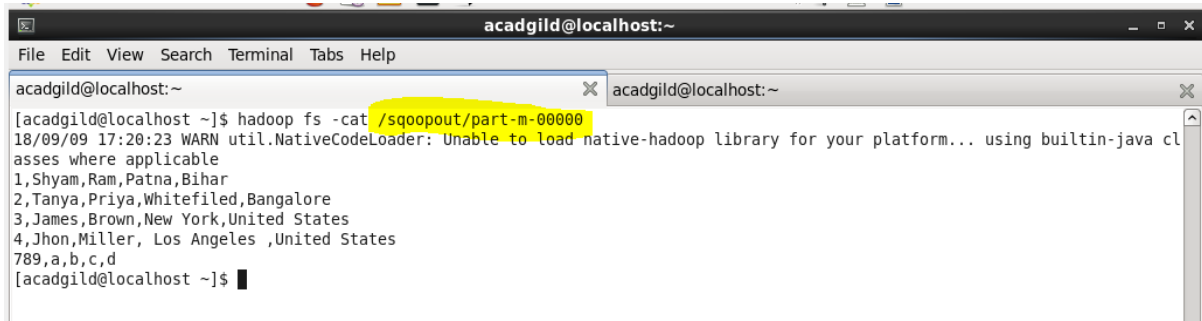


Assignment-6

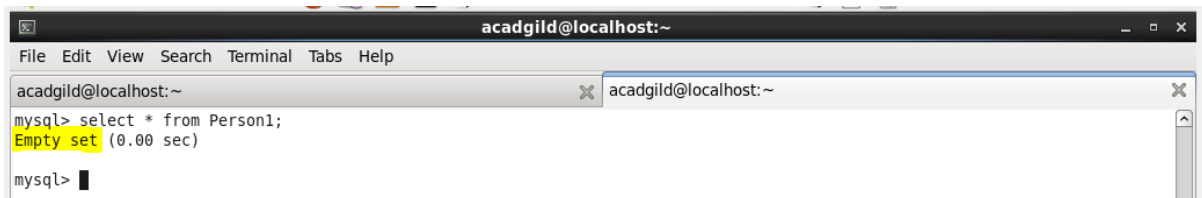
1) Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table

The data present in the sqoopout folder is as shown in the below screenshot.



```
acadmild@localhost:~  
[acadmild@localhost ~]$ hadoop fs -cat /sqoopout/part-m-00000  
18/09/09 17:20:23 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl  
asses where applicable  
1,Shyam,Ram,Patna,Bihar  
2,Tanya,Priya,Whitefiled,Bangalore  
3,James,Brown,New York,United States  
4,Jhon,Miller, Los Angeles ,United States  
789,a,b,c,d  
[acadmild@localhost ~]$
```

A table has been created in mysql with name **Person1**.



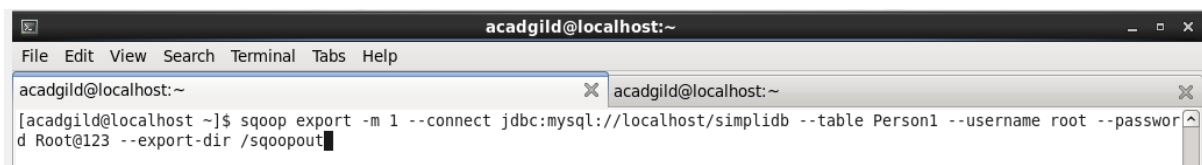
```
acadmild@localhost:~  
mysql> select * from Person1;  
Empty set (0.00 sec)  
mysql>
```

The below screenshot will show the description of table **Person1**.



```
acadmild@localhost:~  
mysql> describe Person1;  
+-----+-----+-----+-----+-----+-----+  
| Field | Type | Null | Key | Default | Extra |  
+-----+-----+-----+-----+-----+-----+  
| person_id | int(11) | NO | PRI | NULL | |  
| lname | varchar(20) | YES | | NULL | |  
| fname | varchar(20) | YES | | NULL | |  
| area | varchar(20) | YES | | NULL | |  
| city | varchar(20) | YES | | NULL | |  
+-----+-----+-----+-----+-----+-----+  
5 rows in set (0.01 sec)  
mysql>
```

Sqoop tool '**export**' is used to export HDFS file data from the **HDFS** to the table in **RDBMS**. The command for the **sqoop export** command for the above task is shown in the below screenshot.



```
acadmild@localhost:~  
[acadmild@localhost ~]$ sqoop export -m 1 --connect jdbc:mysql://localhost/simplidb --table Person1 --username root --password Root@123 --export-dir /sqoopout
```

--connect <jdbc-uri>: Specify JDBC connect string.

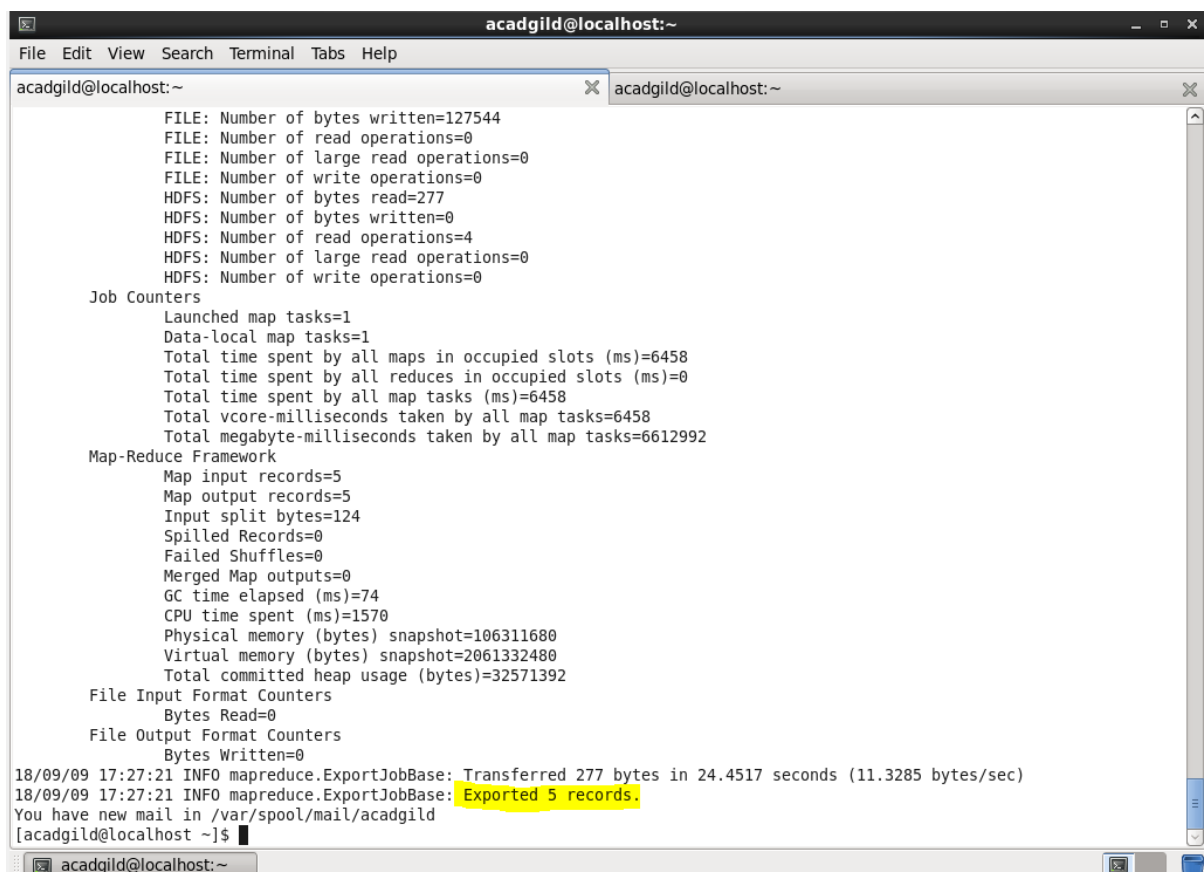
--table: specify table name.

--username: Set authentication username.

--password: Set authentication password.


--export-dir: specify path of the input file.

Highlighted text in the below screenshot shows that 5 records have been exported to the RDBMS table.



```
acacgild@localhost:~  
File Edit View Search Terminal Tabs Help  
acacgild@localhost:~  
acacgild@localhost:~  
FILE: Number of bytes written=127544  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=277  
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)=6458  
  Total time spent by all reduces in occupied slots (ms)=0  
  Total time spent by all map tasks (ms)=6458  
  Total vcore-milliseconds taken by all map tasks=6458  
  Total megabyte-milliseconds taken by all map tasks=6612992  
Map-Reduce Framework  
  Map input records=5  
  Map output records=5  
  Input split bytes=124  
  Spilled Records=0  
  Failed Shuffles=0  
  Merged Map outputs=0  
  GC time elapsed (ms)=74  
  CPU time spent (ms)=1570  
  Physical memory (bytes) snapshot=106311680  
  Virtual memory (bytes) snapshot=2061332480  
  Total committed heap usage (bytes)=32571392  
File Input Format Counters  
  Bytes Read=0  
File Output Format Counters  
  Bytes Written=0  
18/09/09 17:27:21 INFO mapreduce.ExportJobBase: Transferred 277 bytes in 24.4517 seconds (11.3285 bytes/sec)  
18/09/09 17:27:21 INFO mapreduce.ExportJobBase: Exported 5 records  
You have new mail in /var/spool/mail/acacgild  
[acacgild@localhost ~]$
```

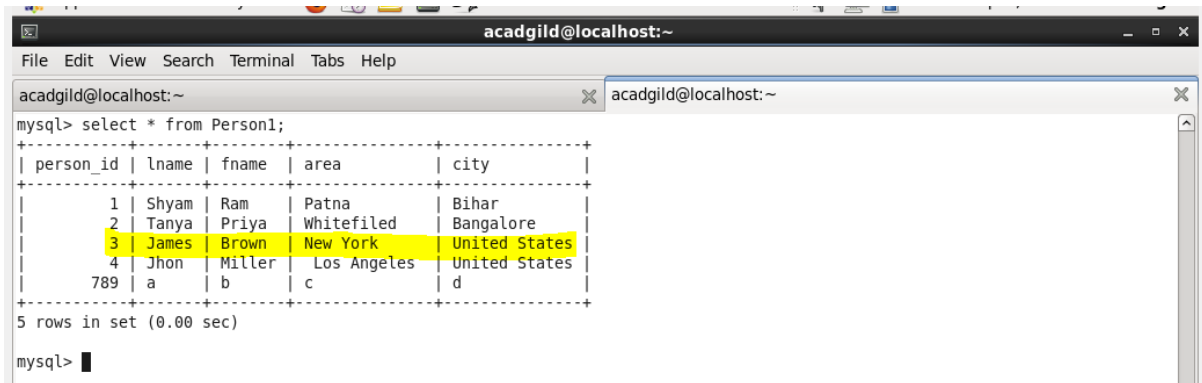
The below screenshot will shows the exported data in the table.



```
acacgild@localhost:~  
File Edit View Search Terminal Tabs Help  
acacgild@localhost:~  
acacgild@localhost:~  
mysql> select * from Person1;  
+-----+-----+-----+-----+-----+  
| person_id | lname | fname | area | city |  
+-----+-----+-----+-----+-----+  
| 1 | Shyam | Ram | Patna | Bihar |  
| 2 | Tanya | Priya | Whitefiled | Bangalore |  
| 3 | James | Brown | New York | United States |  
| 4 | Jhon | Miller | Los Angeles | United States |  
| 789 | a | b | c | d |  
+-----+-----+-----+-----+-----+  
5 rows in set (0.00 sec)  
mysql>
```

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

The data present in the table (Person1) for person_id=3 is highlighted in the screenshot below:



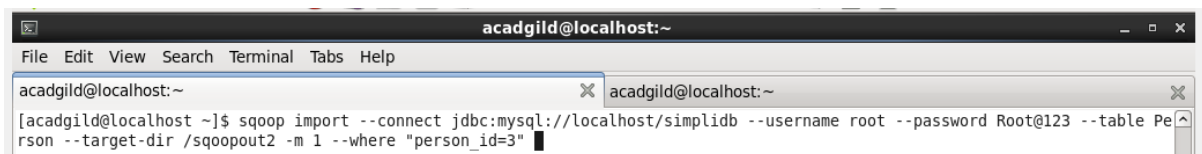
```
mysql> select * from Person1;
```

person_id	lname	fname	area	city
1	Shyam	Ram	Patna	Bihar
2	Tanya	Priya	Whitefield	Bangalore
3	James	Brown	New York	United States
4	Jhon	Miller	Los Angeles	United States
789	a	b	c	d

```
5 rows in set (0.00 sec)
```

```
mysql>
```

The 'Import' imports individual tables data from **RDBMS** to **HDFS**. Each row in a table is treated as a record in HDFS. The **sqoop import** command for the above task is as shown in the below screenshot.



```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/simplidb --username root --password Root@123 --table Person1 --target-dir /sqoopout2 -m 1 --where "person_id=3"
```

--connect <jdbc-uri>: Specify JDBC connect string.

--table: specify table name.

--username: Set authentication username.

--password: Set authentication password.

--target-dir: Specify the path to store the data.

Highlighted text in the below screenshot shows that records imported to HDFS directory.

```
acadgild@localhost:~  
File Edit View Search Terminal Tabs Help  
acadgild@localhost:~  
FILE: Number of bytes read=0  
FILE: Number of bytes written=128016  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=87  
HDFS: Number of bytes written=37  
HDFS: Number of read operations=4  
HDFS: Number of large read operations=0  
HDFS: Number of write operations=2  
Job Counters  
  Launched map tasks=1  
  Other local map tasks=1  
  Total time spent by all maps in occupied slots (ms)=6269  
  Total time spent by all reduces in occupied slots (ms)=0  
  Total time spent by all map tasks (ms)=6269  
  Total vcore-milliseconds taken by all map tasks=6269  
  Total megabyte-milliseconds taken by all map tasks=6419456  
Map-Reduce Framework  
  Map input records=1  
  Map output records=1  
  Input split bytes=87  
  Spilled Records=0  
  Failed Shuffles=0  
  Merged Map outputs=0  
  GC time elapsed (ms)=90  
  CPU time spent (ms)=1720  
  Physical memory (bytes) snapshot=108224512  
  Virtual memory (bytes) snapshot=2063437824  
  Total committed heap usage (bytes)=32571392  
File Input Format Counters  
  Bytes Read=0  
File Output Format Counters  
  Bytes Written=37  
18/09/09 17:37:41 INFO mapreduce.ImportJobBase: Transferred 37 bytes in 25.2009 seconds (1.4682 bytes/sec)  
18/09/09 17:37:41 INFO mapreduce.ImportJobBase: Retrieved 1 records.  
[acadgild@localhost ~]$
```

Record imported to HDFS directory is as shown in the below screenshot.

```
acadgild@localhost:~  
File Edit View Search Terminal Tabs Help  
acadgild@localhost:~  
[acadgild@localhost ~]$ hadoop fs -ls /sqoopout2  
18/09/09 17:39:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
Found 2 items  
-rw-r--r-- 1 acadgild supergroup 0 2018-09-09 17:37 /sqoopout2/ SUCCESS  
-rw-r--r-- 1 acadgild supergroup 37 2018-09-09 17:37 /sqoopout2/part-m-000000  
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout2/part-m-000000  
18/09/09 17:39:29 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
3,James,Brown,New York,United States  
[acadgild@localhost ~]$
```