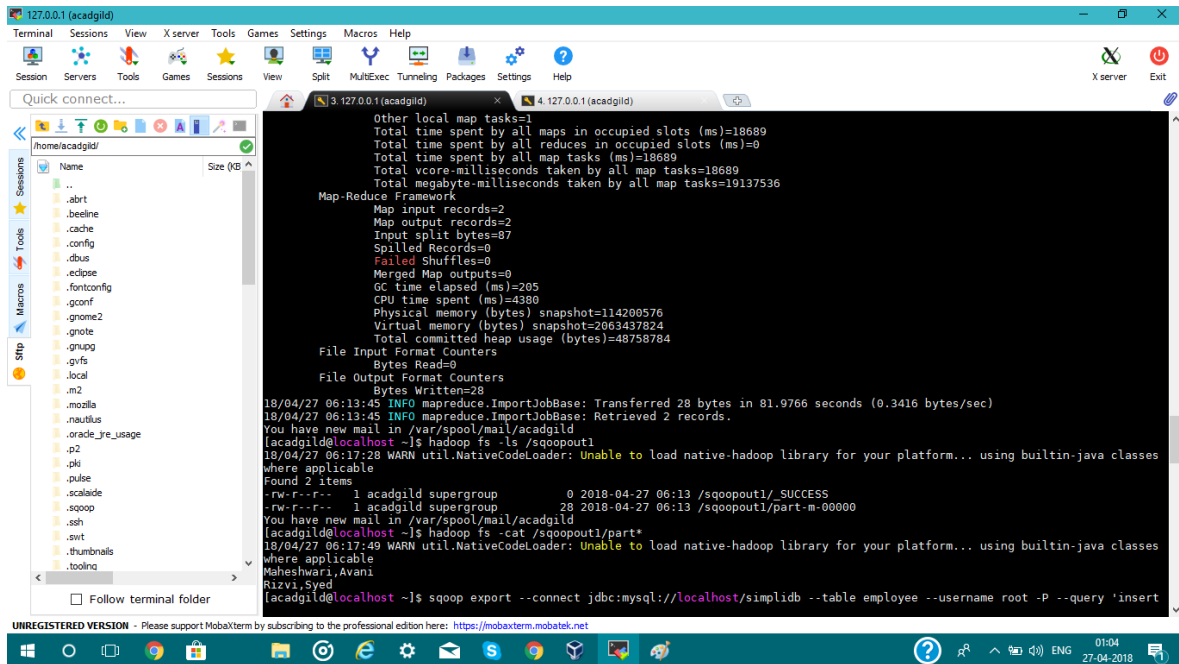


Assignment 6

Task1

Use SQOOP tool to export data present in SQOOPOUT folder made while demo of import table.

Data in SQOOPOUT folder



The screenshot shows a MobaXterm terminal window with a file explorer on the left and a terminal window on the right. The file explorer shows the directory structure of the user's home directory, including folders like .abrt, .beeline, .cache, .config, .dbus, .edpase, .fontconfig, .gnome2, .gnome, .gnupg, .grfs, .local, .m2, .mozilla, .nautilus, .oracle_jre_usage, .p2, .plk, .pulse, .scalade, .ssh, .swt, .thumbnails, and .tooling. The terminal window displays the output of a Hadoop MapReduce job, including statistics on map tasks, reduce tasks, and file input/output. The output shows that the job completed successfully, with 2 records transferred. The terminal also shows the command to export data from the SQOOPOUT folder to a MySQL database using the sqoop tool.

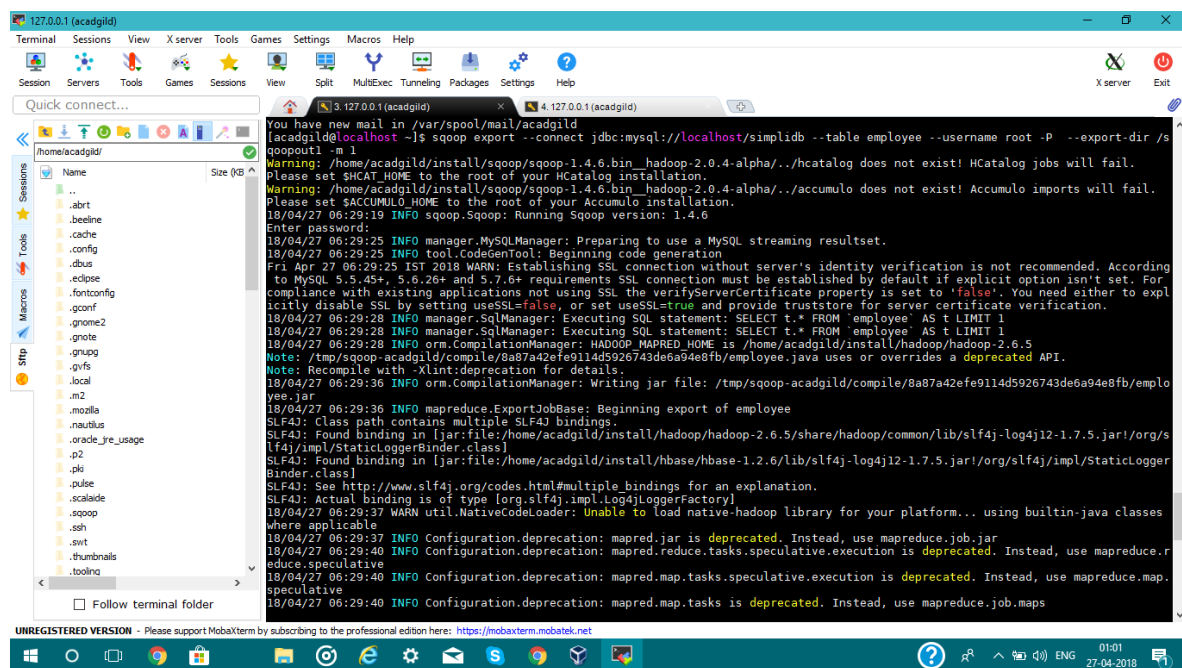
```
Other local map tasks=1
Total time spent by all maps in occupied slots (ms)=18689
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=18689
Total vcore-milliseconds taken by all map tasks=18689
Total megabyte-milliseconds taken by all map tasks=19137536
Map-Reduce Framework
  Map input records=2
  Map output records=2
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=205
  CPU time spent (ms)=4380
  Physical memory (bytes) snapshot=114200576
  Virtual memory (bytes) snapshot=2063437824
  Total committed heap usage (bytes)=48758784
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=28
18/04/27 06:13:45 INFO mapreduce.ImportJobBase: Transferred 28 bytes in 81.9766 seconds (0.3416 bytes/sec)
18/04/27 06:13:45 INFO mapreduce.ImportJobBase: Retrieved 2 records.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /sqoopout1
18/04/27 06:17:28 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 2018-04-27 06:13 /sqoopout1/ SUCCESS
-rw-r--r-- 1 acadgild supergroup          28 2018-04-27 06:13 /sqoopout1/part-m-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout1/part*
18/04/27 06:17:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Maheshwari,Avani
Rizvi,Syed
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/simplidb --table employee --username root -P --query 'insert
```

The data present on sqoopout1 folder is :

Maheshwari, Avani

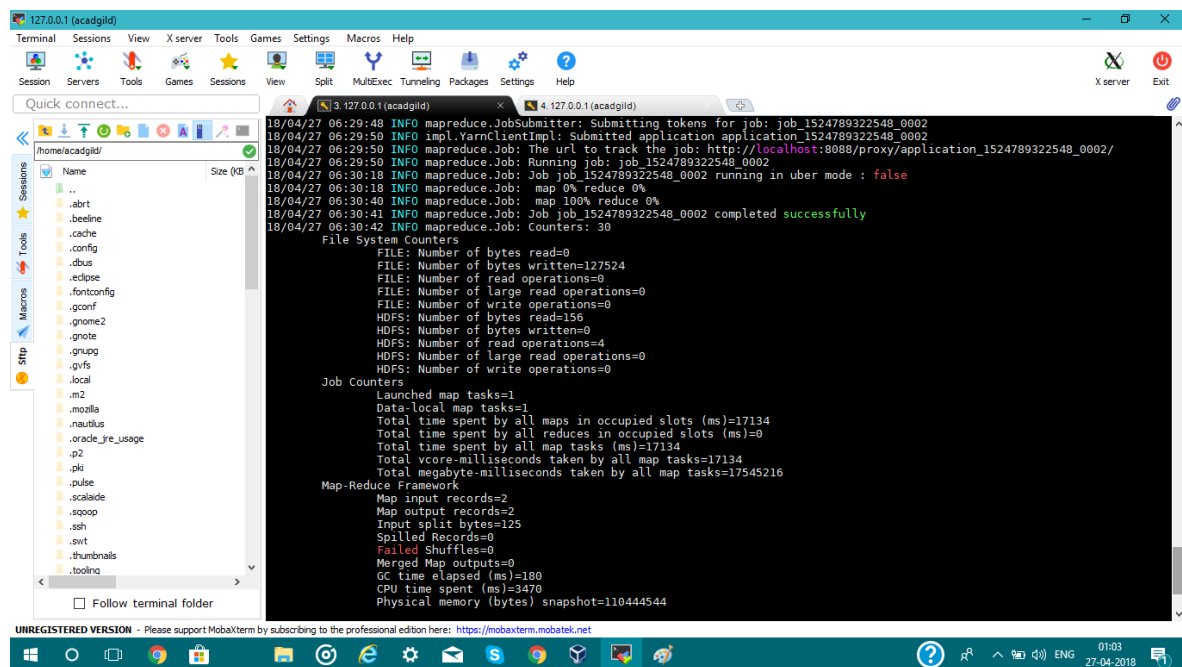
Rizvi,Syed

Command to export data from sqoopout1 folder to employee table



The screenshot shows a MobaXterm terminal window with a file explorer on the left. The terminal displays the following commands and output:

```
18/04/27 06:29:19 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/04/27 06:29:25 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/04/27 06:29:25 INFO tool.CodeGenTool: Beginning code generation
Fri Apr 27 06:29:25 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According
to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For
compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to expl
icly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
18/04/27 06:29:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
18/04/27 06:29:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
18/04/27 06:29:28 INFO orm.CompilationManager: HADOOP MAPRED HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/8a87a42ef9114d5926743de6a94e8fb/employee.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/04/27 06:29:36 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/8a87a42ef9114d5926743de6a94e8fb/emplo
yee.jar
18/04/27 06:29:36 INFO mapreduce.ExportJobBase: Beginning export of employee
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/s
lf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLogger
Binder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/04/27 06:29:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
where applicable
18/04/27 06:29:37 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/04/27 06:29:40 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.r
educe.speculative
18/04/27 06:29:40 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.
speculative
18/04/27 06:29:40 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
```



The screenshot shows a MobaXterm terminal window with a file explorer on the left. The terminal displays the following commands and output:

```
18/04/27 06:29:48 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1524789322548_0002
18/04/27 06:29:50 INFO impl.YarnClientImpl: Submitted application application_1524789322548_0002
18/04/27 06:29:50 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1524789322548_0002/
18/04/27 06:30:18 INFO mapreduce.Job: Job job_1524789322548_0002 running in uber mode : false
18/04/27 06:30:18 INFO mapreduce.Job: map 0% reduce 0%
18/04/27 06:30:40 INFO mapreduce.Job: map 100% reduce 0%
18/04/27 06:30:41 INFO mapreduce.Job: Job job_1524789322548_0002 completed successfully
18/04/27 06:30:42 INFO mapreduce.Job: Counters: 30

File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=127524
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=156
  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)=17134
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=17134
  Total vcore-milliseconds taken by all map tasks=17134
  Total megabyte-milliseconds taken by all map tasks=17545216

Map-Reduce Framework
  Map input records=2
  Map output records=2
  Input split bytes=125
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=180
  CPU time spent (ms)=3470
  Physical memory (bytes) snapshot=1104444544
```

Output

```
mysql> insert into employee values('abc','def');
Query OK, 1 row affected (0.07 sec)

mysql> select * from employee;
+-----+-----+
| lname   | fname |
+-----+-----+
| Maheshwari | Avani |
| Rizvi    | Syed  |
| abc      | def   |
+-----+-----+
3 rows in set (0.00 sec)

mysql> select * from employee;
+-----+-----+
| lname   | fname |
+-----+-----+
| Maheshwari | Avani |
| Rizvi    | Syed  |
| abc      | def   |
| Maheshwari | Avani |
| Rizvi    | Syed  |
+-----+-----+
5 rows in set (0.00 sec)
```

The last two records in this table have been inserted via export command.

Task2

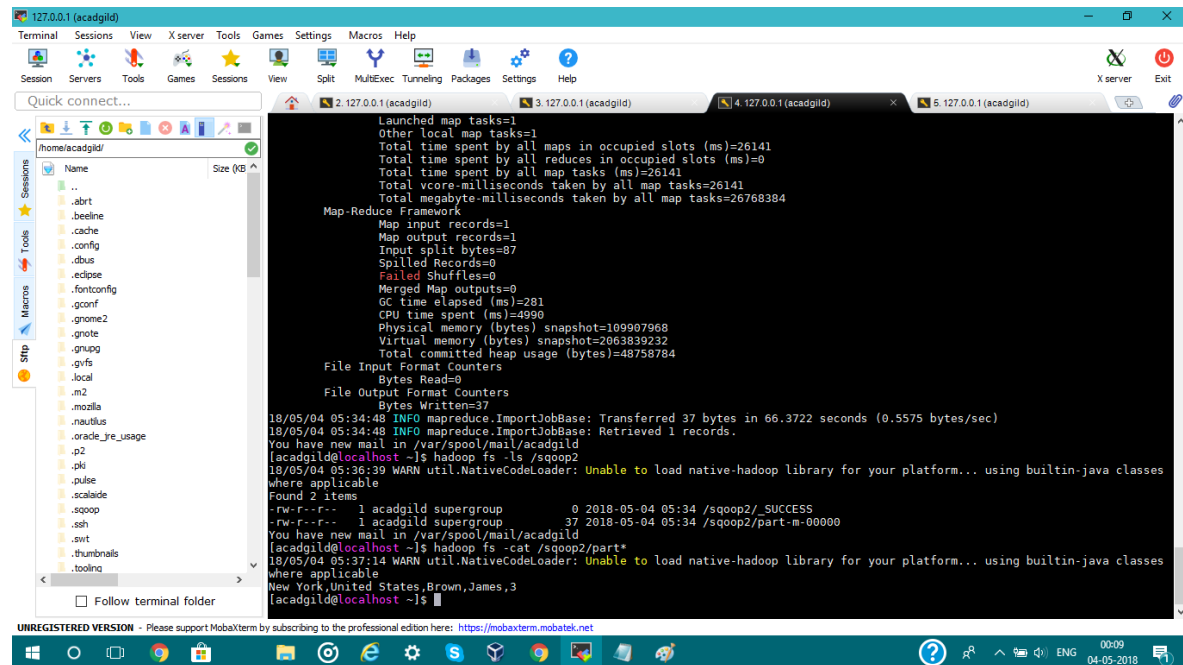
Use SQOOP tool to import data present in SQOOPOUT folder made while demo of import table with parameter person_id =3.

Command for importing data with the condition

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/simplydb --table Person --username root -P --where "person_id=3" --target-dir /sqoop2 -m 1
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6/bin_hadoop-2.0.4-alpha/./hcatalog does not exist! HCatalog jobs will fail.
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6/bin_hadoop-2.0.4-alpha/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
18/05/04 05:33:19 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/05/04 05:33:25 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/05/04 05:33:25 INFO tool.CodeGenTool: Beginning code generation
Fri May 04 05:33:26 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According
to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option isn't set. For
compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to expl
icly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
18/05/04 05:33:28 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1
18/05/04 05:33:28 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/2834a1edede6c1a4d6c116d42cc7e966/Person.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/05/04 05:33:39 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/2834a1edede6c1a4d6c116d42cc7e966/Person.jar
18/05/04 05:33:39 WARN manager.MySQLManager: It looks like you are importing from mysql.
18/05/04 05:33:39 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
18/05/04 05:33:39 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
18/05/04 05:33:39 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
18/05/04 05:33:39 INFO mapreduce.ImportJobBase: Beginning import of Person
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/s
lf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLogger
Binder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
18/05/04 05:33:40 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/05/04 05:33:42 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/05/04 05:33:42 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
```

In this command, we have used 'where' clause to import the data from table Person to the sqoop2 folder, with condition person_id =3

Output



```
Launched map tasks=1
Other local map tasks=1
Total time spent by all maps in occupied slots (ms)=26141
Total time spent by all reduces in occupied slots (ms)=0
Total time spent by all map tasks (ms)=26141
Total vcore-milliseconds taken by all map tasks=26141
Total megabyte-milliseconds taken by all map tasks=26768384

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)=281
  CPU time spent (ms)=4990
  Physical memory (bytes) snapshot=109907968
  Virtual memory (bytes) snapshot=2063839232
  Total committed heap usage (bytes)=48758784

File Input Format Counters
  Bytes Read=0

File Output Format Counters
  Bytes Written=37
18/05/04 05:34:48 INFO mapreduce.ImportJobBase: Transferred 37 bytes in 66.3722 seconds (0.5575 bytes/sec)
18/05/04 05:34:48 INFO mapreduce.ImportJobBase: Retrieved 1 records.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /sqoop2
18/05/04 05:36:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-05-04 05:34 /sqoop2/_SUCCESS
-rw-r--r-- 1 acadgild supergroup        37 2018-05-04 05:34 /sqoop2/part-m-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /sqoop2/part*
18/05/04 05:37:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes
where applicable
New York,United States,Brown,James,3
[acadgild@localhost ~]$
```

Thus, the record where the person_id=3, is inserted into the sqoop2 folder on HDFS.

