

Creating a table in database

```
mysql> insert into Person1 values (1, 'Sam', 'Pam', 'Whitefield', 'bangalore');
^C
mysql> insert into Person1 values (1, 'Sam', 'Pam', 'Whitefield', 'bangalore');
Query OK, 1 row affected (0.01 sec)

mysql> insert into Person1 values (1, 'Tom', 'john', 'Whitefield', 'bangalore');
ERROR 1062 (23000): Duplicate entry '1' for key 'PRIMARY'
mysql> insert into Person1 values (2, 'Tom', 'john', 'Whitefield', 'bangalore');
Query OK, 1 row affected (0.01 sec)

mysql> insert into Person1 values (3, 'ken', 'john', 'Whitefield', 'bangalore');
Query OK, 1 row affected (0.00 sec)

mysql> insert into Person1 values (4, 'ben', 'john', 'Whitefield', 'bangalore');
Query OK, 1 row affected (0.00 sec)

mysql> insert into Person1 values (5, 'hary', 'john', 'Whitefield', 'bangalore');
Query OK, 1 row affected (0.00 sec)

mysql> select * from Person1;
+-----+-----+-----+-----+-----+
| person_id | lname | fname | area      | city      |
+-----+-----+-----+-----+-----+
| 1         | Sam   | Pam   | Whitefield | bangalore |
| 2         | Tom   | john  | Whitefield | bangalore |
| 3         | ken   | john  | Whitefield | bangalore |
| 4         | ben   | john  | Whitefield | bangalore |
| 5         | hary  | john  | Whitefield | bangalore |
+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)
```

Importing table in output directory sqoopout using import command

```
sqoop import --connect jdbc:mysql://localhost/simplidb --table Person --username root --password
Root@123 --target-dir /sqoopout --split-by Person_Id
```

Running the map reduce job and getting data present in sqoopout directory

```
127.0.0.1 (acagild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split Multitex Tunneling Packages Settings Help
Quick connect...
Sessions
Name
...
Videos
Templates
Public
Pictures
Music
install
eclipse-workspace
eclipse
Downloads
Documents
Desktop
webclipse
tooling
thumbnails
.swt
.ssh
.sqoop
.scalade
.pulide
.pki
.p2
.oracle_jre_usage
.nautius
.mozila
.m2
.local
.gifs
.gnupg
.gnote
.gnome2
.gconf
.fontconfig
Remote monitoring
Follow terminal folder

Map output records=12
Input split bytes=452
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=1449
CPU time spent (ms)=7730
Physical memory (bytes) snapshot=435867648
Virtual memory (bytes) snapshot=8253767680
Total committed heap usage (bytes)=130285568
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=436
18/07/26 19:41:37 INFO mapreduce.ImportJobBase: Transferred 436 bytes in 44.2036 seconds (9.8635 bytes/sec)
18/07/26 19:41:37 INFO mapreduce.ImportJobBase: Retrieved 12 records.
You have new mail in /var/spool/mail/acagild
[acagild@localhost ~]$ hadoop fs -cat /sqoopout
cat: /sqoopout: Is a directory
[acagild@localhost ~]$ hadoop fs -ls /sqoopout
18/07/26 19:42:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 5 items
-rw-r--r-- 1 acagild supergroup          0 2018-07-26 19:41 /sqoopout/ SUCCESS
-rw-r--r-- 1 acagild supergroup       108 2018-07-26 19:41 /sqoopout/part-m-00000
-rw-r--r-- 1 acagild supergroup       109 2018-07-26 19:41 /sqoopout/part-m-00001
-rw-r--r-- 1 acagild supergroup       108 2018-07-26 19:41 /sqoopout/part-m-00002
-rw-r--r-- 1 acagild supergroup       111 2018-07-26 19:41 /sqoopout/part-m-00003
[acagild@localhost ~]$ hadoop fs -cat /sqoopout/*
18/07/26 19:42:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1,dutta,marina,BemLayout,Bangalore
2,dutta,marina,BemLayout,Bangalore
3,dutta,marina,BemLayout,Bangalore
4,dutta,marina,BemLayout,Bangalore
5,dutta,marina,BemLayout,Bangalore
6,dutta,marina,BemLayout,Bangalore
7,dutta,marina,BemLayout,Bangalore
8,dutta,marina,BemLayout,Bangalore
9,dutta,marina,BemLayout,Bangalore
10,dutta,marina,BemLayout,Bangalore
11,dutta,marina,BemLayout,Bangalore
12,dutta,marina,BemLayout,Bangalore
[acagild@localhost ~]$
```

```
127.0.0.1 (acagild)
Terminal Sessions View X server Tools Games Settings Macros Help
Session Servers Tools Games Sessions View Split Multitex Tunneling Packages Settings Help
Quick connect...
Sessions
Name
...
Videos
Templates
Public
Pictures
Music
install
eclipse-workspace
eclipse
Downloads
Documents
Desktop
webclipse
tooling
thumbnails
.swt
.ssh
.sqoop
.scalade
.pulide
.pki
.p2
.oracle_jre_usage
.nautius
.mozila
.m2
.local
.gifs
.gnupg
.gnote
.gnome2
.gconf
.fontconfig
Remote monitoring
Follow terminal folder

Map output records=12
Input split bytes=452
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=1449
CPU time spent (ms)=7730
Physical memory (bytes) snapshot=435867648
Virtual memory (bytes) snapshot=8253767680
Total committed heap usage (bytes)=130285568
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=436
18/07/26 19:41:37 INFO mapreduce.ImportJobBase: Transferred 436 bytes in 44.2036 seconds (9.8635 bytes/sec)
18/07/26 19:41:37 INFO mapreduce.ImportJobBase: Retrieved 12 records.
You have new mail in /var/spool/mail/acagild
[acagild@localhost ~]$ hadoop fs -cat /sqoopout
cat: /sqoopout: Is a directory
[acagild@localhost ~]$ hadoop fs -ls /sqoopout
18/07/26 19:42:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 5 items
-rw-r--r-- 1 acagild supergroup          0 2018-07-26 19:41 /sqoopout/ SUCCESS
-rw-r--r-- 1 acagild supergroup       108 2018-07-26 19:41 /sqoopout/part-m-00000
-rw-r--r-- 1 acagild supergroup       109 2018-07-26 19:41 /sqoopout/part-m-00001
-rw-r--r-- 1 acagild supergroup       108 2018-07-26 19:41 /sqoopout/part-m-00002
-rw-r--r-- 1 acagild supergroup       111 2018-07-26 19:41 /sqoopout/part-m-00003
[acagild@localhost ~]$ hadoop fs -cat /sqoopout/*
18/07/26 19:42:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1,dutta,marina,BemLayout,Bangalore
2,dutta,marina,BemLayout,Bangalore
3,dutta,marina,BemLayout,Bangalore
4,dutta,marina,BemLayout,Bangalore
5,dutta,marina,BemLayout,Bangalore
6,dutta,marina,BemLayout,Bangalore
7,dutta,marina,BemLayout,Bangalore
8,dutta,marina,BemLayout,Bangalore
9,dutta,marina,BemLayout,Bangalore
10,dutta,marina,BemLayout,Bangalore
11,dutta,marina,BemLayout,Bangalore
12,dutta,marina,BemLayout,Bangalore
[acagild@localhost ~]$
```

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

```
sqoop export --connect jdbc:mysql://localhost/simplidb --table Person2 --username root -P --export-dir /sqoopout/
```

created a table Person2 in MySQL.

```
mysql> describe Person2;
```

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 (0.02 sec)

Exporting data present in sqoopout folder to Person2 table in Mysql

```
sqoop export --connect jdbc:mysql://localhost/simplidb --table Person2 --username root -P --export-dir /sqoopout/
```

Here we are exporting data from Person table in HDFS to mysql Person2. We are running the mapreduce job to export data

18/07/26 19:48:05 INFO mapreduce.Job: map 100% reduce 0%
18/07/26 19:48:05 INFO mapreduce.Job: Job_1532613584292_0004 completed successfully
18/07/26 19:48:06 INFO mapreduce.Job: Counters: 31

File System Counters
FILE: Number of bytes read=0
FILE: Number of bytes written=382632
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=999
HDFS: Number of bytes written=0
HDFS: Number of read operations=18
HDFS: Number of large read operations=0
HDFS: Number of write operations=0

Job Counters
Launched map tasks=3
Other local map tasks=1
Data-local map tasks=2
Total time spent by all maps in occupied slots (ms)=55503
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=55503
Total core-millisecons taken by all map tasks=55503
Total megabyte-millisecons taken by all map tasks=56835872

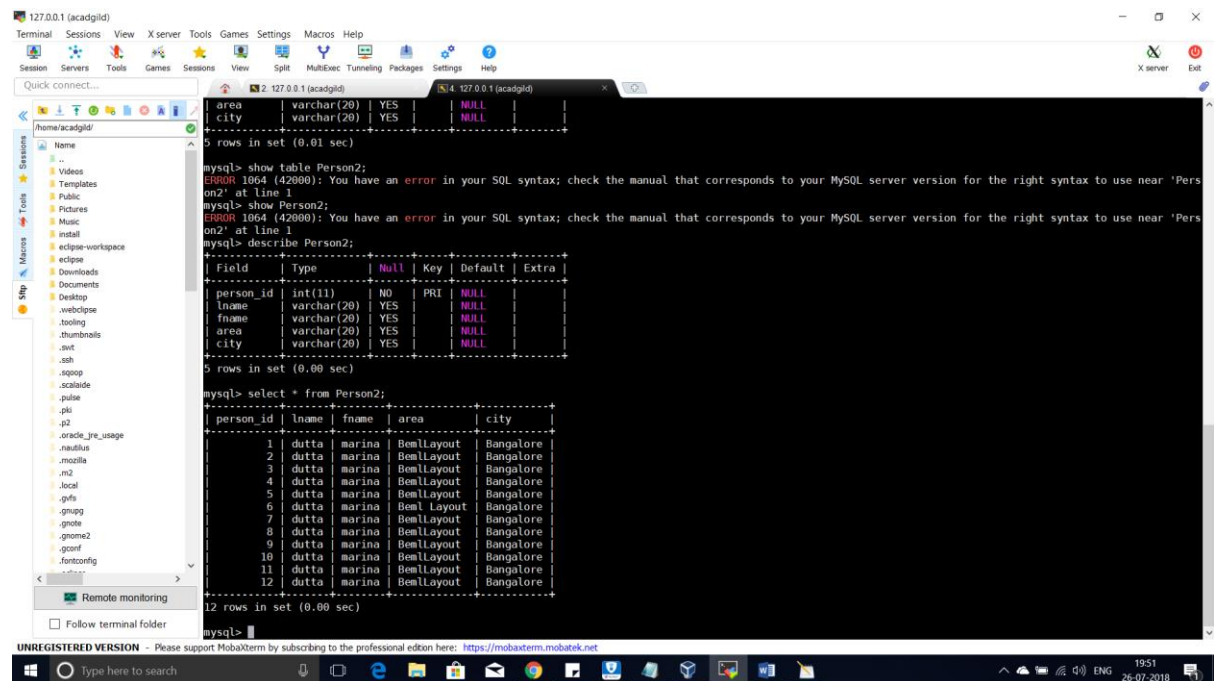
Map-Reduce Framework
Map input records=12
Map output records=12
Input split bytes=492
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=763
CPU time spent (ms)=5090
Physical memory (bytes) snapshot=325513216
Virtual memory (bytes) snapshot=618489728
Total committed heap usage (bytes)=97714176

File Input Format Counters
Bytes Read=0

File Output Format Counters
Bytes Written=0

18/07/26 19:48:06 INFO mapreduce.ExportJobBase: Transferred 999 bytes in 33.8023 seconds (29.5542 bytes/sec)
18/07/26 19:48:06 INFO mapreduce.ExportJobBase: Exported 12 records.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]\$

When we ran the Select* from Person2 our data from HDFS got copied into mysql

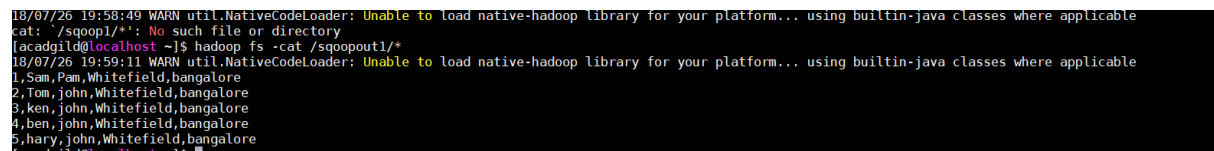


```
mysql> show table Person2;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'Person2' at line 1
mysql> show Person2;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'Person2' at line 1
mysql> describe Person2;
+-----+-----+-----+-----+-----+-----+
| 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 (0.00 sec)

mysql> select * from Person2;
+-----+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+-----+
| 1 | dutta | marina | BemLayout | Bangalore |
| 2 | dutta | marina | BemLayout | Bangalore |
| 3 | dutta | marina | BemLayout | Bangalore |
| 4 | dutta | marina | BemLayout | Bangalore |
| 5 | dutta | marina | BemLayout | Bangalore |
| 6 | dutta | marina | BemLayout | Bangalore |
| 7 | dutta | marina | BemLayout | Bangalore |
| 8 | dutta | marina | BemLayout | Bangalore |
| 9 | dutta | marina | BemLayout | Bangalore |
| 10 | dutta | marina | BemLayout | Bangalore |
| 11 | dutta | marina | BemLayout | Bangalore |
| 12 | dutta | marina | BemLayout | Bangalore |
+-----+-----+-----+-----+-----+
12 rows in set (0.00 sec)
```

Some other examples:

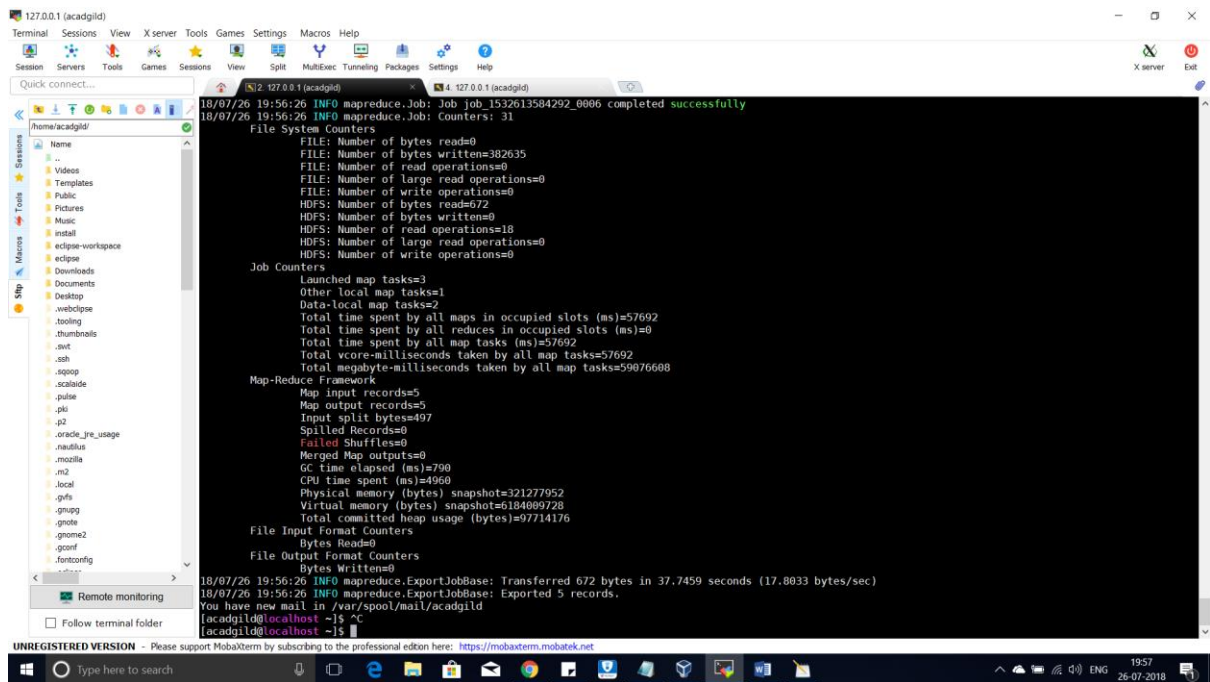
At sqoop1 directory we have following data in hdfs



```
18/07/26 19:58:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
cat: /sqoop1/*: No such file or directory
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout1/*
18/07/26 19:59:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1, Sam, Pam, Whitefield, bangalore
2, Tom, John, Whitefield, bangalore
3, Ken, John, Whitefield, bangalore
4, Ben, John, Whitefield, bangalore
5, Mary, John, Whitefield, bangalore
[acadgild@localhost ~]$
```

Here we created table Person3 in sql and exporting data from sqoop1 directory to table Person3

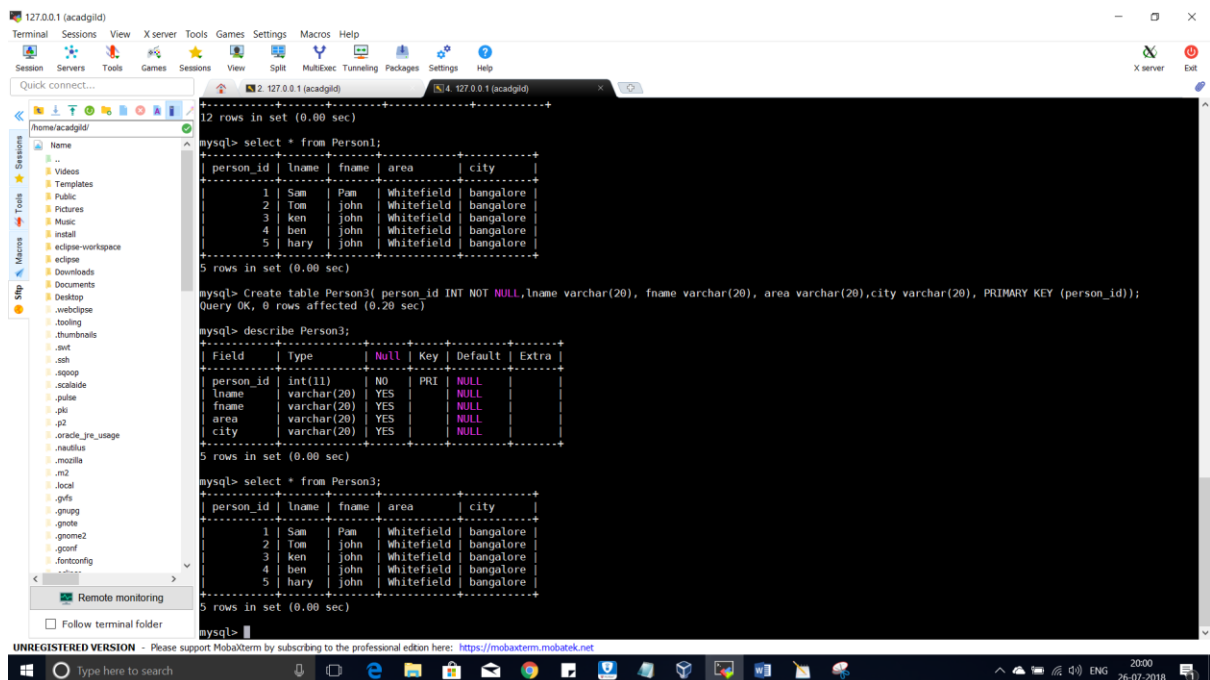
```
sqoop export --connect jdbc:mysql://localhost/simplidb --table Person3 --username root -P --export-dir /sqoopout1
```



The screenshot shows a terminal window with the following output:

```
18/07/26 19:56:26 INFO mapreduce.Job: Job job_1532613584292_0006 completed successfully
18/07/26 19:56:26 INFO mapreduce.Job: Counters: 31
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=382635
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=672
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=18
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0
Job Counters
  Launched map tasks=3
  Other local map tasks=1
  Data-local map tasks=2
  Total time spent by all maps in occupied slots (ms)=57692
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=57692
  Total vcore-milliseconds taken by all map tasks=57692
  Total megabyte-milliseconds taken by all map tasks=59076608
Map-Reduce Framework
  Map input records=5
  Map output records=5
  Input split bytes=497
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=790
  CPU time spent (ms)=4960
  Physical memory (bytes) snapshot=321277952
  Virtual memory (bytes) snapshot=6184009728
  Total committed heap usage (bytes)=97714176
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
18/07/26 19:56:26 INFO mapreduce.ExportJobBase: Transferred 672 bytes in 37.7459 seconds (17.8033 bytes/sec)
18/07/26 19:56:26 INFO mapreduce.ExportJobBase: Exported 5 records.
[acadgild@localhost ~]$ cd
[acadgild@localhost ~]$
```

Checking at mysql using select *from Person3;



The screenshot shows a terminal window with the following output:

```
mysql> select * from Person1;
+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+
| 1 | Sam | Pam | Whitefield | bangalore |
| 2 | Tom | john | Whitefield | bangalore |
| 3 | ken | john | Whitefield | bangalore |
| 4 | ben | john | Whitefield | bangalore |
| 5 | hary | john | Whitefield | bangalore |
+-----+-----+-----+-----+
5 rows in set (0.00 sec)

mysql> Create table Person3( person_id INT NOT NULL,lname varchar(20), fname varchar(20), area varchar(20),city varchar(20), PRIMARY KEY (person_id));
Query OK, 0 rows affected (0.20 sec)

mysql> describe Person3;
+-----+-----+-----+-----+-----+-----+
| 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 (0.00 sec)

mysql> select * from Person3;
+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+
| 1 | Sam | Pam | Whitefield | bangalore |
| 2 | Tom | john | Whitefield | bangalore |
| 3 | ken | john | Whitefield | bangalore |
| 4 | ben | john | Whitefield | bangalore |
| 5 | hary | john | Whitefield | bangalore |
+-----+-----+-----+-----+
5 rows in set (0.00 sec)

mysql>
```

Task2 :

sqoop export --connect jdbc:mysql://localhost/simplidb --table Person_1 --username root -P --export-dir /sqoopout1/part-m-00002 -- where "person_id =3"

```
mysql> select * from Person_1;
+-----+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+-----+
| 3 | ken | john | Whitefield | bangalore |
+-----+-----+-----+-----+-----+
1 row in set (0.00 sec)

mysql>
```

Here we need to export only person id=3 i.e. only one entry. A Person_1 table is created in mysql and one entry person_id =3 is exported from HDFS to mysql by sqoop tool. person_id =3 is stored in directory sqoopout1/part-m-00002