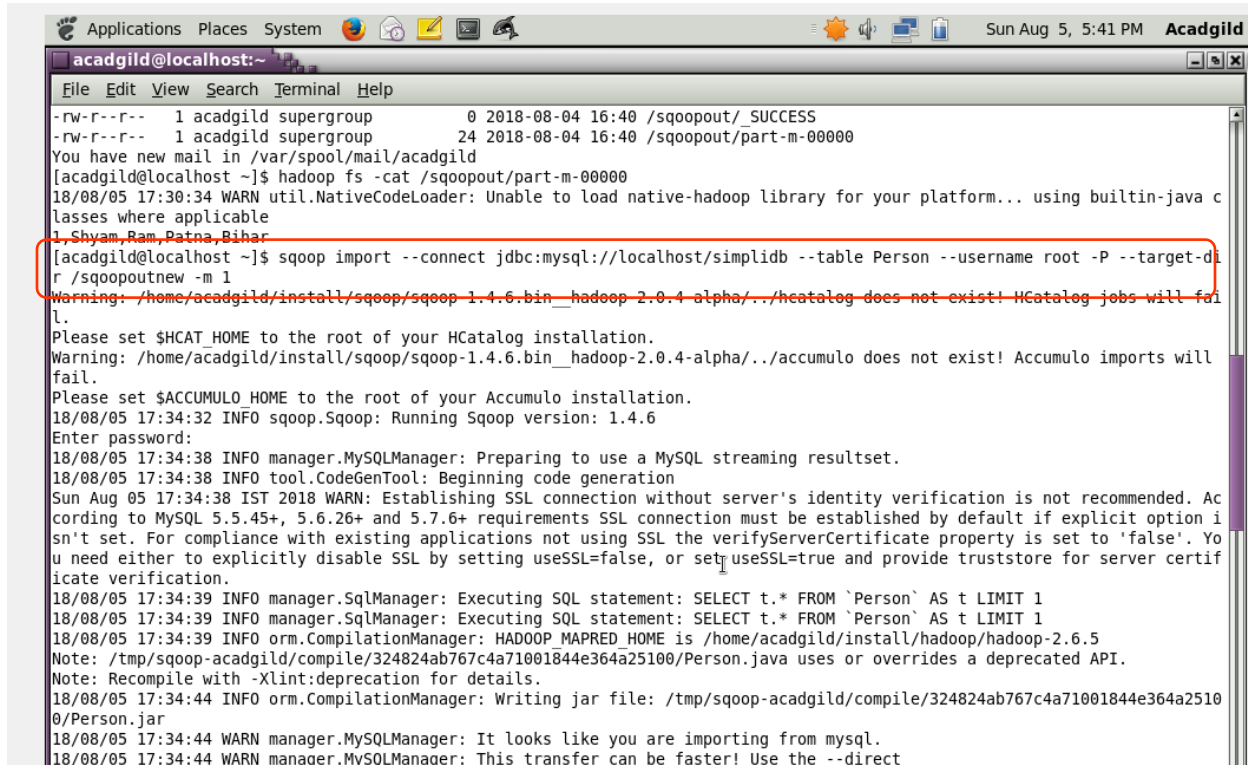


## Big Data Hadoop 'Assignment Six'

Data is imported in Sqoopoutnew folder

- Data import command



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
-rw-r--r-- 1 acadgild supergroup 0 2018-08-04 16:40 /sqoopout/ SUCCESS  
-rw-r--r-- 1 acadgild supergroup 24 2018-08-04 16:40 /sqoopout/part-m-00000  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout/part-m-00000  
18/08/05 17:30:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java c  
lasses where applicable  
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/simplidb --table Person --username root -P --target-di  
r /sqoopoutnew -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.  
Please set $HCAT_HOME to the root of your HCatalog installation.  
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.  
18/08/05 17:34:32 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6  
Enter password:  
18/08/05 17:34:38 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.  
18/08/05 17:34:38 INFO tool.CodeGenTool: Beginning code generation  
Sun Aug 05 17:34:38 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. Ac  
cording to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL connection must be established by default if explicit option i  
sn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. Yo  
u need either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certifi  
cate verification.  
18/08/05 17:34:39 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Person` AS t LIMIT 1  
18/08/05 17:34:39 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Person` AS t LIMIT 1  
18/08/05 17:34:39 INFO orm.CompilationManager: HADOOP MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5  
Note: /tmp/sqoop-acadgild/compile/324824ab767c4a71001844e364a25100/Person.java uses or overrides a deprecated API.  
Note: Recompile with -Xlint:deprecation for details.  
18/08/05 17:34:44 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/324824ab767c4a71001844e364a2510  
0/Person.jar  
18/08/05 17:34:44 WARN manager.MySQLManager: It looks like you are importing from mysql.  
18/08/05 17:34:44 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
```

- Data directory folder

```
Applications Places System Sun Aug 5, 5:43 PM Acadgild
acadmild@localhost:~
File Edit View Search Terminal Help
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)=7114
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=7114
  Total vcore-milliseconds taken by all map tasks=7114
  Total megabyte-milliseconds taken by all map tasks=7284736
Map-Reduce Framework
  Map input records=5
  Map output records=5
  Input split bytes=87
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=82
  CPU time spent (ms)=1620
  Physical memory (bytes) snapshot=105799680
  Virtual memory (bytes) snapshot=2063437824
  Total committed heap usage (bytes)=32571392
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=150
18/08/05 17:35:19 INFO mapreduce.ImportJobBase: Transferred 150 bytes in 33.3791 seconds (4.4938 bytes/sec)
18/08/05 17:35:19 INFO mapreduce.ImportJobBase: Retrieved 5 records.
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$ hadoop fs -ls /sqoopoutnew
18/08/05 17:36:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java c
lasses where applicable
Found 2 items
-rw-r--r-- 1 acadmild supergroup 0 2018-08-05 17:35 /sqoopoutnew/_SUCCESS
-rw-r--r-- 1 acadmild supergroup 150 2018-08-05 17:35 /sqoopoutnew/part-m-00000
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$
```

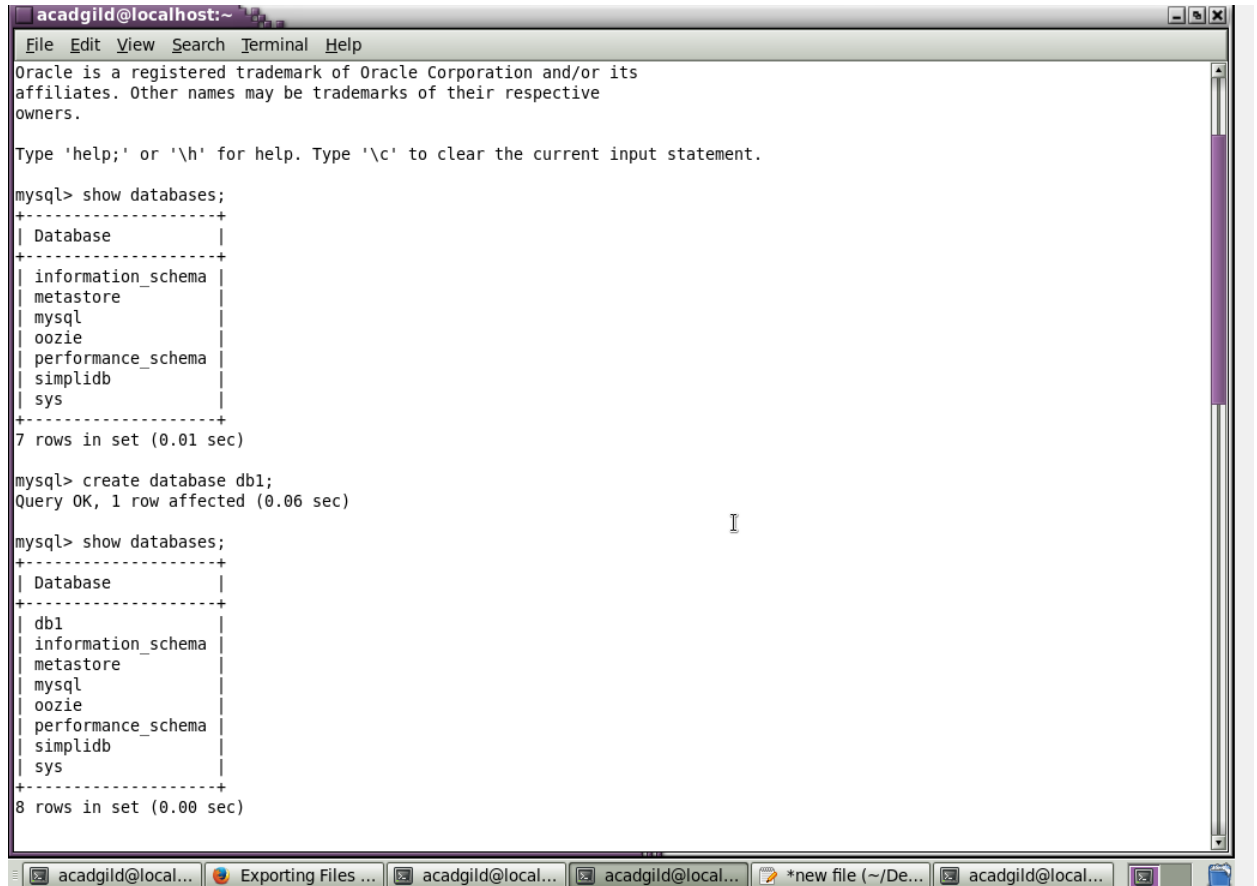
- Data inside the folder part file

```
Bytes Written=150
18/08/05 17:35:19 INFO mapreduce.ImportJobBase: Transferred 150 bytes in 33.3791 seconds (4.4938 bytes/sec)
18/08/05 17:35:19 INFO mapreduce.ImportJobBase: Retrieved 5 records.
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$ hadoop fs -ls /sqoopoutnew
18/08/05 17:36:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java c
lasses where applicable
Found 2 items
-rw-r--r-- 1 acadmild supergroup 0 2018-08-05 17:35 /sqoopoutnew/_SUCCESS
-rw-r--r-- 1 acadmild supergroup 150 2018-08-05 17:35 /sqoopoutnew/part-m-00000
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$ hadoop fs -cat /sqoopoutnew/part-m-00000
18/08/05 17:44:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java c
lasses where applicable
Patna,Bihar,Ram,Shyam,1
Whitefiled,Bangalore,Priya,Tanya,2
New York,United States,Brown,James,3
Los Angeles ,United States,Miller,Jhon,4
c,d,b,a,789
You have new mail in /var/spool/mail/acadmild
[acadmild@localhost ~]$
```

## Tasks

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

- Create database db1



The screenshot shows a terminal window titled 'acadgild@localhost:~'. The terminal displays the MySQL command prompt and the execution of two 'show databases;' queries. The first query shows 7 databases: information\_schema, metastore, mysql, oozie, performance\_schema, simplidb, and sys. The second query, after creating a new database 'db1', shows 8 databases, with 'db1' added to the list. The terminal output is as follows:

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
Oracle is a registered trademark of Oracle Corporation and/or its  
affiliates. Other names may be trademarks of their respective  
owners.  
  
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.  
  
mysql> show databases;  
+-----+  
| Database |  
+-----+  
| information_schema |  
| metastore |  
| mysql |  
| oozie |  
| performance_schema |  
| simplidb |  
| sys |  
+-----+  
7 rows in set (0.01 sec)  
  
mysql> create database db1;  
Query OK, 1 row affected (0.06 sec)  
  
mysql> show databases;  
+-----+  
| Database |  
+-----+  
| db1 |  
| information_schema |  
| metastore |  
| mysql |  
| oozie |  
| performance_schema |  
| simplidb |  
| sys |  
+-----+  
8 rows in set (0.00 sec)
```

- Show table in db1 and data in db1;
  - As this is new database and table we have no records in this.
  - We have described table to see it's structured.

```

line 1
mysql> show tables;
+-----+
| Tables_in_db1 |
+-----+
| exportsqoop   |
+-----+
1 row in set (0.00 sec)

mysql> select * from exportsqoop;
Empty set (0.00 sec)

mysql> describe exportsqoop;
ERROR 1146 (42S02): Table 'db1.exportsqoop' doesn't exist
mysql> describe exportsqoop;
+-----+-----+-----+-----+-----+-----+
| Field      | Type      | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| area_exp   | varchar(10) | YES  |     | NULL    |       |
| city_exp   | varchar(10) | YES  |     | NULL    |       |
| fname_exp  | varchar(10) | YES  |     | NULL    |       |
| lname_exp  | varchar(10) | YES  |     | NULL    |       |
| personid_exp | int(5)      | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)

```

- Write export command to import data in MYSQL exportsqoop table from HDFS
  - Sqoop export --connect jdbc:mysql://localhost/db1 --username root --P table exportsqoop --export-dir /sqoopoutnew/part-m-00000

```

Hadoop 2.6.1.1 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help

acacgild@localhost:~$ sqoop export --connect jdbc:mysql://localhost/db1 --username root --P --table exportsqoop --export-dir /sqoopoutnew/part-m-00000
Warning: /home/acacgild/install/sqoop/sqoop-1.4.6/bin_hadoop-2.0.4-alpha/./hcatalog does not exist! HCatalog jobs will fail.
Please set HCATALOG_HOME to the root of your HCatalog installation.
Warning: /home/acacgild/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.
18/08/05 18:39:36 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/08/05 18:39:41 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/08/05 18:39:41 INFO tool.CodeGenTool: Beginning code generation
Sun Aug 05 18:39:42 IST 2018 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.4+ 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 explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
18/08/05 18:39:44 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'exportsqoop' AS t LIMIT 1
18/08/05 18:39:45 INFO orm.CompilationManager: HADOOP MAPRED HOME is /home/acacgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acacgild/compile/88fa9321c1b0ba599fac5723eb39f2/exportsqoop.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/08/05 18:39:49 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acacgild/compile/88fa9321c1b0ba599fac5723eb39f2/exportsqoop.jar
18/08/05 18:39:49 INFO mapreduce.ExportJobBase: Beginning export of exportsqoop
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acacgild/install/hadoop/hadoop-2.6.5/share/hadoop/common/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acacgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.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.slf4j.impl.Log4jLoggerFactory]
18/08/05 18:39:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/05 18:39:50 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/08/05 18:39:52 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
18/08/05 18:39:52 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative

```

- Data is now loaded in Exportsqoop table;

```
3 rows in set (0.03 sec)
```

```
mysql> select * from exportsqoop;
```

area_exp	city_exp	fname_exp	lname_exp	personid_exp
Patna	Bihar	Ram	Shyam	1
Whitefiled	Bangalore	Priya	Tanya	2
c	d	b	a	789

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

```
mysql> █
```

2. Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table with parameter person\_id =3.

We can use where clause with 'IMPORT' command in Sqoop not with export whole table from that file will get export from HDFS to RDBMS.

*Have found this article saying we can use only with import command*

Unfortunately `--query` argument is not supported with Sqoop Export yet. We can use this with Sqoop import command only.

Please refer [Export control arguments](#) section for detailed information about available option in Sqoop export