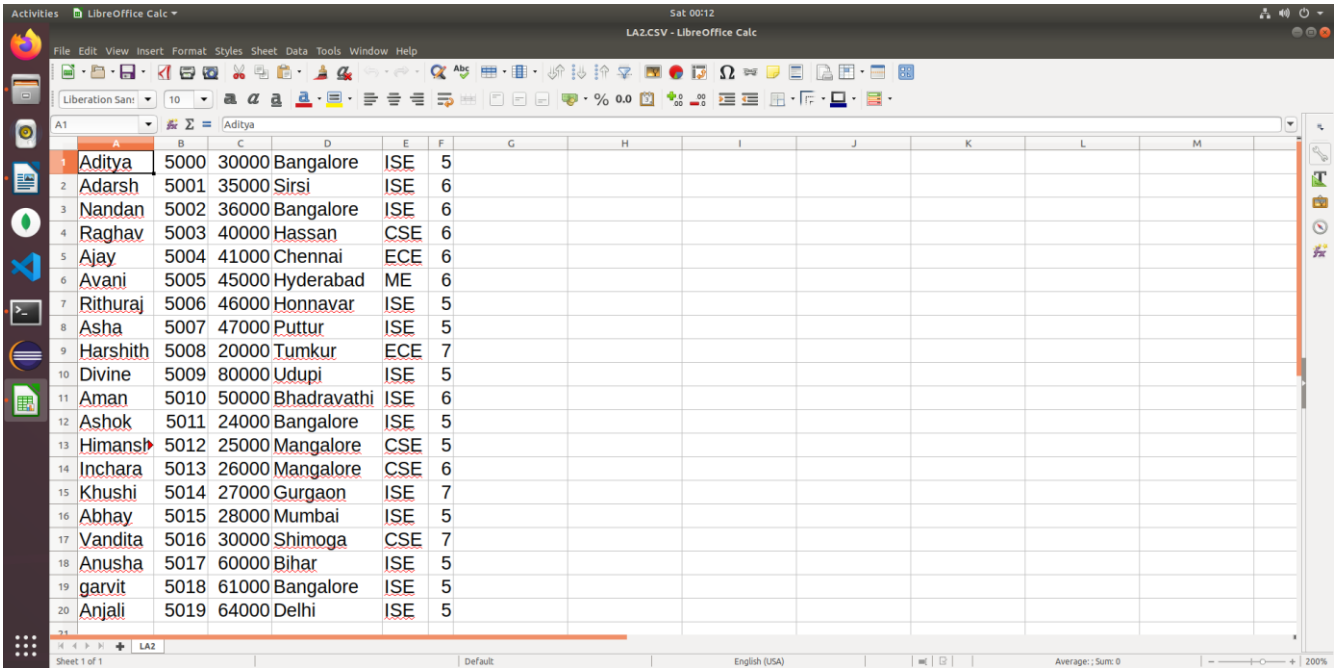
Name	SSN	Salary	Dname	Experience
Harsha	5000	30000	ISE	5

Use the Hadoop MapReduce programming framework to come up with a Program which will take the data from this .csv file and computes the following.

1. Total number of employees who work in ISE department
2. Total number of employees with experience=5 years
3. Count the number of employees who lives in Bangalore

Database Description

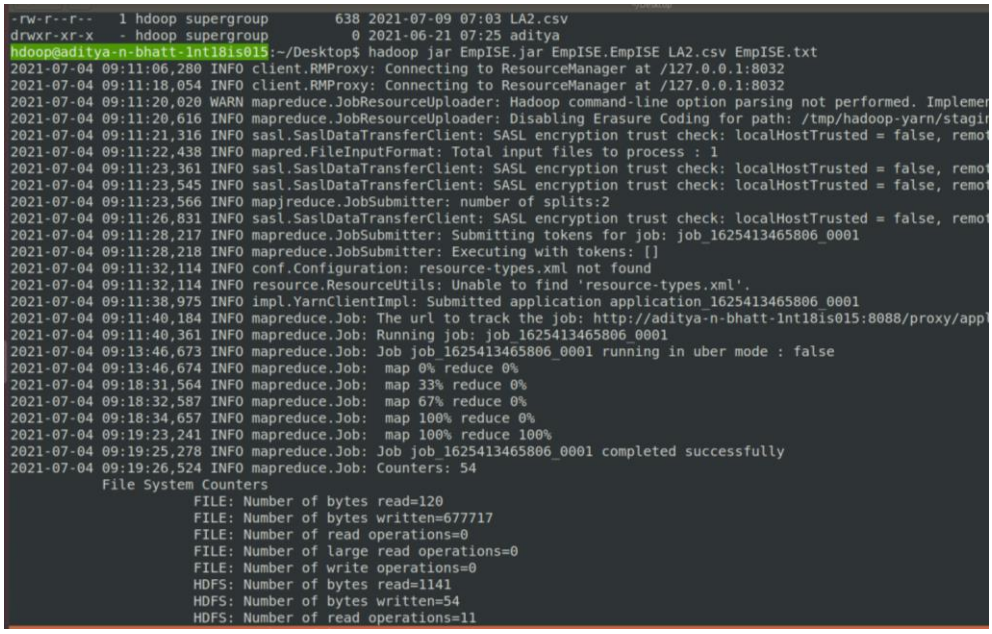
LA2.csv



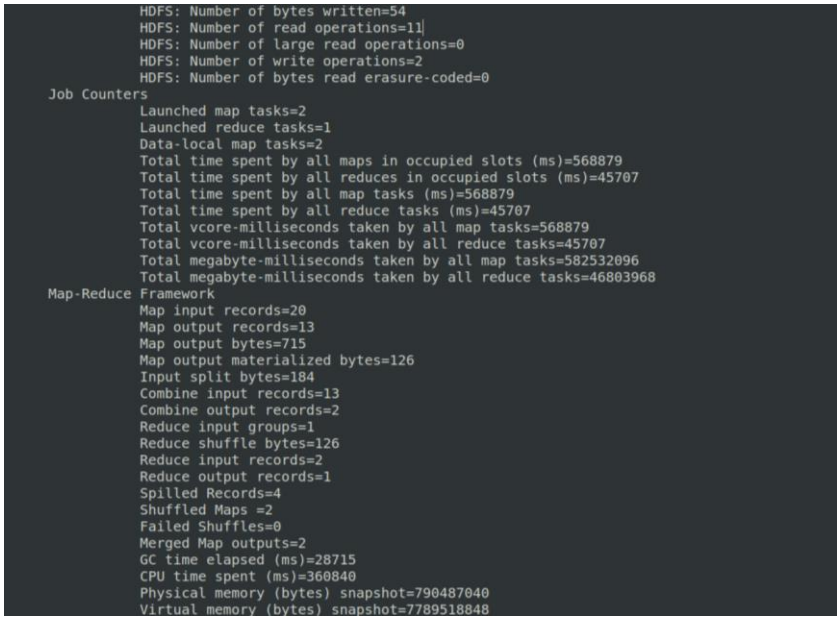
	A	B	C	D	E	F
1	Aditya	5000	30000	Bangalore	ISE	5
2	Adarsh	5001	35000	Sirsi	ISE	6
3	Nandan	5002	36000	Bangalore	ISE	6
4	Raghav	5003	40000	Hassan	CSE	6
5	Ajay	5004	41000	Chennai	ECE	6
6	Avani	5005	45000	Hyderabad	ME	6
7	Rithuraj	5006	46000	Honnavar	ISE	5
8	Asha	5007	47000	Puttur	ISE	5
9	Harshith	5008	20000	Tumkur	ECE	7
10	Divine	5009	80000	Udupi	ISE	5
11	Aman	5010	50000	Bhadravathi	ISE	6
12	Ashok	5011	24000	Bangalore	ISE	5
13	Himansh	5012	25000	Mangalore	CSE	5
14	Inchara	5013	26000	Mangalore	CSE	6
15	Khushi	5014	27000	Gurgaon	ISE	7
16	Abhay	5015	28000	Mumbai	ISE	5
17	Vandita	5016	30000	Shimoga	CSE	7
18	Anusha	5017	60000	Bihar	ISE	5
19	garvit	5018	61000	Bangalore	ISE	5
20	Anjali	5019	64000	Delhi	ISE	5

Results and Snapshot (Hadoop Map-reduce Programming)

1. Total number of employees who work in ISE department



```
-rw-r--r--   1 hdoop supergroup        638 2021-07-09 07:03 LA2.csv
drwxr-xr-x   - hdoop supergroup          0 2021-06-21 07:25 aditya
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop jar EmpISE.jar EmpISE.EmpISE LA2.csv EmpISE.txt
2021-07-04 09:11:06,280 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-04 09:11:18,054 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-04 09:11:20,020 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implemer
2021-07-04 09:11:20,616 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/stagir
2021-07-04 09:11:21,316 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
2021-07-04 09:11:22,438 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-04 09:11:23,361 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
2021-07-04 09:11:23,545 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
2021-07-04 09:11:23,566 INFO mapjreduce.JobSubmitter: number of splits:2
2021-07-04 09:11:26,831 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remot
2021-07-04 09:11:28,217 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625413465806_0001
2021-07-04 09:11:28,218 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-04 09:11:32,114 INFO conf.Configuration: resource-types.xml not found
2021-07-04 09:11:32,114 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-04 09:11:38,975 INFO impl.YarnClientImpl: Submitted application application_1625413465806_0001
2021-07-04 09:11:40,184 INFO mapreduce.Job: The url to track the job: http://aditya-n-bhatt-1nt18is015:8088/proxy/appl
2021-07-04 09:11:40,361 INFO mapreduce.Job: Running job: job_1625413465806_0001
2021-07-04 09:13:46,673 INFO mapreduce.Job: Job job_1625413465806_0001 running in uber mode : false
2021-07-04 09:13:46,674 INFO mapreduce.Job:  map 0% reduce 0%
2021-07-04 09:18:31,564 INFO mapreduce.Job:  map 33% reduce 0%
2021-07-04 09:18:32,587 INFO mapreduce.Job:  map 67% reduce 0%
2021-07-04 09:18:34,657 INFO mapreduce.Job:  map 100% reduce 0%
2021-07-04 09:19:23,241 INFO mapreduce.Job:  map 100% reduce 100%
2021-07-04 09:19:25,278 INFO mapreduce.Job: Job job_1625413465806_0001 completed successfully
2021-07-04 09:19:26,524 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=120
                FILE: Number of bytes written=677717
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1141
                HDFS: Number of bytes written=54
                HDFS: Number of read operations=11
```



```
                HDFS: Number of bytes written=54
                HDFS: Number of read operations=11
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
        Job Counters
                Launched map tasks=2
                Launched reduce tasks=1
                Data-local map tasks=2
                Total time spent by all maps in occupied slots (ms)=568879
                Total time spent by all reduces in occupied slots (ms)=45707
                Total time spent by all map tasks (ms)=568879
                Total time spent by all reduce tasks (ms)=45707
                Total vcore-milliseconds taken by all map tasks=568879
                Total vcore-milliseconds taken by all reduce tasks=45707
                Total megabyte-milliseconds taken by all map tasks=582532096
                Total megabyte-milliseconds taken by all reduce tasks=46803968
        Map-Reduce Framework
                Map input records=20
                Map output records=13
                Map output bytes=715
                Map output materialized bytes=126
                Input split bytes=184
                Combine input records=13
                Combine output records=2
                Reduce input groups=1
                Reduce shuffle bytes=126
                Reduce input records=2
                Reduce output records=1
                Spilled Records=4
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=28715
                CPU time spent (ms)=360840
                Physical memory (bytes) snapshot=790487040
                Virtual memory (bytes) snapshot=7789518848
```

```
                Map input records=20
                Map output records=13
                Map output bytes=715
                Map output materialized bytes=126
                Input split bytes=184
                Combine input records=13
                Combine output records=2
                Reduce input groups=1
                Reduce shuffle bytes=126
                Reduce input records=2
                Reduce output records=1
                Spilled Records=4
                Shuffled Maps =2
                Failed Shuffles=0
                Merged Map outputs=2
                GC time elapsed (ms)=28715
                CPU time spent (ms)=360840
                Physical memory (bytes) snapshot=790487040
                Virtual memory (bytes) snapshot=7789518848
                Total committed heap usage (bytes)=644874240
                Peak Map Physical memory (bytes)=304123904
                Peak Map Virtual memory (bytes)=2595966976
                Peak Reduce Physical memory (bytes)=183201792
                Peak Reduce Virtual memory (bytes)=2604199936
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=957
        File Output Format Counters
                Bytes Written=54
doop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls EmpISE.txt
ound 2 items
```

```
                Bytes Written=54
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls EmpISE.txt
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2021-07-09 06:36 EmpISE.txt/_SUCCESS
-rw-r--r--   1 hdoop supergroup         54 2021-07-09 09:36 EmpISE.txt/part-00000
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -cat EmpISE.txt/part-00000
2021-07-04 09:20:31,372 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = fal
Total no.of employees working in ISE Department :      13
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop jar EmpExp.jar EmpExp.EmpExp LA2.csv EmpExp.txt
```

2. Total number of employees with experience=5 years

```
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -cat EmpISE.txt/part-00000
2021-07-04 09:20:31,372 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrus
Total no.of employees working in ISE Department :      13
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop jar EmpExp.jar EmpExp.EmpExp LA2.csv EmpExp.txt
2021-07-09 06:34:04,465 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 06:34:09,024 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 06:34:10,405 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the To
2021-07-09 06:34:10,709 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/
2021-07-09 06:34:11,286 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrus
2021-07-09 06:34:12,643 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-09 06:34:12,836 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrus
2021-07-09 06:34:12,862 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrus
2021-07-09 06:34:12,989 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-09 06:34:13,318 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrus
2021-07-09 06:34:13,366 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625413465806_0002
2021-07-09 06:34:13,366 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-09 06:34:13,903 INFO conf.Configuration: resource-types.xml not found
2021-07-09 06:34:13,904 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-09 06:34:16,566 INFO impl.YarnClientImpl: Submitted application application_1625413465806_0002
2021-07-09 06:34:18,577 INFO mapreduce.Job: The url to track the job: http://aditya-n-bhatt-1nt18is015:8088/proxy/application_
2021-07-09 06:34:18,712 INFO mapreduce.Job: Running job: job_1625413465806_0002
2021-07-09 06:35:23,814 INFO mapreduce.Job: Job job_1625413465806_0002 running in uber mode : false
2021-07-09 06:35:23,818 INFO mapreduce.Job:  map 0% reduce 0%
2021-07-09 06:36:21,633 INFO mapreduce.Job:  map 100% reduce 0%
2021-07-09 06:36:26,679 INFO mapreduce.Job:  map 100% reduce 100%
2021-07-09 06:36:27,708 INFO mapreduce.Job: Job job_1625413465806_0002 completed successfully
2021-07-09 06:36:27,841 INFO mapreduce.Job: Counters: 55
        File System Counters
                FILE: Number of bytes read=126
                FILE: Number of bytes written=677738
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1141
                HDFS: Number of bytes written=57
                HDFS: Number of read operations=11
                HDFS: Number of large read operations=0
```

```
            HDFS: Number of bytes read=1141
            HDFS: Number of bytes written=57
            HDFS: Number of read operations=11
            HDFS: Number of large read operations=0
            HDFS: Number of write operations=2
            HDFS: Number of bytes read erasure-coded=0
    Job Counters
            Killed map tasks=1
            Launched map tasks=2
            Launched reduce tasks=1
            Data-local map tasks=2
            Total time spent by all maps in occupied slots (ms)=107735
            Total time spent by all reduces in occupied slots (ms)=2063
            Total time spent by all map tasks (ms)=107735
            Total time spent by all reduce tasks (ms)=2063
            Total vcore-milliseconds taken by all map tasks=107735
            Total vcore-milliseconds taken by all reduce tasks=2063
            Total megabyte-milliseconds taken by all map tasks=110320640
            Total megabyte-milliseconds taken by all reduce tasks=2112512
    Map-Reduce Framework
            Map input records=20
            Map output records=10
            Map output bytes=580
            Map output materialized bytes=132
            Input split bytes=184
            Combine input records=10
            Combine output records=2
            Reduce input groups=1
            Reduce shuffle bytes=132
            Reduce input records=2
            Reduce output records=1
            Spilled Records=4
            Shuffled Maps =2
            Failed Shuffles=0
            Merged Map outputs=2
            GC time elapsed (ms)=4467
```

```
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls EmpExp.txt
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2021-07-09 06:36 EmpExp.txt/_SUCCESS
-rw-r--r--   1 hdoop supergroup         57 2021-07-09 06:36 EmpExp.txt/part-00000
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -cat EmpExp.txt/part-00000
2021-07-09 06:36:55,206 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localH
Total no.of employees having 5 years of experience :   10
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop jar EmpAddress.jar EmpAddress.EmpAddress LA
```

3. Count the number of employees who lives in Bangalore

```
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop jar EmpAddress.jar EmpAddress.EmpAddress LA2.csv EmpAddress.txt
2021-07-09 06:38:36,404 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 06:38:36,566 INFO client.RMProxy: Connecting to ResourceManager at /127.0.0.1:8032
2021-07-09 06:38:36,727 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the T
2021-07-09 06:38:36,781 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop
2021-07-09 06:38:36,873 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTr
2021-07-09 06:38:37,008 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-09 06:38:37,032 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTr
2021-07-09 06:38:37,069 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTr
2021-07-09 06:38:37,077 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-09 06:38:37,177 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTr
2021-07-09 06:38:37,657 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625413465806_0003
2021-07-09 06:38:37,658 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-09 06:38:37,833 INFO conf.Configuration: resource-types.xml not found
2021-07-09 06:38:37,834 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-09 06:38:37,935 INFO impl.YarnClientImpl: Submitted application application_1625413465806_0003
2021-07-09 06:38:38,102 INFO mapreduce.Job: The url to track the job: http://aditya-n-bhatt-1nt18is015:8088/proxy/applicatior
2021-07-09 06:38:38,103 INFO mapreduce.Job: Running job: job_1625413465806_0003
2021-07-09 06:38:43,215 INFO mapreduce.Job: Job job_1625413465806_0003 running in uber mode : false
2021-07-09 06:38:43,218 INFO mapreduce.Job:  map 0% reduce 0%
2021-07-09 06:38:48,291 INFO mapreduce.Job:  map 100% reduce 0%
2021-07-09 06:38:52,327 INFO mapreduce.Job:  map 100% reduce 100%
2021-07-09 06:38:53,361 INFO mapreduce.Job: Job job_1625413465806_0003 completed successfully
2021-07-09 06:38:53,449 INFO mapreduce.Job: Counters: 54
        File System Counters
                FILE: Number of bytes read=114
                FILE: Number of bytes written=677780
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=1141
                HDFS: Number of bytes written=50
                HDFS: Number of read operations=11
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
                HDFS: Number of bytes read erasure-coded=0
```

```
            HDFS: Number of bytes read erasure-coded=0
  Job Counters
            Launched map tasks=2
            Launched reduce tasks=1
            Data-local map tasks=2
            Total time spent by all maps in occupied slots (ms)=4934
            Total time spent by all reduces in occupied slots (ms)=1767
            Total time spent by all map tasks (ms)=4934
            Total time spent by all reduce tasks (ms)=1767
            Total vcore-milliseconds taken by all map tasks=4934
            Total vcore-milliseconds taken by all reduce tasks=1767
            Total megabyte-milliseconds taken by all map tasks=5052416
            Total megabyte-milliseconds taken by all reduce tasks=1809408
  Map-Reduce Framework
            Map input records=20
            Map output records=4
            Map output bytes=208
            Map output materialized bytes=120
            Input split bytes=184
            Combine input records=4
            Combine output records=2
            Reduce input groups=1
            Reduce shuffle bytes=120
            Reduce input records=2
            Reduce output records=1
            Spilled Records=4
            Shuffled Maps =2
            Failed Shuffles=0
            Merged Map outputs=2
            GC time elapsed (ms)=171
            CPU time spent (ms)=1660
            Physical memory (bytes) snapshot=729804800
            Virtual memory (bytes) snapshot=7796326400
            Total committed heap usage (bytes)=606601216
            Peak Map Physical memory (bytes)=276217856
```

```
Text Editor                                                    Sat 05:48
Open                                                  *LA2-HadoopExercise.txt
                                                                ~/Desktop
                    Shuffled Maps =2
                    Failed Shuffles=0
                    Merged Map outputs=2
                    GC time elapsed (ms)=171
                    CPU time spent (ms)=1660
                    Physical memory (bytes) snapshot=729804800
                    Virtual memory (bytes) snapshot=7796326400
                    Total committed heap usage (bytes)=606601216
                    Peak Map Physical memory (bytes)=276217856
                    Peak Map Virtual memory (bytes)=2597658624
                    Peak Reduce Physical memory (bytes)=181227520
                    Peak Reduce Virtual memory (bytes)=2602868736
        Shuffle Errors
                    BAD_ID=0
                    CONNECTION=0
                    IO_ERROR=0
                    WRONG_LENGTH=0
                    WRONG_MAP=0
                    WRONG_REDUCE=0
        File Input Format Counters
                    Bytes Read=957
        File Output Format Counters
                    Bytes Written=50
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls EmpAddress.txt
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2021-07-09 06:36 EmpAddress.txt/_SUCCESS
-rw-r--r--   1 hdoop supergroup         50 2021-07-09 06:3 EmpAddress.txt/part-00000
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -cat EmpAddress.txt/part-00000
2021-07-04 10:01:18,780 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHos
Total no.of employees who lives in Bangalore :          4
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls
Found 4 items
drwxr-xr-x   - hdoop supergroup          0 2021-07-09 06:36 EmpExp.txt
drwxr-xr-x   - hdoop supergroup          0 2021-07-09 06:39 EmpISE.txt
-rw-r--r--   1 hdoop supergroup        638 2021-07-09 07:03 LA2.csv
drwxr-xr-x   - hdoop supergroup          0 2021-06-21 07:25 aditya
```

HIVE

Hive is a data warehouse infrastructure tool to process structured data in Hadoop. It resides on top of Hadoop to summarize Big Data, and makes querying and analyzing easy. Initially Hive was developed by Facebook, later the Apache Software Foundation took it up and developed it further as an open source under the name Apache Hive. It is used by different companies. For example, Amazon uses it in Amazon Elastic MapReduce.

Hive is not

 A relational database

 A design for OnLine Transaction Processing (OLTP)

 A language for real-time queries and row-level updates

Features of Hive

 It stores schema in a database and processed data into HDFS.

 It is designed for OLAP.

 It provides SQL type language for querying called HiveQL or HQL.

 It is familiar, fast, scalable, and extensible.

Hive Problem Statement

Use the above dataset in .csv file and create a database called as EmployeeDB. Create a table under the database called as Employee using HIVEQL. The table fields are same, that is,

Name	SSN	Salary	Dname	Experience
Harsha	5000	30000	ISE	5

Use the HiveQL language to perform the following Query based Map-reduce operations,

1. Insert 5 records using INSERT command.
2. Demonstrate the Alter command for the following cases,

a. Rename the table name to “Emp”.

b. Rename the column name “Dname” to “Dept_name”.

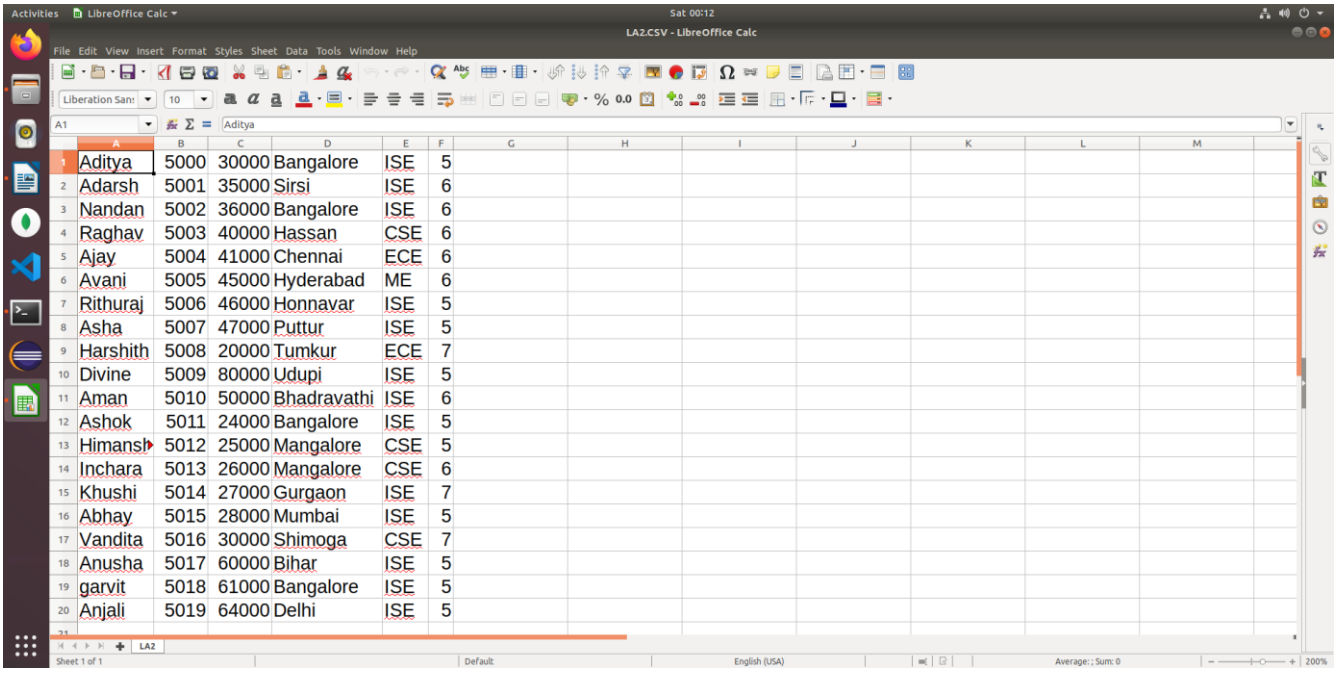
3. Retrieve all the employees who’s salary is not less than 50000.
4. Extract all employees who live in Bangalore but having less than 5 years of experience
5. Create separate view containing Name, Dept_name of employees
6. Display Name and SSN and use group by SSN and order by Name
7. Retrieve Maximum salary, minimum salary and Average salary of the employees

8. Create Another table called Department with the following fields (Dname = Dept_name and perform the following joins (outer, left outer, right outer) over Dname

Dno	Dname
6	ISE

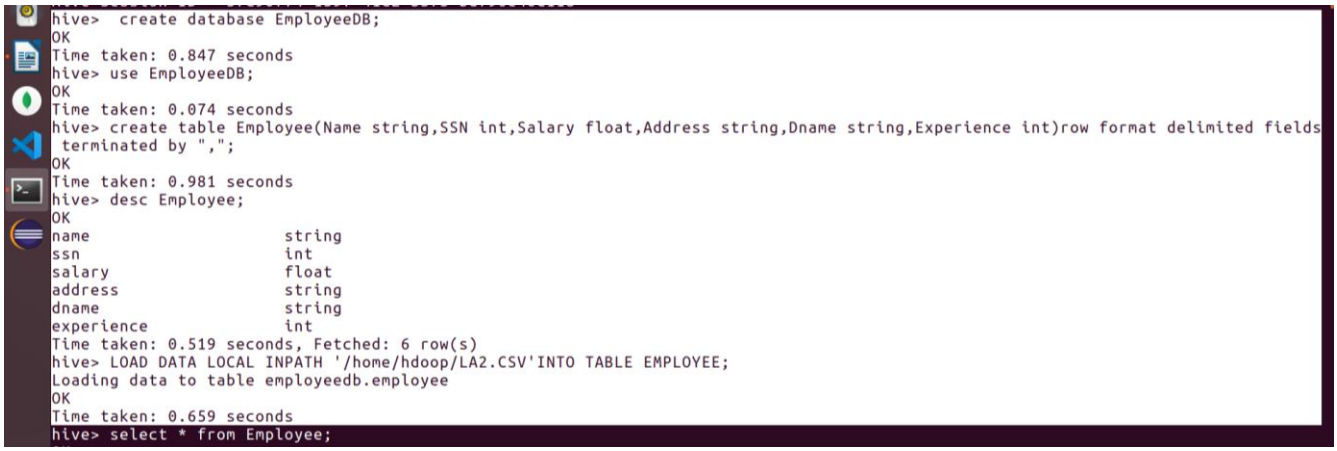
Database Description

LA2.csv



	A	B	C	D	E	F
1	Aditya	5000	30000	Bangalore	ISE	5
2	Adarsh	5001	35000	Sirsi	ISE	6
3	Nandan	5002	36000	Bangalore	ISE	6
4	Raghav	5003	40000	Hassan	CSE	6
5	Ajay	5004	41000	Chennai	ECE	6
6	Avani	5005	45000	Hyderabad	ME	6
7	Rithuraj	5006	46000	Honnavar	ISE	5
8	Asha	5007	47000	Puttur	ISE	5
9	Harshith	5008	20000	Tumkur	ECE	7
10	Divine	5009	80000	Udupi	ISE	5
11	Aman	5010	50000	Bhadravathi	ISE	6
12	Ashok	5011	24000	Bangalore	ISE	5
13	Himansh	5012	25000	Mangalore	CSE	5
14	Inchara	5013	26000	Mangalore	CSE	6
15	Khushi	5014	27000	Gurgaon	ISE	7
16	Abhay	5015	28000	Mumbai	ISE	5
17	Vandita	5016	30000	Shimoga	CSE	7
18	Anusha	5017	60000	Bihar	ISE	5
19	garvit	5018	61000	Bangalore	ISE	5
20	Anjali	5019	64000	Delhi	ISE	5

Results and Snapshots:

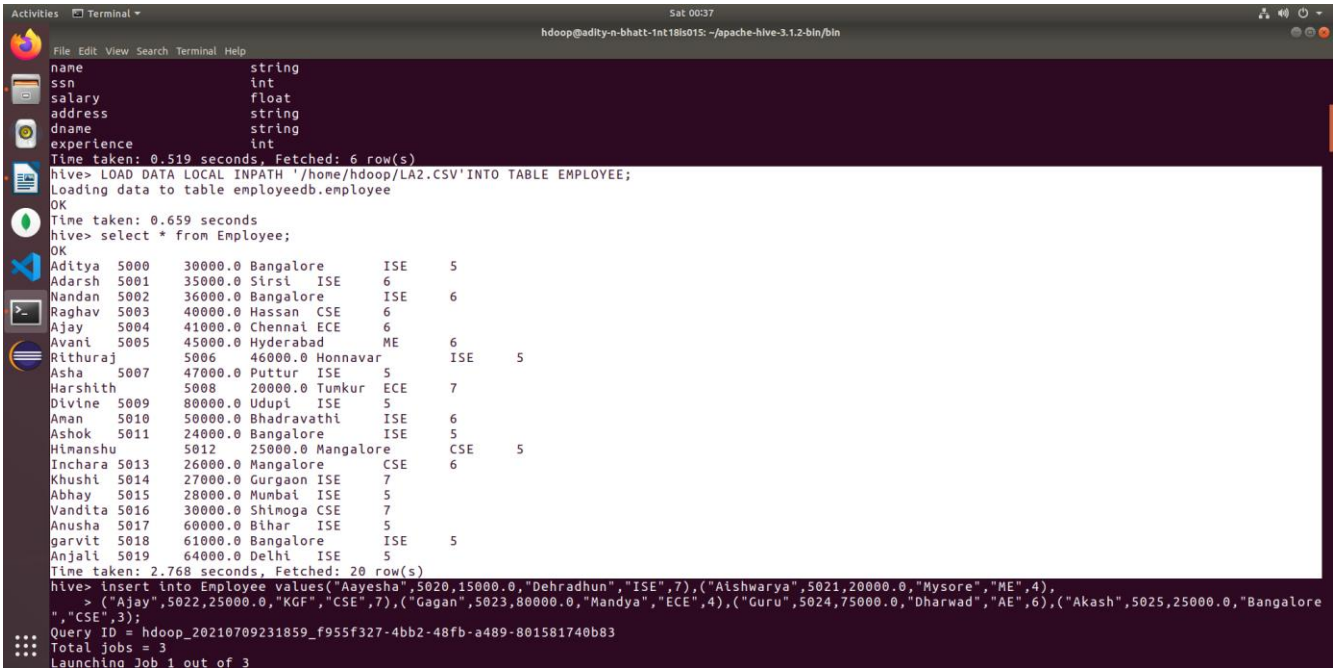


```
hive>  create database EmployeeDB;
OK
Time taken: 0.847 seconds
hive> use EmployeeDB;
OK
Time taken: 0.074 seconds
hive> create table Employee(Name string,SSN int,Salary float,Address string,Dname string,Experience int)row format delimited fields
 terminated by ",";
OK
Time taken: 0.981 seconds
hive> desc Employee;
OK
name                    string
ssn                     int
salary                  float
address                 string
dname                   string
experience              int
Time taken: 0.519 seconds, Fetched: 6 row(s)
hive> LOAD DATA LOCAL INPATH '/home/hdoop/LA2.CSV'INTO TABLE EMPLOYEE;
Loading data to table employeedb.employee
OK
Time taken: 0.659 seconds
hive> select * from Employee;
```

Loading the database:

hive> LOAD DATA LOCAL INPATH '/home/hdoop/LA2.CSV'INTO TABLE EMPLOYEE;

hive>select * from Employee;



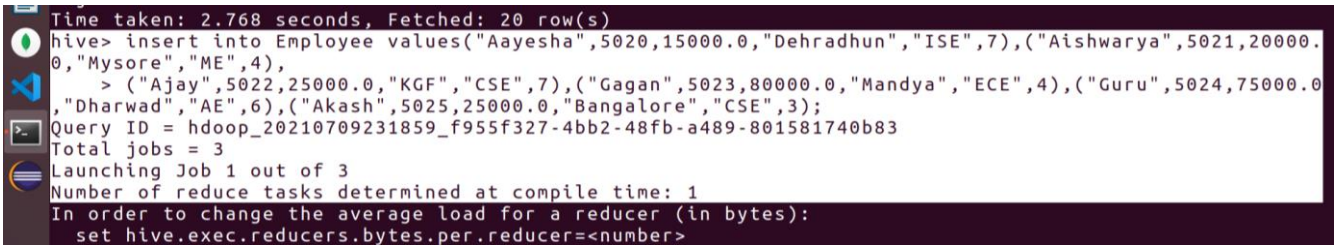
```
Activities  Terminal                    Sat 00:37
hdoop@adity-n-bhatt-1nt18is015: ~/apache-hive-3.1.2-bin/bin
File  Edit  View  Search  Terminal  Help
name                    string
ssn                     int
salary                  float
address                 string
dname                   string
experience              int
Time taken: 0.519 seconds, Fetched: 6 row(s)
hive> LOAD DATA LOCAL INPATH '/home/hdoop/LA2.CSV'INTO TABLE EMPLOYEE;
Loading data to table employeedb.employee
OK
Time taken: 0.659 seconds
hive> select * from Employee;
OK
Aditya  5000    30000.0 Bangalore       ISE     5
Adarsh  5001    35000.0 Sirsi   ISE     6
Nandan  5002    36000.0 Bangalore       ISE     6
Raghav  5003    40000.0 Hassan  CSE     6
Ajay    5004    41000.0 Chennai ECE     6
Avani   5005    45000.0 Hyderabad       ME      6
Rithuraj        5006    46000.0 Honnavar        ISE     5
Asha    5007    47000.0 Puttur  ISE     5
Harshith        5008    20000.0 Tumkur  ECE     7
Divine  5009    80000.0 Udupi   ISE     5
Aman    5010    50000.0 Bhadravathi     ISE     6
Ashok   5011    24000.0 Bangalore       ISE     5
Himanshu        5012    25000.0 Mangalore       CSE     5
Inchara 5013    26000.0 Mangalore       CSE     6
Khushi  5014    27000.0 Gurgaon ISE     7
Abhay   5015    28000.0 Mumbai  ISE     5
Vandita 5016    30000.0 Shimoga CSE     7
Anusha  5017    60000.0 Bihar   ISE     5
garvit  5018    61000.0 Bangalore       ISE     5
Anjali  5019    64000.0 Delhi   ISE     5
Time taken: 2.768 seconds, Fetched: 20 row(s)
hive> insert into Employee values("Aayesha",5020,15000.0,"Dehradhun","ISE",7),("Aishwarya",5021,20000.0,"Mysore","ME",4),
    > ("Ajay",5022,25000.0,"KGF","CSE",7),("Gagan",5023,80000.0,"Mandya","ECE",4),("Guru",5024,75000.0,"Dharwad","AE",6),("Akash",5025,25000.0,"Bangalore
","CSE",3);
Query ID = hdoop_20210709231859_f955f327-4bb2-48fb-a489-801581740b83
Total jobs = 3
Launching Job 1 out of 3
```

Hive Queries:

1. Insert 5 records using INSERT command.

General Format: INSERT INTO tablename VALUES(Col_name Type);

```
hive> insert into Employee
values("Aayesha",5020,15000.0,"Dehradhun","ISE",7),
      ("Aishwarya",5021,20000.0,"Mysore","ME",4),
      ("Ajay",5022,25000.0,"KGF","CSE",7),
      ("Gagan",5023,80000.0,"Mandya","ECE",4),
      ("Guru",5024,75000.0,"Dharwad","AE",6),
      ("Akash",5025,25000.0,"Bangalore","CSE",3);
```



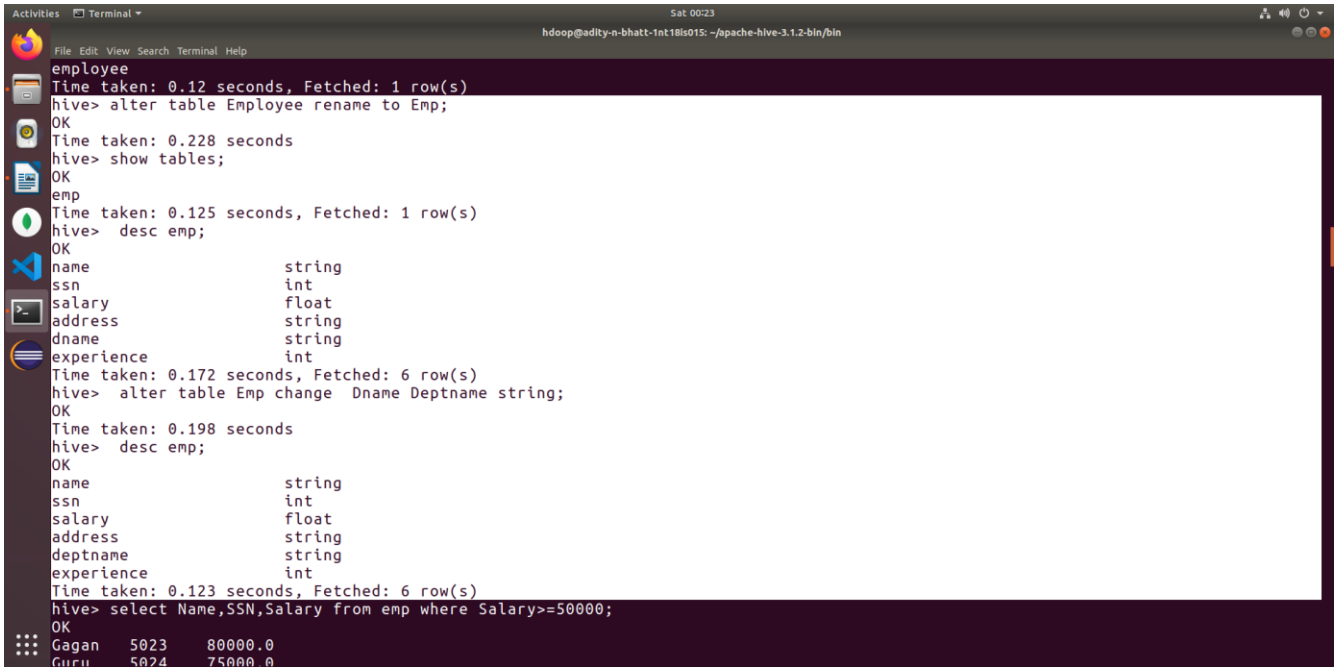
2. Demonstrate the Alter command for the following cases,

a. Rename the table name to “Emp”.

```
hive>alter table Employee rename to Emp;
```

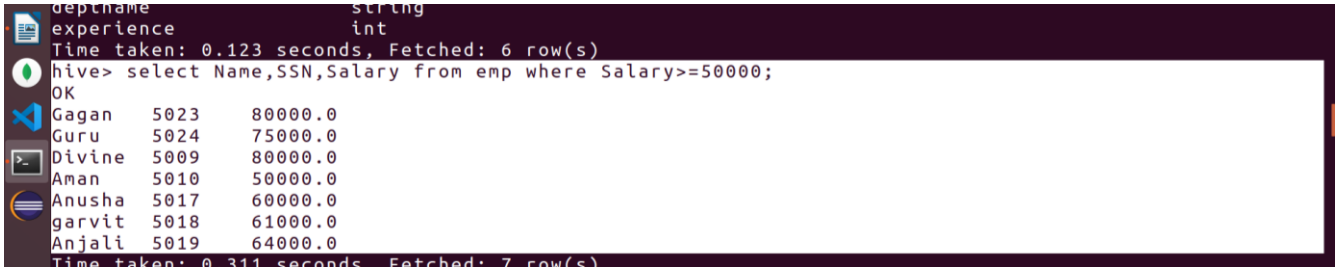
b. Rename the column name “Dname” to “Dept_name”.

```
hive>  alter table Emp change  Dname Deptname string;
```

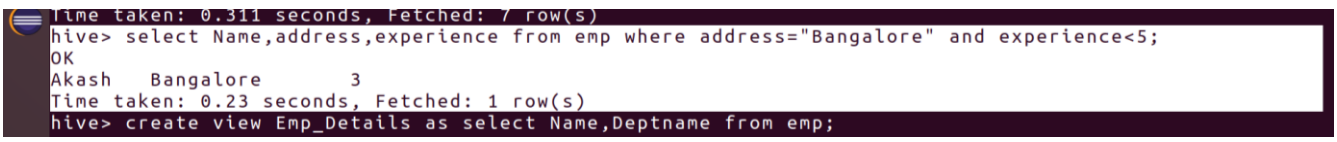


3. Retrieve all the employees who's salary is not less than 50000.

hive> select Name,SSN,Salary from emp where Salary>=50000;



4. Extract all employees who live in Bangalore but having less than 5 years of experience.
hive> select Name,address,experience from emp where address="Bangalore" and experience<5;

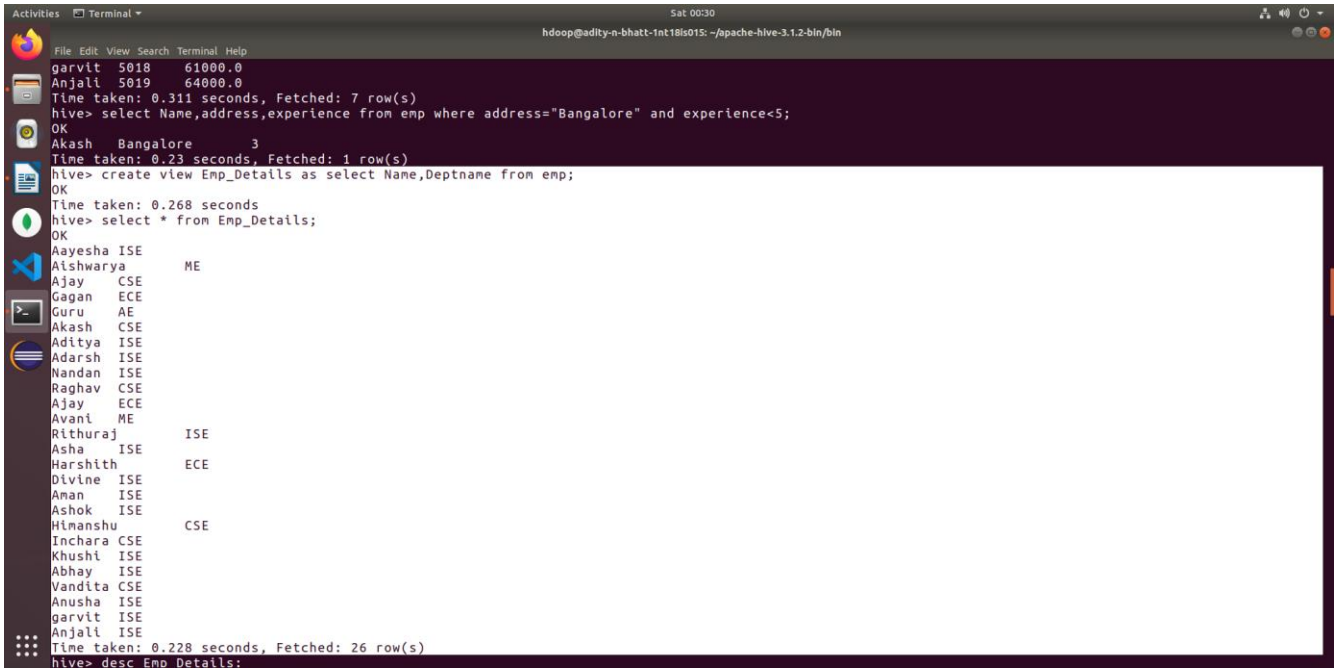


5. Create separate view containing Name, Dept_name of employees

hive> create view Emp_Details as select Name,Deptname from emp;

hive> select * from Emp_Details;

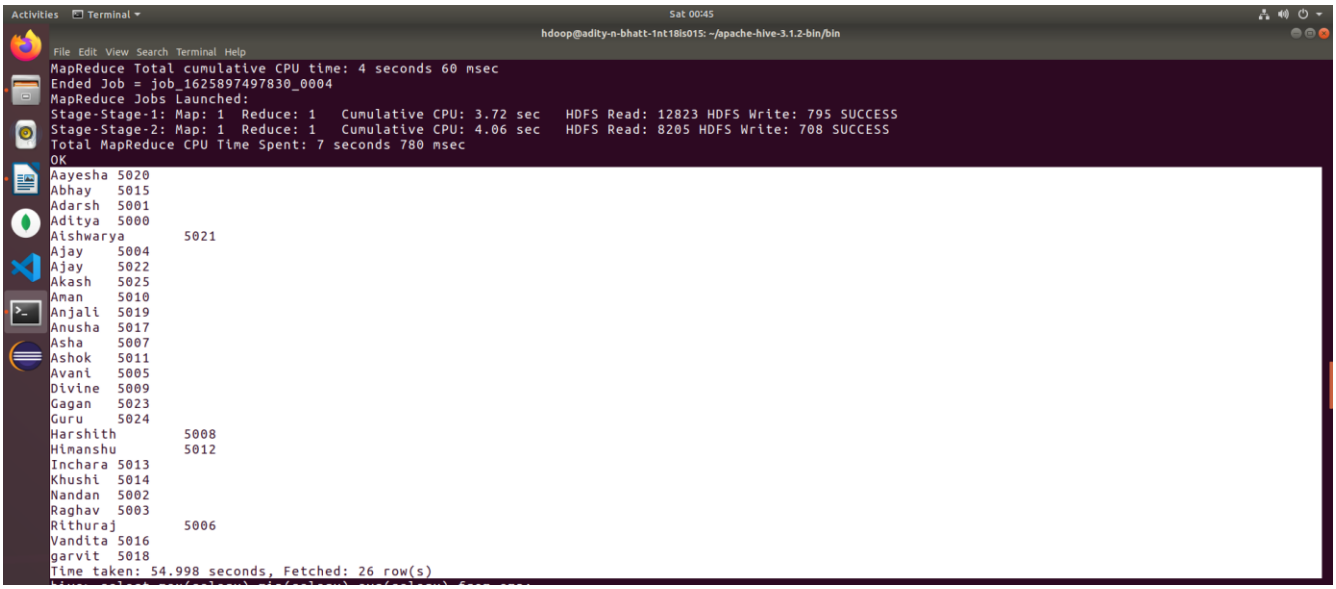
(Apache Hive View is a searchable object in a database which we can define by the query. However, we cannot store data in the view. Still, some refer to as a view as “virtual tables”)



6. Display Name and SSN and use group by SSN and order by Name.

hive> select name,ssn from emp group by name,ssn order by name;

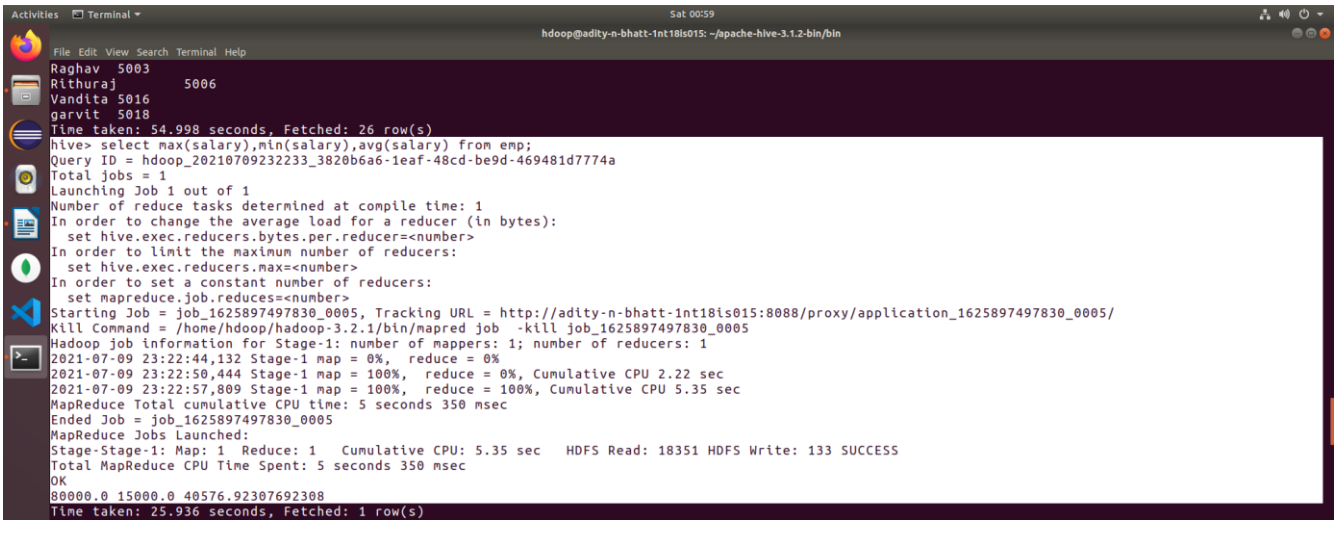
The GROUP BY clause is used to group all the records in a result set using a particular collection column. It is used to query a group of records



7. Retrieve Maximum salary, minimum salary and Average salary of the employees.

hive> select max(salary),min(salary),avg(salary) from emp;

80000.0(max)
15000.0(min)
40576.92307692308(average)



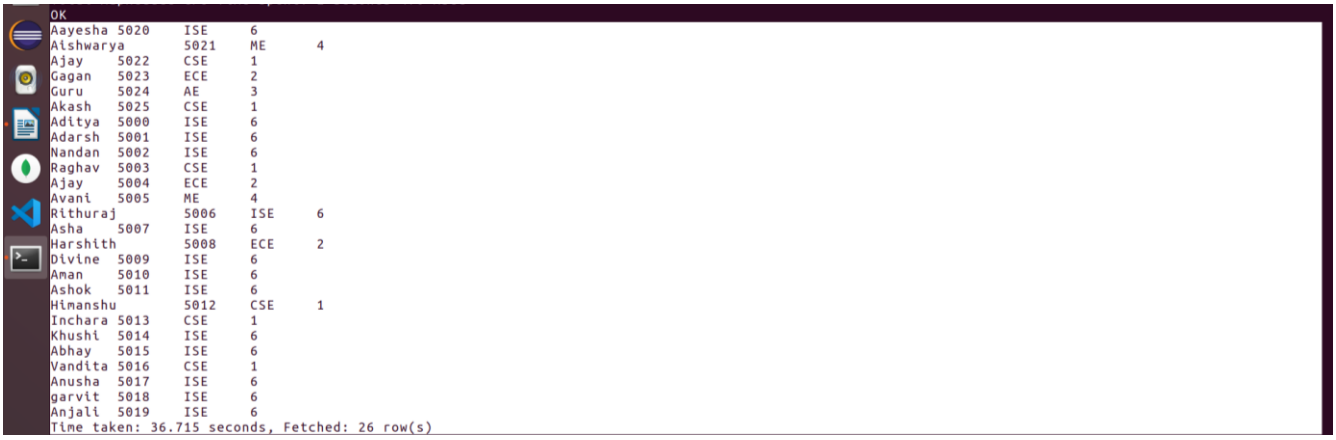
8. Create Another table called Department with the following fields (Dname = Dept_name and perform the following joins (outer, left outer, right outer) over Dname

Dno	Dname
6	ISE

JOIN is a clause that is used for combining specific fields from two tables by using values common to each one. It is used to combine records from two or more tables in the database.

hive> select name,ssn,deptname,dno from emp e full outer join department d on e.deptname=d.dname;

hive> select name,ssn,deptname,dno from emp e left outer join department d on e.deptname=d.dname;



hive> select name,ssn,deptname,dno from emp e right outer join department d on e.deptname=d.dname;

References

Hadoop & Mapreduce :

https://www.youtube.com/watch?v=U3fkWvaqgl8
https://www.youtube.com/watch?v=K0aDh_sfVrc

Hive :

https://www.youtube.com/watch?v=SAX8b3AN3Uc

Hadoop Terminal execution:

```
hdoop@aditya-n-bhatt-1nt18is015:~$ cd hadoop-3.2.1/sbin
hdoop@aditya-n-bhatt-1nt18is015:~/hadoop-3.2.1/sbin$ ./start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [aditya-n-bhatt-1nt18is015]
Starting resourcemanager
Starting nodemanagers
hdoop@aditya-n-bhatt-1nt18is015:~/hadoop-3.2.1/sbin$ jps
5202 SecondaryNameNode
6612 Jps
5700 NodeManager
5525 ResourceManager
4982 DataNode
3386 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
4831 NameNode
hdoop@aditya-n-bhatt-1nt18is015:~/hadoop-3.2.1/sbin$ cd Desktop
bash: cd: Desktop: No such file or directory
hdoop@aditya-n-bhatt-1nt18is015:~/hadoop-3.2.1/sbin$ cd ~
hdoop@aditya-n-bhatt-1nt18is015:~$ cd Desktop
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ ls
  LA2.csv       EmpISE.jar
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -copyFromLocal LA2.csv
2021-07-04 09:03:22,579 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls
Found 3 items
drwxr-xr-x   - hdoop supergroup          0 2021-06-27 07:34 Adityanbhatt
-rw-r--r--   1 hdoop supergroup        638 2021-07-09 07:03 LA2.csv
drwxr-xr-x   - hdoop supergroup          0 2021-06-21 07:25 aditya
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop jar EmpISE.jar EmpISE.EmpISE
LA2.csv EmpISE.txt
2021-07-04 09:11:06,280 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
2021-07-04 09:11:18,054 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
```

2021-07-04 09:11:20,020 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
2021-07-04 09:11:20,616 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1625413465806_0001
2021-07-04 09:11:21,316 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-04 09:11:22,438 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-04 09:11:23,361 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-04 09:11:23,545 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-04 09:11:23,566 INFO mapjreduce.JobSubmitter: number of splits:2
2021-07-04 09:11:26,831 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2021-07-04 09:11:28,217 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625413465806_0001
2021-07-04 09:11:28,218 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-04 09:11:32,114 INFO conf.Configuration: resource-types.xml not found
2021-07-04 09:11:32,114 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-04 09:11:38,975 INFO impl.YarnClientImpl: Submitted application application_1625413465806_0001
2021-07-04 09:11:40,184 INFO mapreduce.Job: The url to track the job: http://aditya-n-bhatt-1nt18is015:8088/proxy/application_1625413465806_0001/
2021-07-04 09:11:40,361 INFO mapreduce.Job: Running job: job_1625413465806_0001
2021-07-04 09:13:46,673 INFO mapreduce.Job: Job job_1625413465806_0001 running in uber mode : false
2021-07-04 09:13:46,674 INFO mapreduce.Job: map 0% reduce 0%
2021-07-04 09:18:31,564 INFO mapreduce.Job: map 33% reduce 0%
2021-07-04 09:18:32,587 INFO mapreduce.Job: map 67% reduce 0%
2021-07-04 09:18:34,657 INFO mapreduce.Job: map 100% reduce 0%
2021-07-04 09:19:23,241 INFO mapreduce.Job: map 100% reduce 100%
2021-07-04 09:19:25,278 INFO mapreduce.Job: Job job_1625413465806_0001 completed successfully
2021-07-04 09:19:26,524 INFO mapreduce.Job: Counters: 54
	File System Counters
		FILE: Number of bytes read=120
		FILE: Number of bytes written=677717
		FILE: Number of read operations=0
		FILE: Number of large read operations=0

FILE: Number of write operations=0
HDFS: Number of bytes read=1141
HDFS: Number of bytes written=54
HDFS: Number of read operations=11
HDFS: Number of large read operations=0
HDFS: Number of write operations=2
HDFS: Number of bytes read erasure-coded=0
Job Counters
Launched map tasks=2
Launched reduce tasks=1
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=568879
Total time spent by all reduces in occupied slots (ms)=45707
Total time spent by all map tasks (ms)=568879
Total time spent by all reduce tasks (ms)=45707
Total vcore-milliseconds taken by all map tasks=568879
Total vcore-milliseconds taken by all reduce tasks=45707
Total megabyte-milliseconds taken by all map tasks=582532096
Total megabyte-milliseconds taken by all reduce tasks=46803968
Map-Reduce Framework
Map input records=20
Map output records=13
Map output bytes=715
Map output materialized bytes=126
Input split bytes=184
Combine input records=13
Combine output records=2
Reduce input groups=1
Reduce shuffle bytes=126
Reduce input records=2
Reduce output records=1
Spilled Records=4
Shuffled Maps =2
Failed Shuffles=0
Merged Map outputs=2
GC time elapsed (ms)=28715
CPU time spent (ms)=360840
Physical memory (bytes) snapshot=790487040
Virtual memory (bytes) snapshot=7789518848
Total committed heap usage (bytes)=644874240
Peak Map Physical memory (bytes)=304123904
Peak Map Virtual memory (bytes)=2595966976

```
		Peak Reduce Physical memory (bytes)=183201792
		Peak Reduce Virtual memory (bytes)=2604199936
	Shuffle Errors
		BAD_ID=0
		CONNECTION=0
		IO_ERROR=0
		WRONG_LENGTH=0
		WRONG_MAP=0
		WRONG_REDUCE=0
	File Input Format Counters
		Bytes Read=957
	File Output Format Counters
		Bytes Written=54
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls EmpISE.txt
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2021-07-09 06:36 EmpISE.txt/_SUCCESS
-rw-r--r--   1 hdoop supergroup         54 2021-07-09 09:36 EmpISE.txt/part-00000
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -cat EmpISE.txt/part-00000
2021-07-04 09:20:31,372 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
Total no.of employees working in ISE Department :	13
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop jar EmpExp.jar
EmpExp.EmpExp LA2.csv EmpExp.txt
2021-07-09 06:34:04,465 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
2021-07-09 06:34:09,024 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
2021-07-09 06:34:10,405 WARN mapreduce.JobResourceUploader: Hadoop command-
line option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
2021-07-09 06:34:10,709 INFO mapreduce.JobResourceUploader: Disabling Erasure
Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1625413465806_0002
2021-07-09 06:34:11,286 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 06:34:12,643 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-09 06:34:12,836 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 06:34:12,862 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 06:34:12,989 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-09 06:34:13,318 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
```

2021-07-09 06:34:13,366 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1625413465806_0002
2021-07-09 06:34:13,366 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-09 06:34:13,903 INFO conf.Configuration: resource-types.xml not found
2021-07-09 06:34:13,904 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2021-07-09 06:34:16,566 INFO impl.YarnClientImpl: Submitted application application_1625413465806_0002
2021-07-09 06:34:18,577 INFO mapreduce.Job: The url to track the job: http://aditya-n-bhatt-1nt18is015:8088/proxy/application_1625413465806_0002/
2021-07-09 06:34:18,712 INFO mapreduce.Job: Running job: job_1625413465806_0002
2021-07-09 06:35:23,814 INFO mapreduce.Job: Job job_1625413465806_0002 running in uber mode : false
2021-07-09 06:35:23,818 INFO mapreduce.Job: map 0% reduce 0%
2021-07-09 06:36:21,633 INFO mapreduce.Job: map 100% reduce 0%
2021-07-09 06:36:26,679 INFO mapreduce.Job: map 100% reduce 100%
2021-07-09 06:36:27,708 INFO mapreduce.Job: Job job_1625413465806_0002 completed successfully
2021-07-09 06:36:27,841 INFO mapreduce.Job: Counters: 55
	File System Counters
		FILE: Number of bytes read=126
		FILE: Number of bytes written=677738
		FILE: Number of read operations=0
		FILE: Number of large read operations=0
		FILE: Number of write operations=0
		HDFS: Number of bytes read=1141
		HDFS: Number of bytes written=57
		HDFS: Number of read operations=11
		HDFS: Number of large read operations=0
		HDFS: Number of write operations=2
		HDFS: Number of bytes read erasure-coded=0
	Job Counters
		Killed map tasks=1
		Launched map tasks=2
		Launched reduce tasks=1
		Data-local map tasks=2
		Total time spent by all maps in occupied slots (ms)=107735
		Total time spent by all reduces in occupied slots (ms)=2063
		Total time spent by all map tasks (ms)=107735
		Total time spent by all reduce tasks (ms)=2063
		Total vcore-milliseconds taken by all map tasks=107735

```
		Total vcore-milliseconds taken by all reduce tasks=2063
		Total megabyte-milliseconds taken by all map tasks=110320640
		Total megabyte-milliseconds taken by all reduce tasks=2112512
	Map-Reduce Framework
		Map input records=20
		Map output records=10
		Map output bytes=580
		Map output materialized bytes=132
		Input split bytes=184
		Combine input records=10
		Combine output records=2
		Reduce input groups=1
		Reduce shuffle bytes=132
		Reduce input records=2
		Reduce output records=1
		Spilled Records=4
		Shuffled Maps =2
		Failed Shuffles=0
		Merged Map outputs=2
		GC time elapsed (ms)=4467
		CPU time spent (ms)=45330
		Physical memory (bytes) snapshot=812605440
		Virtual memory (bytes) snapshot=7793565696
		Total committed heap usage (bytes)=626524160
		Peak Map Physical memory (bytes)=314843136
		Peak Map Virtual memory (bytes)=2597437440
		Peak Reduce Physical memory (bytes)=182988800
		Peak Reduce Virtual memory (bytes)=2600173568
	Shuffle Errors
		BAD_ID=0
		CONNECTION=0
		IO_ERROR=0
		WRONG_LENGTH=0
		WRONG_MAP=0
		WRONG_REDUCE=0
	File Input Format Counters
		Bytes Read=957
	File Output Format Counters
		Bytes Written=57
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls EmpExp.txt
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2021-07-09 06:36 EmpExp.txt/_SUCCESS
```

```
-rw-r--r--   1 hdoop supergroup         57 2021-07-09 06:36 EmpExp.txt/part-00000
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -cat EmpExp.txt/part-00000
2021-07-09 06:36:55,206 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
Total no.of employees having 5 years of experience :        10
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop jar EmpAddress.jar
EmpAddress.EmpAddress LA2.csv EmpAddress.txt
2021-07-09 06:38:36,404 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
2021-07-09 06:38:36,566 INFO client.RMProxy: Connecting to ResourceManager at
/127.0.0.1:8032
2021-07-09 06:38:36,727 WARN mapreduce.JobResourceUploader: Hadoop command-
line option parsing not performed. Implement the Tool interface and execute your
application with ToolRunner to remedy this.
2021-07-09 06:38:36,781 INFO mapreduce.JobResourceUploader: Disabling Erasure
Coding for path: /tmp/hadoop-yarn/staging/hdoop/.staging/job_1625413465806_0003
2021-07-09 06:38:36,873 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 06:38:37,008 INFO mapred.FileInputFormat: Total input files to process : 1
2021-07-09 06:38:37,032 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 06:38:37,069 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 06:38:37,077 INFO mapreduce.JobSubmitter: number of splits:2
2021-07-09 06:38:37,177 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
2021-07-09 06:38:37,657 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1625413465806_0003
2021-07-09 06:38:37,658 INFO mapreduce.JobSubmitter: Executing with tokens: []
2021-07-09 06:38:37,833 INFO conf.Configuration: resource-types.xml not found
2021-07-09 06:38:37,834 INFO resource.ResourceUtils: Unable to find 'resource-
types.xml'.
2021-07-09 06:38:37,935 INFO impl.YarnClientImpl: Submitted application
application_1625413465806_0003
2021-07-09 06:38:38,102 INFO mapreduce.Job: The url to track the job: http://aditya-n-
bhatt-1nt18is015:8088/proxy/application_1625413465806_0003/
2021-07-09 06:38:38,103 INFO mapreduce.Job: Running job:
job_1625413465806_0003
2021-07-09 06:38:43,215 INFO mapreduce.Job: Job job_1625413465806_0003 running
in uber mode : false
2021-07-09 06:38:43,218 INFO mapreduce.Job:  map 0% reduce 0%
2021-07-09 06:38:48,291 INFO mapreduce.Job:  map 100% reduce 0%
```

2021-07-09 06:38:52,327 INFO mapreduce.Job: map 100% reduce 100%
2021-07-09 06:38:53,361 INFO mapreduce.Job: Job job_1625413465806_0003 completed successfully
2021-07-09 06:38:53,449 INFO mapreduce.Job: Counters: 54
	File System Counters
		FILE: Number of bytes read=114
		FILE: Number of bytes written=677780
		FILE: Number of read operations=0
		FILE: Number of large read operations=0
		FILE: Number of write operations=0
		HDFS: Number of bytes read=1141
		HDFS: Number of bytes written=50
		HDFS: Number of read operations=11
		HDFS: Number of large read operations=0
		HDFS: Number of write operations=2
		HDFS: Number of bytes read erasure-coded=0
	Job Counters
		Launched map tasks=2
		Launched reduce tasks=1
		Data-local map tasks=2
		Total time spent by all maps in occupied slots (ms)=4934
		Total time spent by all reduces in occupied slots (ms)=1767
		Total time spent by all map tasks (ms)=4934
		Total time spent by all reduce tasks (ms)=1767
		Total vcore-milliseconds taken by all map tasks=4934
		Total vcore-milliseconds taken by all reduce tasks=1767
		Total megabyte-milliseconds taken by all map tasks=5052416
		Total megabyte-milliseconds taken by all reduce tasks=1809408
	Map-Reduce Framework
		Map input records=20
		Map output records=4
		Map output bytes=208
		Map output materialized bytes=120
		Input split bytes=184
		Combine input records=4
		Combine output records=2
		Reduce input groups=1
		Reduce shuffle bytes=120
		Reduce input records=2
		Reduce output records=1
		Spilled Records=4
		Shuffled Maps =2

```
		Failed Shuffles=0
		Merged Map outputs=2
		GC time elapsed (ms)=171
		CPU time spent (ms)=1660
		Physical memory (bytes) snapshot=729804800
		Virtual memory (bytes) snapshot=7796326400
		Total committed heap usage (bytes)=606601216
		Peak Map Physical memory (bytes)=276217856
		Peak Map Virtual memory (bytes)=2597658624
		Peak Reduce Physical memory (bytes)=181227520
		Peak Reduce Virtual memory (bytes)=2602868736
	Shuffle Errors
		BAD_ID=0
		CONNECTION=0
		IO_ERROR=0
		WRONG_LENGTH=0
		WRONG_MAP=0
		WRONG_REDUCE=0
	File Input Format Counters
		Bytes Read=957
	File Output Format Counters
		Bytes Written=50
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls EmpAddress.txt
Found 2 items
-rw-r--r--   1 hdoop supergroup          0 2021-07-09 06:36 EmpAddress.txt/_SUCCESS
-rw-r--r--   1 hdoop supergroup         50 2021-07-09 06:3 EmpAddress.txt/part-00000
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -cat EmpAddress.txt/part-
00000
2021-07-04 10:01:18,780 INFO sasl.SaslDataTransferClient: SASL encryption trust
check: localHostTrusted = false, remoteHostTrusted = false
Total no.of employees who lives in Bangalore :      4
hdoop@aditya-n-bhatt-1nt18is015:~/Desktop$ hadoop fs -ls
Found 4 items
drwxr-xr-x   - hdoop supergroup          0 2021-07-09 06:36 EmpExp.txt
drwxr-xr-x   - hdoop supergroup          0 2021-07-09 06:39 EmpISE.txt
-rw-r--r--   1 hdoop supergroup        638 2021-07-09 07:03 LA2.csv
drwxr-xr-x   - hdoop supergroup          0 2021-06-21 07:25 aditya
```

Hive terminal execution:

```
hdoop@adity-n-bhatt-1nt18is015:~/apache-hive-3.1.2-bin/bin$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hdoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
localhost: namenode is running as process 2149.  Stop it first.
Starting datanodes
localhost: datanode is running as process 2304.  Stop it first.
Starting secondary namenodes [adity-n-bhatt-1nt18is015]
adity-n-bhatt-1nt18is015: secondarynamenode is running as process 2543.  Stop it first.
2021-07-09 23:17:15,566 WARN util.NativeCodeLoader: Unable to load native-hadoop
library for your platform... using builtin-java classes where applicable
Starting resourcemanager
resourcemanager is running as process 2873.  Stop it first.
Starting nodemanagers
localhost: nodemanager is running as process 3036.  Stop it first.

hdoop@adity-n-bhatt-1nt18is015:~/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 = 276ef455-36f8-41dd-a7bf-198818efb270

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 = 37c90f74-139f-41c2-b375-bdf96e46c118
```

2)creating,selecting Database

```
hive>  create database EmployeeDB;
```

```
OK
Time taken: 0.847 seconds

hive> use EmployeeDB;
OK
Time taken: 0.074 seconds

hive> create table Employee(Name string,SSN int,Salary float,Address string,Dname
string,Experience int)row format delimited fields terminated by ",";
OK
Time taken: 0.981 seconds

hive> desc Employee;
OK
name                    string
ssn                     int
salary                  float
address                 string
dname                   string
experience              int
Time taken: 0.519 seconds, Fetched: 6 row(s)

hive> LOAD DATA LOCAL INPATH '/home/hdoop/LA2.CSV'INTO TABLE
EMPLOYEE;
Loading data to table employeedb.employee
OK
Time taken: 0.659 seconds

hive> select * from Employee;
OK
Aditya      5000  30000.0     Bangalore   ISE   5
Adarsh      5001  35000.0     Sirsi  ISE  6
Nandan      5002  36000.0     Bangalore   ISE   6
Raghav      5003  40000.0     Hassan      CSE   6
Ajay  5004  41000.0      Chennai     ECE   6
Avani 5005  45000.0      Hyderabad   ME    6
Rithuraj    5006  46000.0     Honnavar    ISE   5
Asha  5007  47000.0      Puttur ISE   5
Harshith    5008  20000.0     Tumkur      ECE   7
Divine      5009  80000.0     Udupi ISE   5
Aman 5010  50000.0      Bhadravathi ISE   6
Ashok       5011  24000.0     Bangalore   ISE   5
```

```
Himanshu    5012  25000.0      Mangalore   CSE   5
Inchara     5013  26000.0      Mangalore   CSE   6
Khushi      5014  27000.0      Gurgaon     ISE   7
Abhay       5015  28000.0      Mumbai      ISE   5
Vandita     5016  30000.0      Shimoga     CSE   7
Anusha      5017  60000.0      Bihar ISE   5
garvit 5018  61000.0      Bangalore   ISE   5
Anjali5019  64000.0      Delhi ISE   5
Time taken: 2.768 seconds, Fetched: 20 row(s)
hive> insert into Employee
values("Aayesha",5020,15000.0,"Dehradhun","ISE",7),("Aishwarya",5021,20000.0,"My
sore","ME",4),
    >
("Ajay",5022,25000.0,"KGF","CSE",7),("Gagan",5023,80000.0,"Mandya","ECE",4),("
Guru",5024,75000.0,"Dharwad","AE",6),("Akash",5025,25000.0,"Bangalore","CSE",3);
Query ID = hdoop_20210709231859_f955f327-4bb2-48fb-a489-801581740b83
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_1625897497830_0002, Tracking URL = http://adity-n-bhatt-
1nt18is015:8088/proxy/application_1625897497830_0002/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill
job_1625897497830_0002
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-09 23:19:11,954 Stage-1 map = 0%,  reduce = 0%
2021-07-09 23:19:21,811 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 4.44 sec
2021-07-09 23:19:31,333 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 6.42
sec
MapReduce Total cumulative CPU time: 6 seconds 420 msec
Ended Job = job_1625897497830_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/employeedb.db/employee/.hive-
staging_hive_2021-07-09_23-18-59_179_1635479131614468431-1/-ext-10000
```

```
Loading data to table employeedb.employee
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 6.42 sec   HDFS Read: 22857
HDFS Write: 774 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 420 msec
OK
Time taken: 35.006 seconds
hive> select * from Employee;
OK
Aayesha     5020  15000.0     Dehradhun  ISE   7
Aishwarya   5021  20000.0     Mysore     ME    4
Ajay  5022  25000.0     KGF   CSE   7
Gagan       5023  80000.0     Mandya     ECE   4
Guru  5024  75000.0     Dharwad    AE    6
Akash5025  25000.0     Bangalore  CSE   3
Aditya      5000  30000.0     Bangalore  ISE   5
Adarsh      5001  35000.0     Sirsi ISE    6
Nandan      5002  36000.0     Bangalore  ISE   6
Raghav      5003  40000.0     Hassan     CSE   6
Ajay  5004  41000.0     Chennai    ECE   6
Avani 5005  45000.0     Hyderabad  ME    6
Rithuraj    5006  46000.0     Honnavar   ISE   5
Asha  5007  47000.0     Puttur ISE    5
Harshith    5008  20000.0     Tumkur     ECE   7
Divine      5009  80000.0     UdupiISE    5
Aman5010  50000.0     Bhadravathi ISE    6
Ashok       5011  24000.0     Bangalore  ISE   5
Himanshu    5012  25000.0     Mangalore  CSE   5
Inchara     5013  26000.0     Mangalore  CSE   6
Khushi      5014  27000.0     Gurgaon    ISE   7
Abhay       5015  28000.0     Mumbai     ISE   5
Vandita     5016  30000.0     Shimoga    CSE   7
Anusha      5017  60000.0     Bihar ISE    5
garvit 5018  61000.0     Bangalore  ISE   5
Anjali5019  64000.0     Delhi ISE    5
Time taken: 0.187 seconds, Fetched: 26 row(s)
hive>  show tables;
OK
employee
Time taken: 0.12 seconds, Fetched: 1 row(s)
hive> alter table Employee rename to Emp;
OK
```

```
Time taken: 0.228 seconds
hive> show tables;
OK
emp
Time taken: 0.125 seconds, Fetched: 1 row(s)
hive>  desc emp;
OK
name                    string
ssn                     int
salary                  float
address                 string
dname                   string
experience              int
Time taken: 0.172 seconds, Fetched: 6 row(s)
hive>  alter table Emp change  Dname Deptname string;
OK
Time taken: 0.198 seconds
hive>  desc emp;
OK
name                    string
ssn                     int
salary                  float
address                 string
deptname                string
experience              int
Time taken: 0.123 seconds, Fetched: 6 row(s)
hive> select Name,SSN,Salary from emp where Salary>=50000;
OK
Gagan           5023    80000.0
Guru    5024    75000.0
Divine          5009    80000.0
Aman    5010    50000.0
Anusha          5017    60000.0
garvit  5018    61000.0
Anjali  5019    64000.0
Time taken: 0.311 seconds, Fetched: 7 row(s)
hive> select Name,address,experience from emp where address="Bangalore" and
experience<5;
OK
AkashBangalore    3
Time taken: 0.23 seconds, Fetched: 1 row(s)
hive> create view Emp_Details as select Name,Deptname from emp;
```

```
OK
Time taken: 0.268 seconds
hive> select * from Emp_Details;
OK
Aayesha	ISE
Aishwarya	ME
Ajay	CSE
Gagan	ECE
Guru	AE
AkashCSE
Aditya	ISE
Adarsh	ISE
Nandan	ISE
Raghav	CSE
Ajay	ECE
Avani	ME
Rithuraj	ISE
Asha	ISE
Harshith	ECE
Divine	ISE
Aman	ISE
Ashok	ISE
Himanshu	CSE
Inchara	CSE
Khushi	ISE
Abhay	ISE
Vandita	CSE
Anusha	ISE
garvit	ISE
Anjali	ISE
Time taken: 0.228 seconds, Fetched: 26 row(s)
hive> desc Emp_Details;
OK
name                    string
deptname                string
Time taken: 0.124 seconds, Fetched: 2 row(s)
hive> select name,ssn from emp group by name,ssn order by name;
Query ID = hdoop_20210709232132_f4f3ceb4-493d-4835-bcbb-05e32e8de6e4
Total jobs = 2
Launching Job 1 out of 2
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_1625897497830_0003, Tracking URL = http://adity-n-bhatt-
1nt18is015:8088/proxy/application_1625897497830_0003/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill
job_1625897497830_0003
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-09 23:21:43,008 Stage-1 map = 0%,  reduce = 0%
2021-07-09 23:21:49,313 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 1.75 sec
2021-07-09 23:21:56,780 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 3.72
sec
MapReduce Total cumulative CPU time: 3 seconds 720 msec
Ended Job = job_1625897497830_0003
Launching Job 2 out of 2
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_1625897497830_0004, Tracking URL = http://adity-n-bhatt-
1nt18is015:8088/proxy/application_1625897497830_0004/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill
job_1625897497830_0004
Hadoop job information for Stage-2: number of mappers: 1; number of reducers: 1
2021-07-09 23:22:11,798 Stage-2 map = 0%,  reduce = 0%
2021-07-09 23:22:18,156 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 1.79 sec
2021-07-09 23:22:26,575 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 4.06
sec
MapReduce Total cumulative CPU time: 4 seconds 60 msec
Ended Job = job_1625897497830_0004
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 3.72 sec   HDFS Read: 12823
HDFS Write: 795 SUCCESS
Stage-Stage-2: Map: 1  Reduce: 1   Cumulative CPU: 4.06 sec   HDFS Read: 8205
HDFS Write: 708 SUCCESS
Total MapReduce CPU Time Spent: 7 seconds 780 msec
OK
```

Aayesha 5020
Abhay 5015
Adarsh 5001
Aditya 5000
Aishwarya 5021
Ajay 5004
Ajay 5022
Akash5025
Aman 5010
Anjali5019
Anusha 5017
Asha 5007
Ashok 5011
Avani 5005
Divine 5009
Gagan 5023
Guru 5024
Harshith 5008
Himanshu 5012
Inchara 5013
Khushi 5014
Nandan 5002
Raghav 5003
Rithuraj 5006
Vandita 5016
garvit 5018
Time taken: 54.998 seconds, Fetched: 26 row(s)
hive> select max(salary),min(salary),avg(salary) from emp;
Query ID = hdoop_20210709232233_3820b6a6-1eaf-48cd-be9d-469481d7774a
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_1625897497830_0005, Tracking URL = http://adity-n-bhatt-1nt18is015:8088/proxy/application_1625897497830_0005/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625897497830_0005

```
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-09 23:22:44,132 Stage-1 map = 0%,  reduce = 0%
2021-07-09 23:22:50,444 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 2.22 sec
2021-07-09 23:22:57,809 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.35
sec
MapReduce Total cumulative CPU time: 5 seconds 350 msec
Ended Job = job_1625897497830_0005
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.35 sec   HDFS Read: 18351
HDFS Write: 133 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 350 msec
OK
80000.0        15000.0        40576.92307692308
Time taken: 25.936 seconds, Fetched: 1 row(s)
hive>  create table department(dno int,dname string)row format delimited fields
terminated by ",";
OK
Time taken: 0.112 seconds
hive> insert into department
values(6,"ISE"),(1,"CSE"),(2,"ECE"),(5,"EEE"),(3,"AE"),(4,"ME");
Query ID = hdoop_20210709232316_0a1c6a63-ed1e-4a53-84b5-547c1ee88e35
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_1625897497830_0006, Tracking URL = http://adity-n-bhatt-
1nt18is015:8088/proxy/application_1625897497830_0006/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill
job_1625897497830_0006
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2021-07-09 23:23:28,709 Stage-1 map = 0%,  reduce = 0%
2021-07-09 23:23:36,139 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 3.69 sec
2021-07-09 23:23:44,584 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 5.85
sec
MapReduce Total cumulative CPU time: 5 seconds 850 msec
Ended Job = job_1625897497830_0006
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/employeedb.db/department/.hive-
staging_hive_2021-07-09_23-23-17_032_2257761327028071332-1/-ext-10000
Loading data to table employeedb.department
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1  Reduce: 1   Cumulative CPU: 5.85 sec   HDFS Read: 15866
HDFS Write: 342 SUCCESS
Total MapReduce CPU Time Spent: 5 seconds 850 msec
OK
Time taken: 29.136 seconds
hive> select * from department;
OK
6       ISE
1       CSE
2       ECE
5       EEE
3       AE
4       ME
Time taken: 0.175 seconds, Fetched: 6 row(s)
hive> select name,ssn,deptname,dno from emp e full outer join department d on
e.deptname=d.dname;
Query ID = hdoop_20210709232357_c5da02dc-5ffa-4c4d-b788-3e87175fb55d
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_1625897497830_0007, Tracking URL = http://adity-n-bhatt-
1nt18is015:8088/proxy/application_1625897497830_0007/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill
job_1625897497830_0007
Hadoop job information for Stage-1: number of mappers: 2; number of reducers: 1
2021-07-09 23:24:08,907 Stage-1 map = 0%,  reduce = 0%
2021-07-09 23:24:23,353 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 6.59 sec
2021-07-09 23:24:31,730 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 8.42
sec
```

MapReduce Total cumulative CPU time: 8 seconds 420 msec
Ended Job = job_1625897497830_0007
MapReduce Jobs Launched:
Stage-Stage-1: Map: 2 Reduce: 1   Cumulative CPU: 8.42 sec   HDFS Read: 17545 HDFS Write: 884 SUCCESS
Total MapReduce CPU Time Spent: 8 seconds 420 msec
OK
Guru 5024 AE 3
Akash 5025 CSE 1
Inchara 5013 CSE 1
Ajay 5022 CSE 1
Raghav 5003 CSE 1
Himanshu 5012 CSE 1
Vandita 5016 CSE 1
Gagan 5023 ECE 2
Ajay 5004 ECE 2
Harshith 5008 ECE 2
NULL NULL NULL 5
Anjali 5019 ISE 6
garvit 5018 ISE 6
Anusha 5017 ISE 6
Abhay 5015 ISE 6
Khushi 5014 ISE 6
Ashok 5011 ISE 6
Aman 5010 ISE 6
Divine 5009 ISE 6
Asha 5007 ISE 6
Rithuraj 5006 ISE 6
Nandan 5002 ISE 6
Adarsh 5001 ISE 6
Aditya 5000 ISE 6
Aayesha 5020 ISE 6
Aishwarya 5021 ME 4
Avani 5005 ME 4
Time taken: 36.134 seconds, Fetched: 27 row(s)
hive> select name,ssn,deptname,dno from emp e left outer join department d on e.deptname=d.dname;
Query ID = hdoop_20210709232459_124c7b66-59e9-431b-a9b3-62d18ae33113
Total jobs = 1
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]2021-07-09 23:25:15
        Dump the side-table for tag: 1 with group count: 6 into file:

file:/tmp/hive/java/hdoop/276ef455-36f8-41dd-a7bf-198818efb270/hive_2021-07-09_23-24-59_935_6371910031447898687-1/-local-10004/HashTable-Stage-3/MapJoin-mapfile01--.hashtable
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_1625897497830_0008, Tracking URL = http://adity-n-bhatt-1nt18is015:8088/proxy/application_1625897497830_0008/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job -kill job_1625897497830_0008
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-09 23:25:27,084 Stage-3 map = 0%, reduce = 0%
2021-07-09 23:25:34,418 Stage-3 map = 100%, reduce = 0%, Cumulative CPU 2.47 sec
MapReduce Total cumulative CPU time: 2 seconds 470 msec
Ended Job = job_1625897497830_0008
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 2.47 sec   HDFS Read: 9942 HDFS Write: 861 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 470 msec
OK
Aayesha	5020	ISE	6
Aishwarya	5021	ME	4
Ajay	5022	CSE	1
Gagan	5023	ECE	2
Guru	5024	AE	3
Akash	5025	CSE	1
Aditya	5000	ISE	6
Adarsh	5001	ISE	6
Nandan	5002	ISE	6
Raghav	5003	CSE	1
Ajay	5004	ECE	2
Avani	5005	ME	4
Rithuraj	5006	ISE	6
Asha	5007	ISE	6
Harshith	5008	ECE	2
Divine	5009	ISE	6
Aman	5010	ISE	6
Ashok	5011	ISE	6
Himanshu	5012	CSE	1
Inchara	5013	CSE	1
Khushi	5014	ISE	6

```
Abhay       5015  ISE    6
Vandita     5016  CSE    1
Anusha      5017  ISE    6
garvit 5018  ISE    6
Anjali5019  ISE    6
Time taken: 36.715 seconds, Fetched: 26 row(s)
hive> select name,ssn,deptname,dno from emp e right outer join department d on
e.deptname=d.dname;
Query ID = hdoop_20210709232604_bfa87d1f-ad4f-4aed-9c80-00766ba32bc0
Total jobs = 1
SLF4J: Class path contains multiple SLF4J bindings.


2021-07-09 23:26:15      End of local task; Time Taken: 1.648 sec.
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_1625897497830_0009, Tracking URL = http://adity-n-bhatt-
1nt18is015:8088/proxy/application_1625897497830_0009/
Kill Command = /home/hdoop/hadoop-3.2.1/bin/mapred job  -kill
job_1625897497830_0009
Hadoop job information for Stage-3: number of mappers: 1; number of reducers: 0
2021-07-09 23:26:29,342 Stage-3 map = 0%,  reduce = 0%
2021-07-09 23:26:35,973 Stage-3 map = 100%,  reduce = 0%, Cumulative CPU 2.41 sec
MapReduce Total cumulative CPU time: 2 seconds 410 msec
Ended Job = job_1625897497830_0009
MapReduce Jobs Launched:
Stage-Stage-3: Map: 1   Cumulative CPU: 2.41 sec   HDFS Read: 8558 HDFS Write:
884 SUCCESS
Total MapReduce CPU Time Spent: 2 seconds 410 msec
OK
Aayesha     5020  ISE    6
Aditya      5000  ISE    6
Adarsh      5001  ISE    6
Nandan      5002  ISE    6
Rithuraj    5006  ISE    6
Asha  5007  ISE    6
Divine      5009  ISE    6
Aman 5010  ISE    6
Ashok       5011  ISE    6
Khushi      5014  ISE    6
```

```
Abhay       5015  ISE   6
Anusha      5017  ISE   6
garvit 5018  ISE   6
Anjali5019  ISE   6
Ajay  5022  CSE   1
Akash5025  CSE   1
Raghav      5003  CSE   1
Himanshu    5012  CSE   1
Inchara     5013  CSE   1
Vandita     5016  CSE   1
Gagan       5023  ECE   2
Ajay  5004  ECE   2
Harshith    5008  ECE   2
NULL        NULL        NULL        5
Guru  5024  AE    3
Aishwarya   5021  ME    4
Avani 5005  ME    4
Time taken: 33.359 seconds, Fetched: 27 row(s)
hive>
```