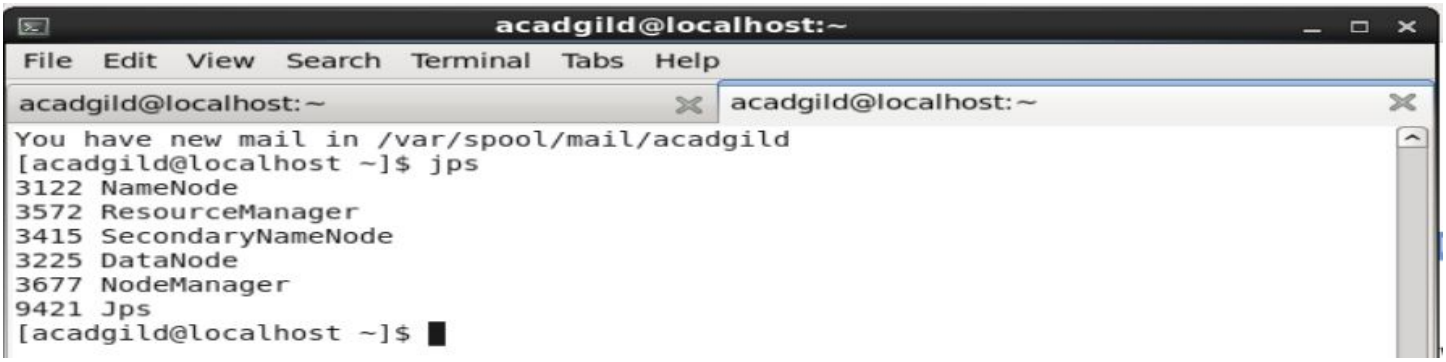
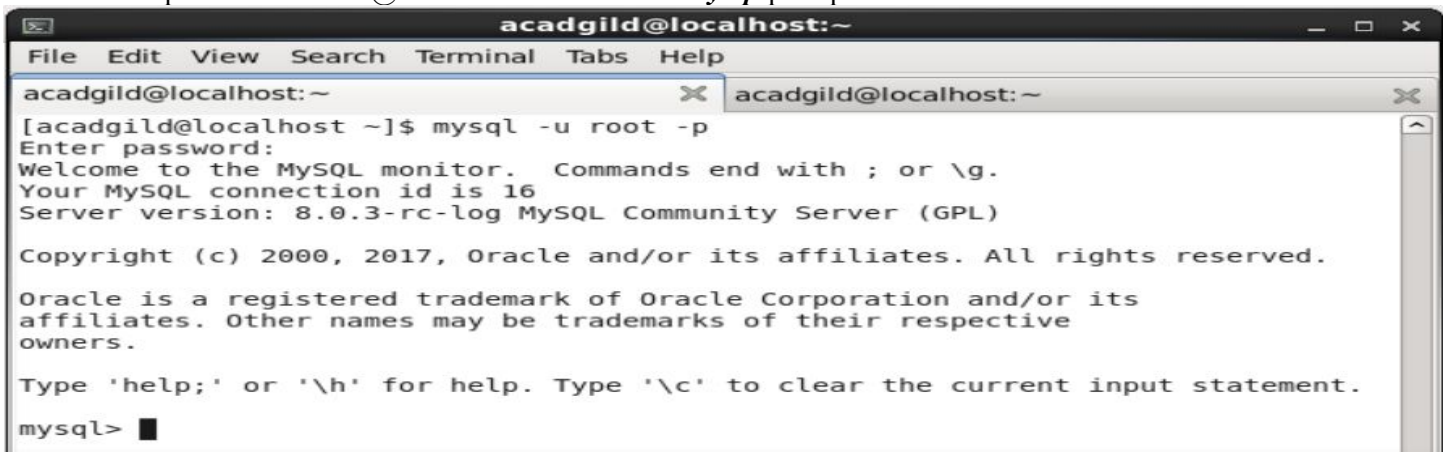
Session 6: Data Ingestion tool Sqoop & Introduction to Case Study 1.

1. Start the Hadoop daemons in the VM with the following command:
   $start-all.sh
2. Check all the hadoop daemons started in the VM with the following command:
   $jps

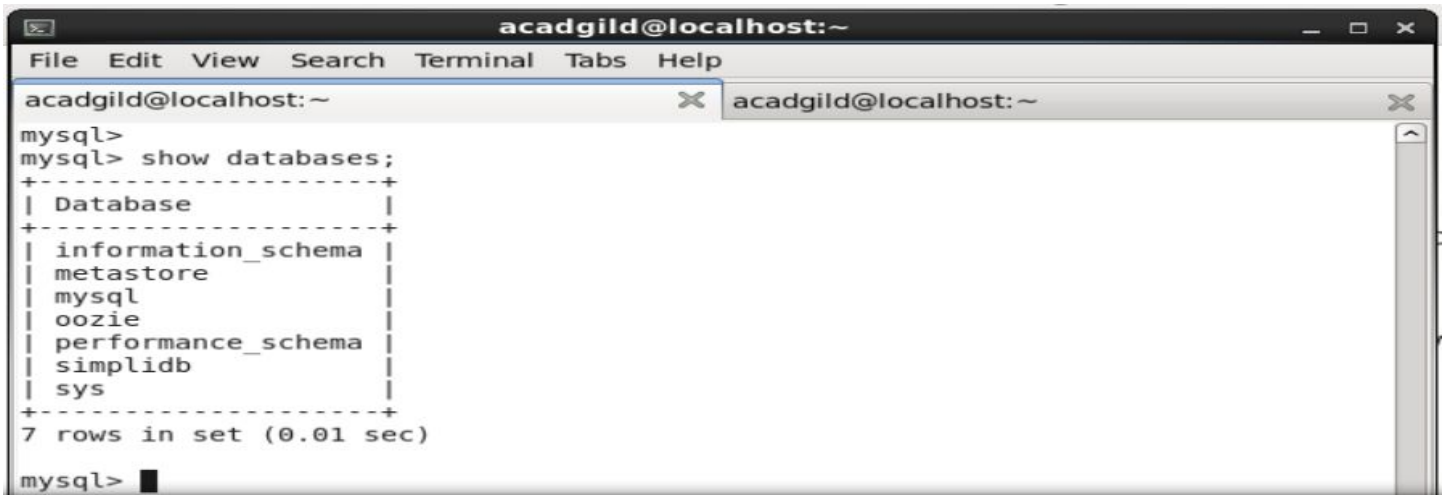


1. Now start the ***mySql*** with the following command:
   $mysql –u root –p
2. It will prompt the user to enter the password:
   Enter password:
3. Enter the password as ***Root@123*** it will enter into the ***mySql*** prompt as follows:



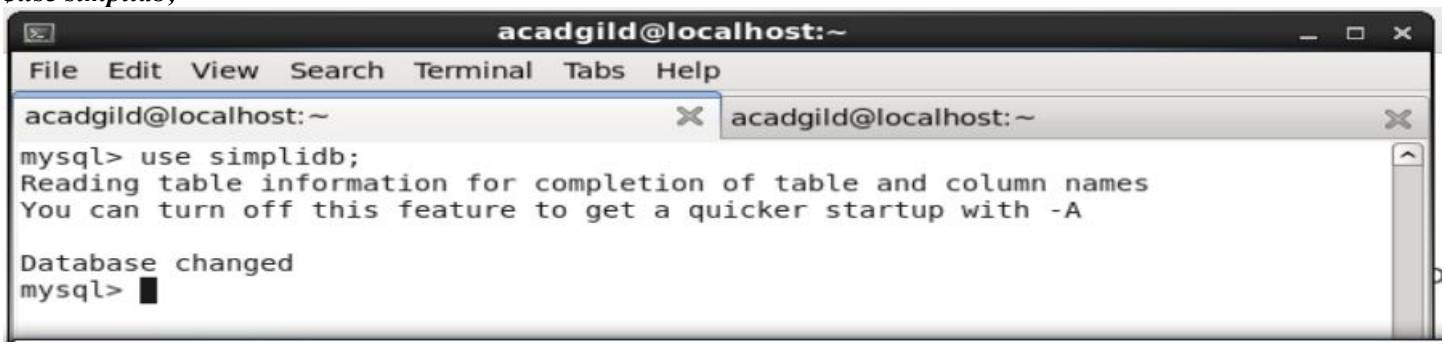
4. Now enter the command to show data base as follows:
   mysql> *show databases;*

5. It will show the list of databases as follows:



6. Now select the ***simplidb*** schema database as follows:

$use simplidb;

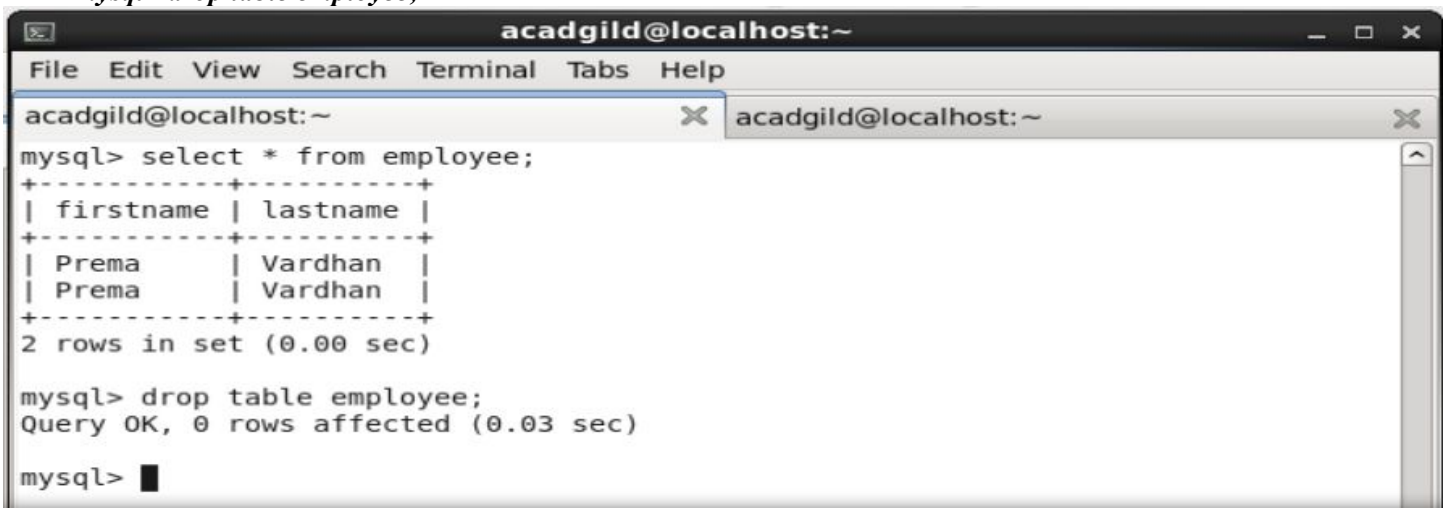


7. We can observe that the message Database changed for selecting the database simplidb.
8. Now list all the tables present in the ***simplidb*** database, as follows:

$show tables;

9. Now we can observe that there are two tables with **Person** and ***employee*** tables present in the database ***simplidb.***
10. Now for the demo from the beginning, I am going to drop the table employee to start from the beginning.
11. To drop the table we can use the command:

mysql> drop table employee;



12. We can see the message Query OK, 0 rows affected (0.03 sec) for deletion of an employee table.
13. Now we can create the table again with the following command:

mysql>create table employee (firstname varchar(10), lastname varchar(10));

14. We can see that the table is create with the name employee.
15. Now we can add records into the table employee using the following command:

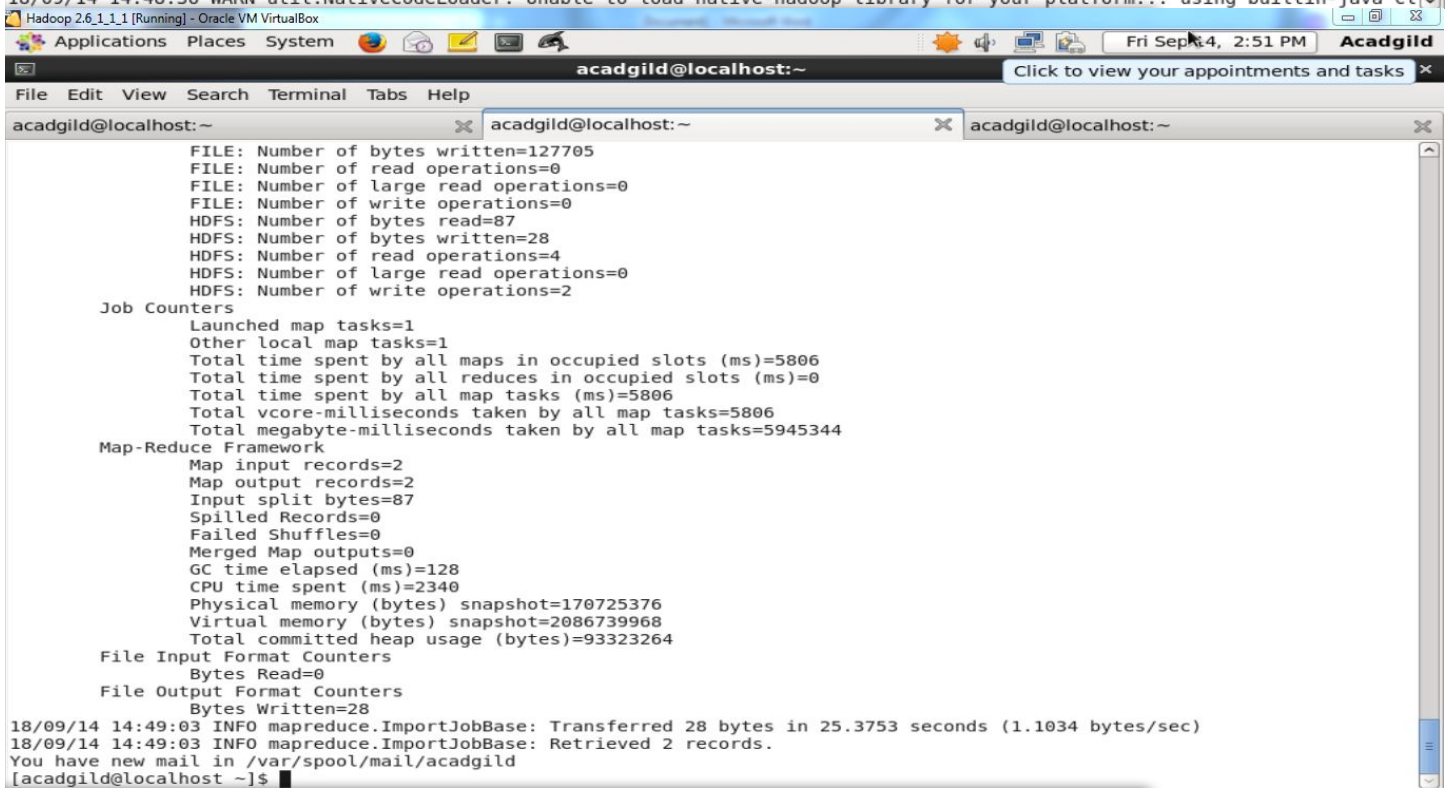
$insert into employee values ("Prema", "Vardhan");
$insert into employee values ("Bhaskar", "Reddy");

16. Now list the records present in the employee table as follows:

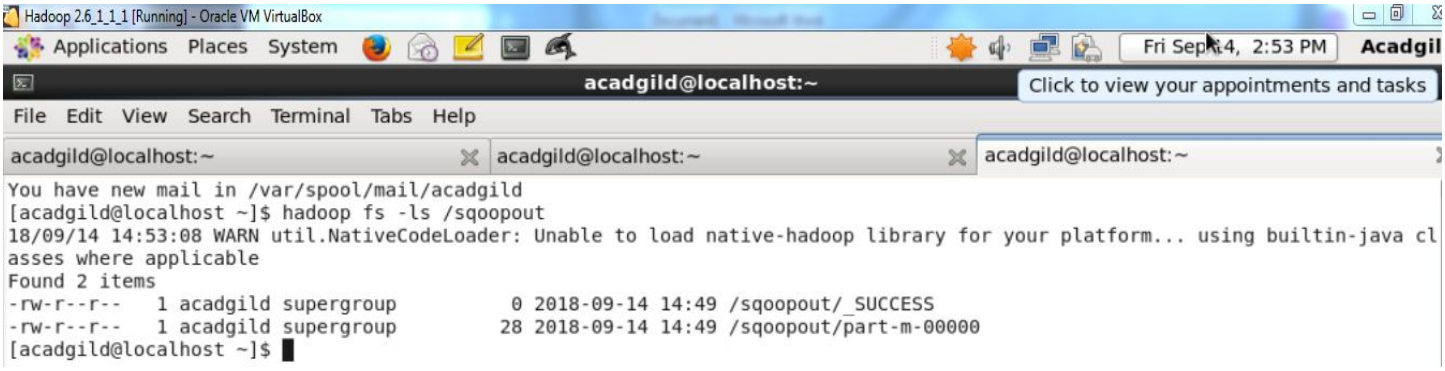
$select * from employee;

17. Now we can observe that there are two records present in the employee table.
18. Now we can use **sqoop** tool to import the table ***employee*** into hadoop file system using the following command:

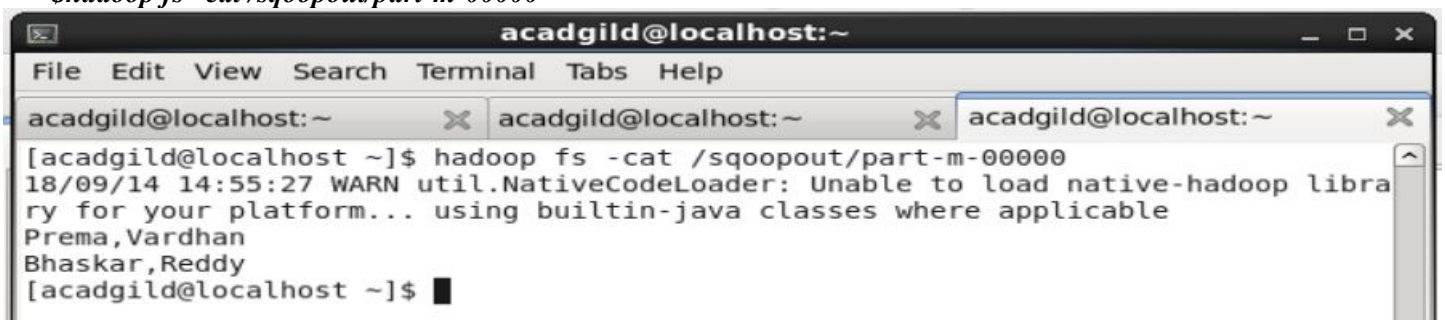
$sqoop import –connect jdbc:mysql://localhost/simplidb –table employee –usrname root –password Root@123 –target-dir /sqoopout –m 1



19. Now the output can be present in the ***/sqoopout*** folder present in the file system:
    $hadoop fs –ls /sqoopout



20. Now check the output present in the mapper output file: ***/sqoopout/part-m-00000.*** As follows:
    $hadoop fs –cat /sqoopout/part-m-00000

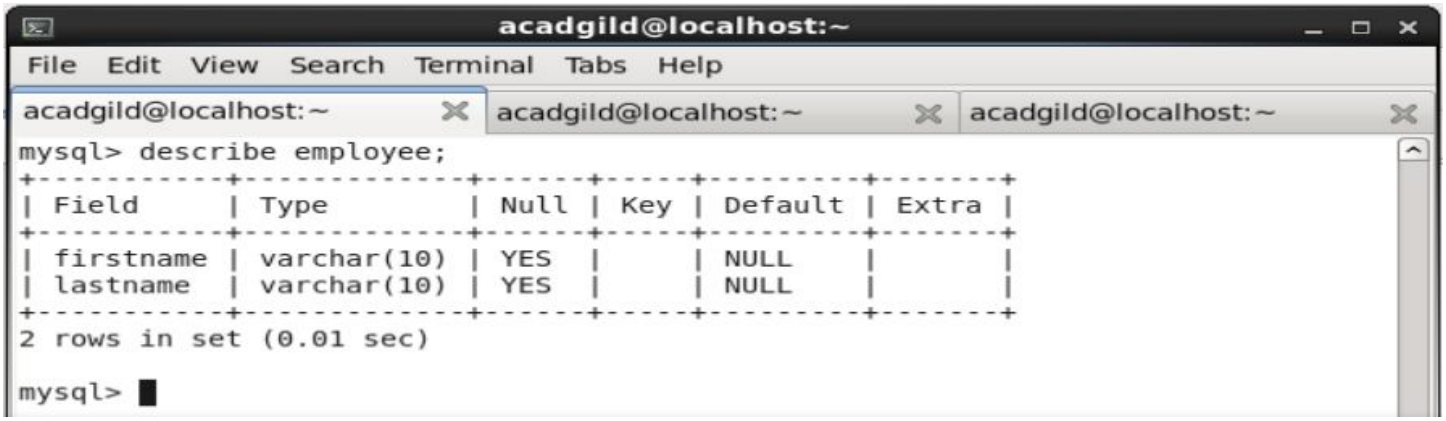


21. With this we can observe that the table is imported from mysql to hadoop.

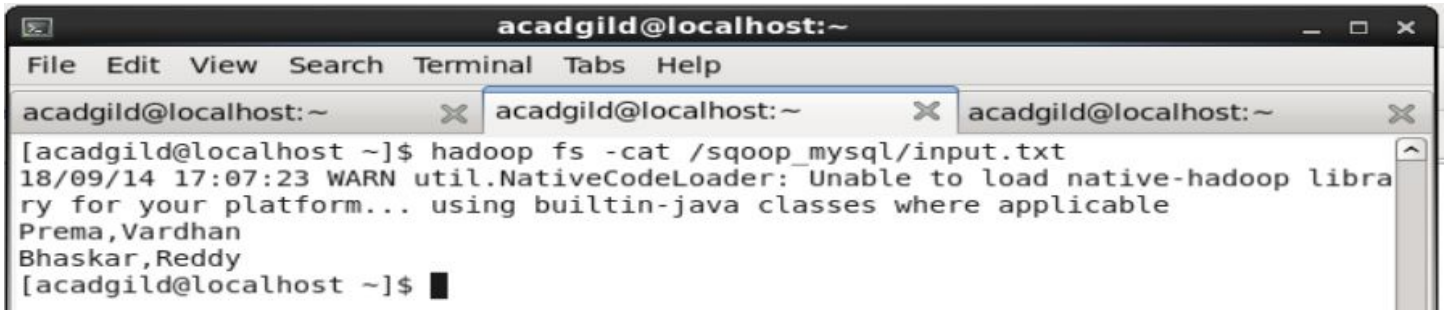
Task 1: Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table.

1. The above demo is used to import the table ***employee*** from the mysql to sqoop.
2. Now we are exporting the same employee records into the employee table in mysql.
3. To do that first we need to create a new database, or we can use an existing database also. In the above example we used the database as ***simplidb***.
4. Create table named ***employee*** for us the table employee is also present in the database ***simplidb***.
5. Now truncate if any records present the employee table to make table empty, as follows:
   mysql> truncate employee;

6. We also checked that if there are any records present, we got an Empty set represents no records present in the table
7. Now give a command, ***describe <tablename>*** to show the various fields and types of it. This will help in comparing the type of data present inside HDFS which is ready to be mapped.



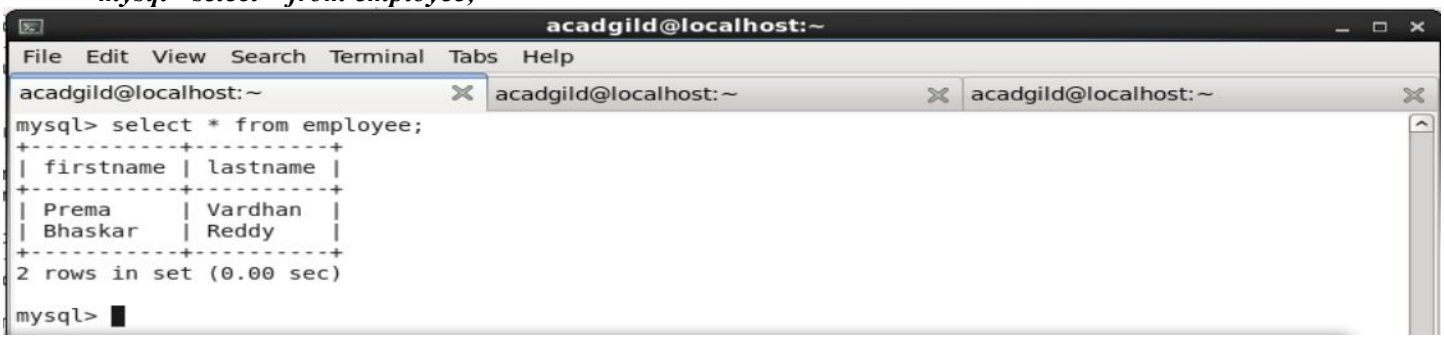
8. Now the file in the HDFS must have the same format as that of MYSql table, to enable the mapping of the data. The file will be like this:



9. Now we have a file named ***input.txt*** in HDFS with two records of *prema, Vardhan* and *Bhaskar, Reddy* with first name and second name separated.
10. Now export the ***input.txt*** file from HDFS to MySQL, as follows:
$sqoop export –connect jdbc:mysql://localhost/simplidb –username root –password Root@123 –table employee –export-dir /sqoop_mysql/ -m 1

```
                FILE: Number of bytes read=0
                FILE: Number of bytes written=127537
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=155
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=5499
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=5499
                Total vcore-milliseconds taken by all map tasks=5499
                Total megabyte-milliseconds taken by all map tasks=5630976
        Map-Reduce Framework
                Map input records=2
                Map output records=2
                Input split bytes=124
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=103
                CPU time spent (ms)=2150
                Physical memory (bytes) snapshot=178573312
                Virtual memory (bytes) snapshot=2084880384
                Total committed heap usage (bytes)=93323264
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
18/09/14 17:19:47 INFO mapreduce.ExportJobBase: Transferred 155 bytes in 22.4289 seconds (6.9107 bytes/sec)
18/09/14 17:19:47 INFO mapreduce.ExportJobBase: Exported 2 records.
[acadgild@localhost ~]$
```

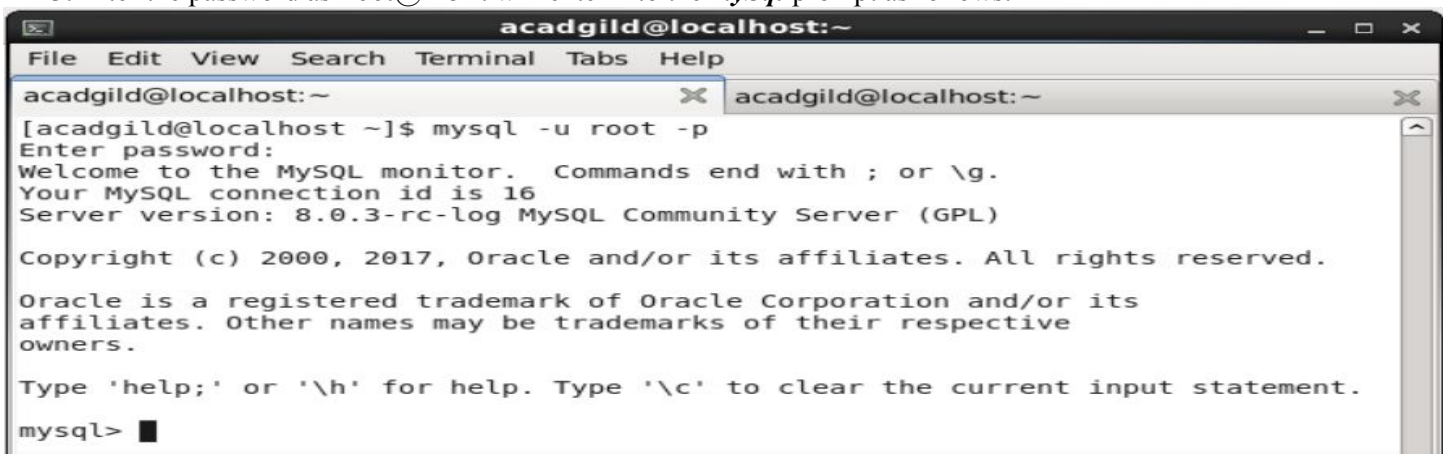
11. The message at the end represents “Exported 2 records” into the simplidb database into the table employee.
12. Now we will verify the list of records present in the employee table as:
    mysql> select * from employee;



13. Here we can observe that the two record are exported from HDFS to mysql.

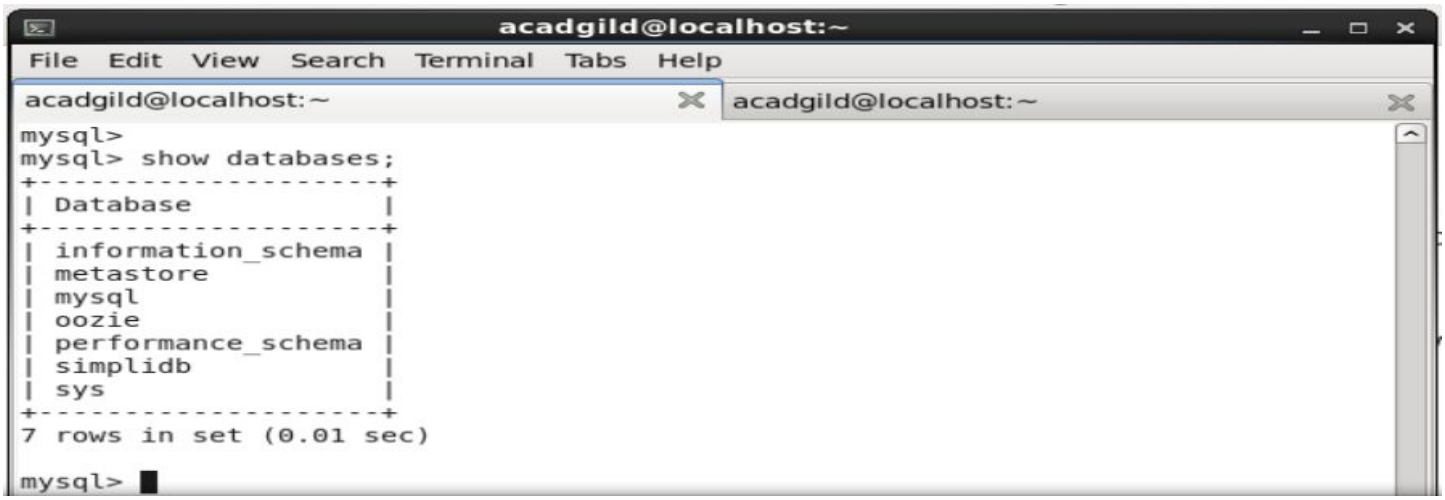
Task 2: Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table with parameter person_id = 3.

1. Now start the ***mySql*** with the following command:
$mysql –u root –p
2.It will prompt the user to enter the password:
Enter password:
3.Enter the password as ***Root@123*** it will enter into the ***mySql*** prompt as follows:

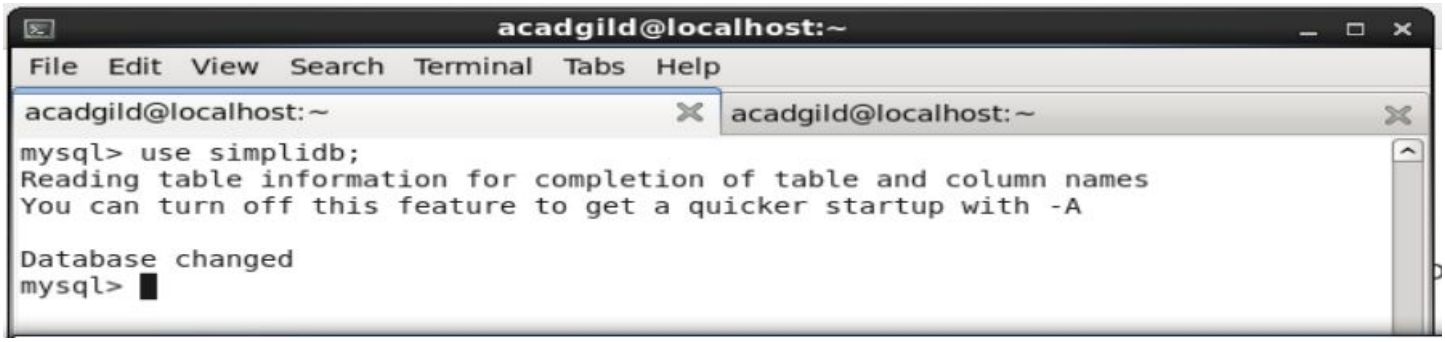


4.Now enter the command to show data base as follows:
mysql> ***show databases;***

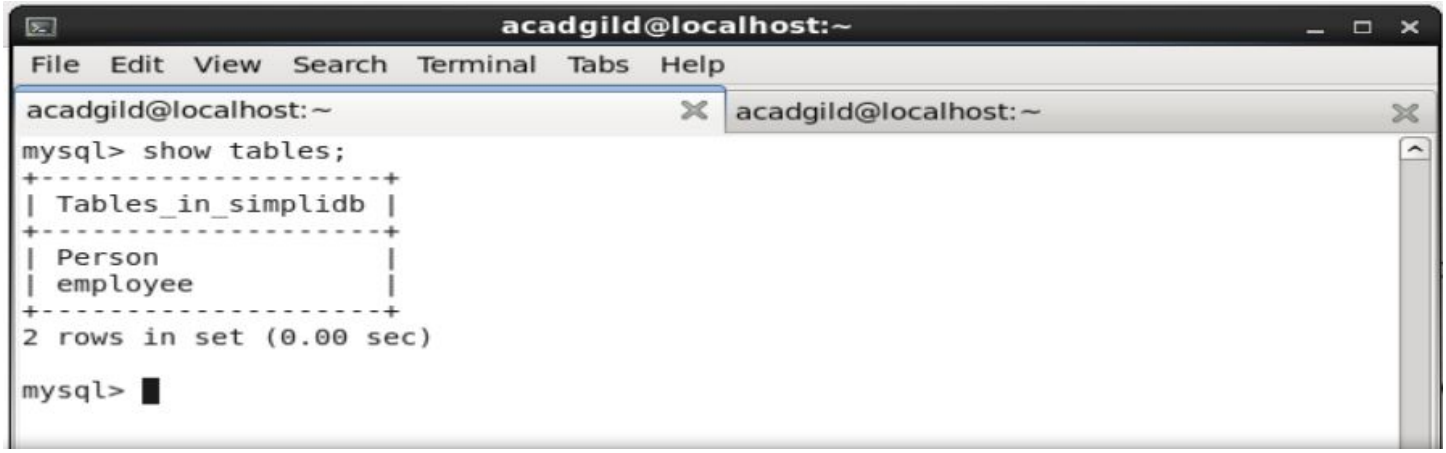
5.It will show the list of databases as follows:



6.Now select the ***simplidb*** schema database as follows:
mysql> use simplidb;



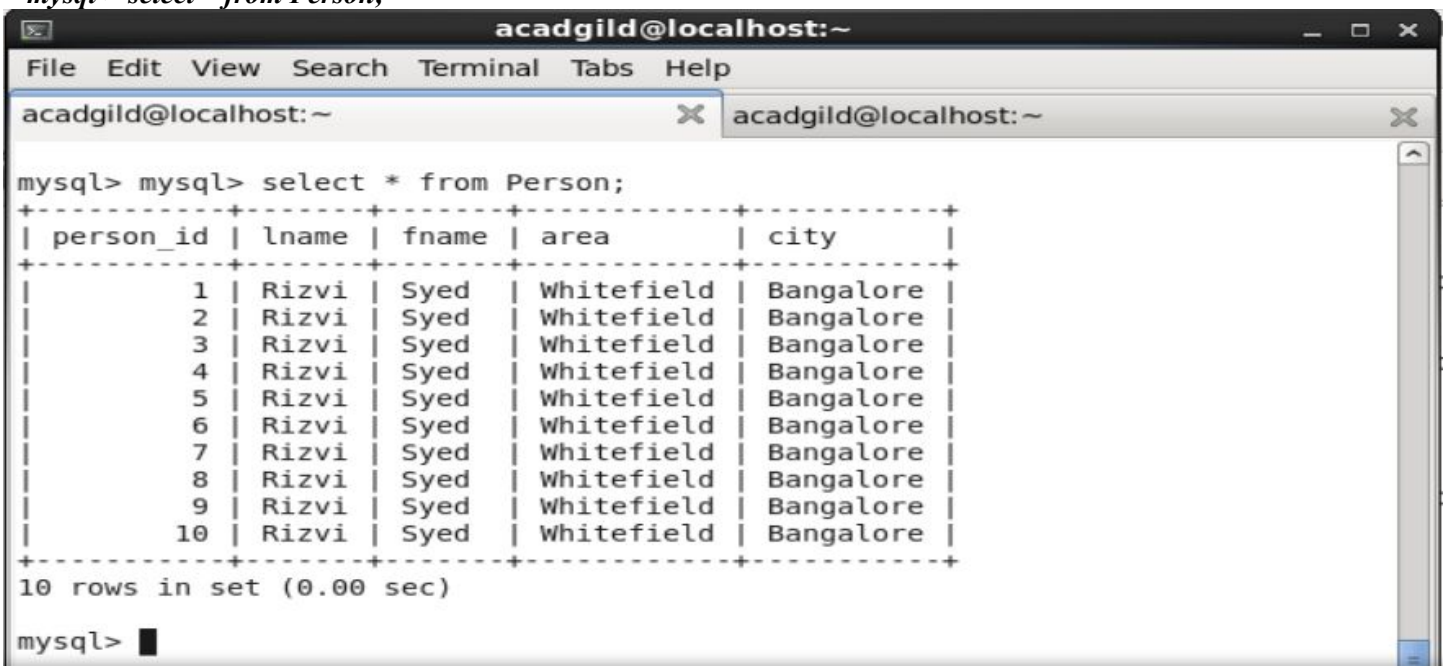
7.We can observe that the message Database changed for selecting the database simplidb.
8.Now list all the tables present in the ***simplidb*** database, as follows:
mysql>show tables;



9.Now we can observe that there are two tables with **Person** and ***employee*** tables present in the database ***simplidb.***

10. Print the list of records present in the Person table with the following command:

mysql > select * from Person;

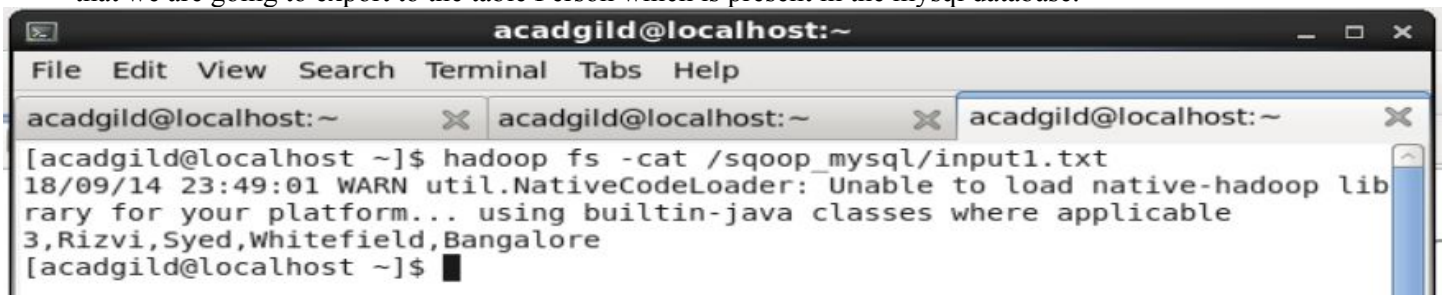


11. Now delete a record form the table ***Person****;*

mysql> delete from Person where person_id=3;

12. And here we can observe that after deletion, we used command ***select * from Person;*** to display all the records present in the Person table. But the record with person_id=3 is deleted.

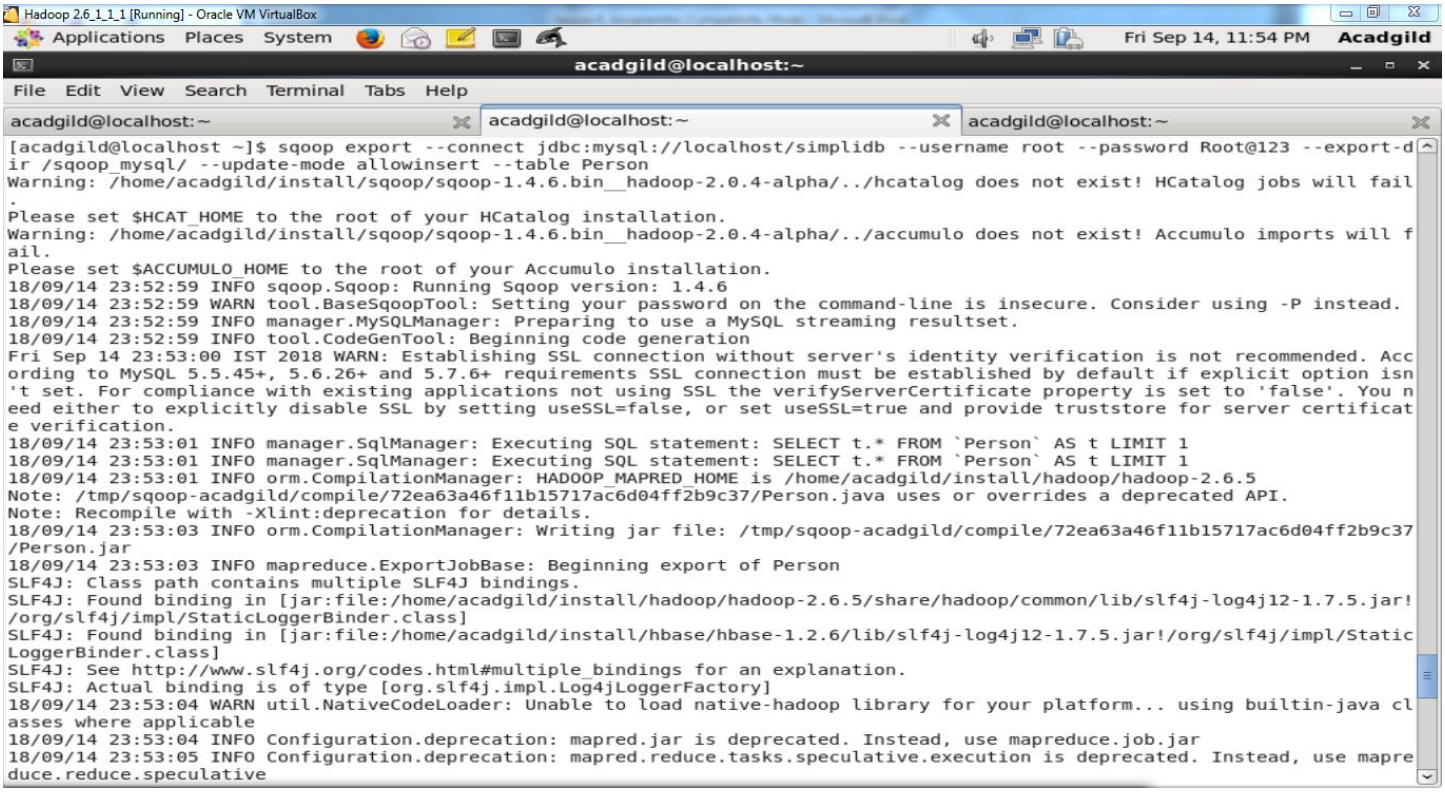
1. Now in the hadoop file system we can create a file named with ***/sqoop_mysql/input1.txt*** which contains a record that we are going to export to the table Person which is present in the mysql database.



2. We can observe that the record ***3, Rizvi, Syed, Whitefield, Bangalore*** is present in the directory ***/sqoop_mysql/input1.txt*** file.

3. Now export the file to update the record into the Person table with the following command:

$sqoop export –connect jdbc:mysql://localhost/simplidb –username root –password Root@123 –export-dir /sqoop_mysql/ --update-mode allowinsert –table Person



```
                FILE: Number of bytes written=510176
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=669
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=19
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=0
        Job Counters
                Launched map tasks=4
                Data-local map tasks=4
                Total time spent by all maps in occupied slots (ms)=79999
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=79999
                Total vcore-milliseconds taken by all map tasks=79999
                Total megabyte-milliseconds taken by all map tasks=81918976
        Map-Reduce Framework
                Map input records=1
                Map output records=1
                Input split bytes=561
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=1365
                CPU time spent (ms)=10500
                Physical memory (bytes) snapshot=668778496
                Virtual memory (bytes) snapshot=8344002560
                Total committed heap usage (bytes)=369098752
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
18/09/14 23:53:43 INFO mapreduce.ExportJobBase: Transferred 669 bytes in 38.1442 seconds (17.5387 bytes/sec)
18/09/14 23:53:43 INFO mapreduce.ExportJobBase: Exported 1 records.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

4. We can observe that at the end the message says ‘Exported 1 Records’ show the record is updated in the Person table.
5. We can see the Person table to check the update with the following command:
   mysql>select * from Person;

6. Here we can observe that before update the table contains only 9 rows, and after update, the table contains 10 rows with updating record person_id=3.