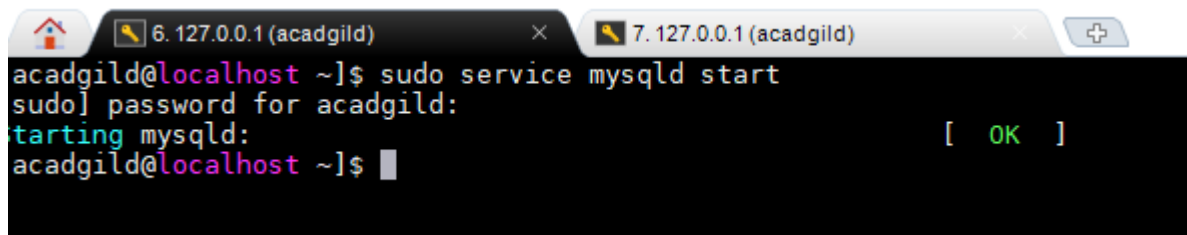


Session 6: DATA INGESTION TOOL SQOOP & INTRODUCTION TO CASE STUDY 1

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

Step: 1 Start the mysql Database

A terminal window with two tabs. The first tab is titled '6. 127.0.0.1 (acadgild)' and the second is '7. 127.0.0.1 (acadgild)'. The terminal shows the command 'sudo service mysqld start' being executed. The output shows 'starting mysqld:' followed by '[OK]' in green. The prompt 'acadgild@localhost ~\$' is visible at the bottom.

```
acadgild@localhost ~]$ sudo service mysqld start
sudo] password for acadgild:
starting mysqld: [ OK ]
acadgild@localhost ~]$
```

Step 2: Login to Mysql

```
[acadgild@localhost ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 48
Server version: 8.0.3-rc-log MySQL Community Server (GPL)

Copyright (c) 2000, 2017, Oracle and/or its affiliates. All rights reserved.

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> █
```

Step 3: See all the Database Present in DB;

```
mysql> Show databases;
+-----+
| Database |
+-----+
| information_schema |
| metastore |
| mysql |
| oozie |
| performance_schema |
| simplidb |
| sys |
+-----+
7 rows in set (0.00 sec)

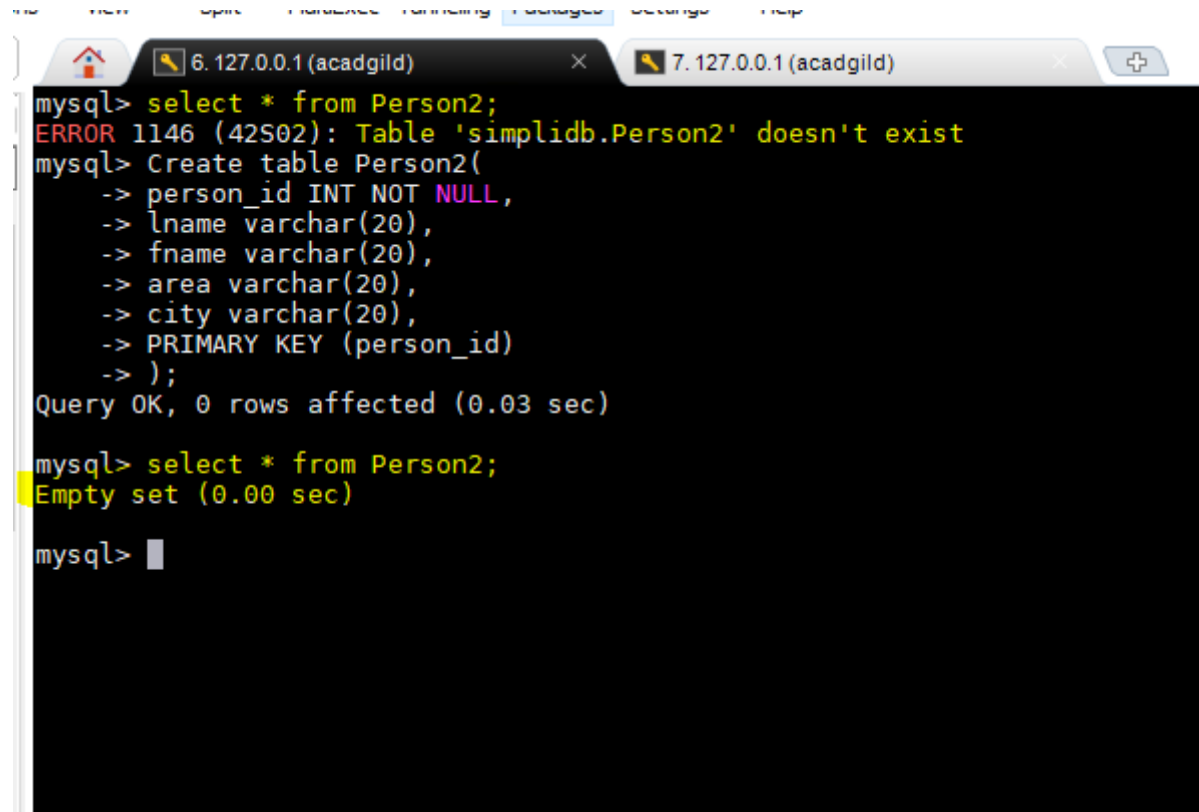
mysql> █
```

Step :4 Use Simplidb database.

```
mysql> use simplidb
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> █
```

Step 5 : Create a new table Person2.



The screenshot shows a MySQL terminal window with two tabs. The first tab is titled '6. 127.0.0.1 (acadgild)' and the second is '7. 127.0.0.1 (acadgild)'. The terminal output is as follows:

```
mysql> select * from Person2;
ERROR 1146 (42S02): Table 'simplidb.Person2' doesn't exist
mysql> Create table Person2(
-> 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 (0.03 sec)

mysql> select * from Person2;
Empty set (0.00 sec)

mysql> █
```

Step 6 : Check the sqoopoutbyid present in HDFS.

```

[acadgild@localhost ~]$ hadoop fs -ls /
18/07/06 22:30:45 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
re applicable
Found 7 items
-rw-r--r-- 1 acadgild supergroup 1056906185 2018-07-06 16:25 /file.txt
-rw-r--r-- 1 acadgild supergroup 327155712 2018-07-06 16:21 /file327.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-06 19:00 /sqoopout
drwxr-xr-x - acadgild supergroup 0 2018-07-06 20:10 /sqoopoutbyid
-rw-r--r-- 1 acadgild supergroup 252 2018-07-06 16:16 /test.txt
drwxrwx--- - acadgild supergroup 0 2018-07-06 20:39 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-07-06 07:38 /user
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /sqoopoutbyid
18/07/06 22:33:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
re applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 321 2018-07-06 20:06 /sqoopoutbyid/part-m-00000
-rw-r--r-- 1 acadgild supergroup 30 2018-07-06 20:10 /sqoopoutbyid/part-m-00001
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /sqoopoutbyid/part*
18/07/06 22:33:41 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java
re applicable
1,Manish,Anand,BTM,Bangalore
2,Manish,Anand,BTM,Bangalore
3,Manish,Anand,BTM,Bangalore
4,Manish,Anand,BTM,Bangalore
5,Manish,Anand,BTM,Bangalore
6,Manish,Anand,BTM,Bangalore
7,Manish,Anand,BTM,Bangalore
8,Manish,Anand,BTM,Bangalore
9,Manish,Anand,BTM,Bangalore
10,Manish,Anand,BTM,Bangalore
11,Manish,Anand,BTM,Bangalore
12,Manish,Anand,BTM,Bangalore
[acadgild@localhost ~]$ █

```

Step 7: Use Export query to fetch the data to mysql table Person2.

```

[acadgild@localhost ~]$
[acadgild@localhost ~]$ sqoop export -m 1 --connect jdbc:mysql://localhost/simplidb --username root --password Root@123 --table Person
2 --export-dir /sqoopoutbyid/
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.
18/07/06 22:35:57 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/07/06 22:35:57 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/07/06 22:35:58 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/07/06 22:35:58 INFO tool.CodeGenTool: Beginning code generation
Fri Jul 06 22:35:59 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 compli
ance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need either to explicitly di
sable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
18/07/06 22:36:02 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person2' AS t LIMIT 1
18/07/06 22:36:02 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person2' AS t LIMIT 1
18/07/06 22:36:02 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/2e68a075a59269a0a4b3f524ee866764/Person2.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/07/06 22:36:09 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/2e68a075a59269a0a4b3f524ee866764/Person2.
jar
18/07/06 22:36:09 INFO mapreduce.ExportJobBase: Beginning export of Person2

```

```

Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=191
CPU time spent (ms)=3350
Physical memory (bytes) snapshot=119177216
Virtual memory (bytes) snapshot=2061332480
Total committed heap usage (bytes)=62980096

File Input Format Counters
Bytes Read=0
File Output Format Counters
Bytes Written=0
18/07/06 22:37:12 INFO mapreduce.ExportJobBase: Transferred 549 bytes in 57.3675 seconds (9.5699 bytes/sec)
18/07/06 22:37:13 INFO mapreduce.ExportJobBase: Exported 12 records.
You have new mail in /var/spool/mail/acadgild

```

Step 8: Check all the exported data in mysql table Person2.

```
mysql> select * from Person2;
```

person_id	lname	fname	area	city
1	Manish	Anand	BTM	Bangalore
2	Manish	Anand	BTM	Bangalore
3	Manish	Anand	BTM	Bangalore
4	Manish	Anand	BTM	Bangalore
5	Manish	Anand	BTM	Bangalore
6	Manish	Anand	BTM	Bangalore
7	Manish	Anand	BTM	Bangalore
8	Manish	Anand	BTM	Bangalore
9	Manish	Anand	BTM	Bangalore
10	Manish	Anand	BTM	Bangalore
11	Manish	Anand	BTM	Bangalore
12	Manish	Anand	BTM	Bangalore

```
12 rows in set (0.00 sec)

mysql> █
```

Task2 Use Sqoop tool to export data present in **SQOOPOUTBYID** folder made while demo of Import table with parameter person_id =3.

Step: 1 Check Person2 table is empty or not .

```
mysql> truncate Person2;
Query OK, 0 rows affected (0.03 sec)

mysql> select * from Person2;
Empty set (0.00 sec)

mysql>
```

Step 2 : Check Sqoopoutbyid folder present in hadoop .

```
[acadgild@localhost ~]$ hadoop fs -ls /sqoopoutbyid
18/07/18 06:39:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using bui
ltin-java classes where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup          0 2018-07-18 06:39 /sqoopoutbyid/_SUCCESS
-rw-r--r-- 1 acadgild supergroup       29 2018-07-18 06:39 /sqoopoutbyid/part-m-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Step : 3 Check the Export Data present in Sqoopoutbyid folder.

```
[acadgild@localhost ~]$ hadoop fs -cat /sqoopoutbyid/part-m-00000
18/07/18 06:43:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using bui
ltin-java classes where applicable
BTM,Bangalore,Anand,Manish,3
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

Step :4 Now we will write export command to fetch data from sqoopoutbyid folder to our mysql PERSON2 table.

```

[acadgild@localhost ~]$ sqoop export -m 1 --connect jdbc:mysql://localhost/simplidb --username root --password Root@123 --table Person2 --export-dir /sqoopoutbyid/
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.
18/07/18 06:47:08 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/07/18 06:47:08 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/07/18 06:47:09 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/07/18 06:47:09 INFO tool.CodeGenTool: Beginning code generation
Wed Jul 18 06:47:09 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

```

```

Total vcore-milliseconds taken by all map tasks=7033
Total megabyte-milliseconds taken by all map tasks=7222272
Map-Reduce Framework
  Map input records=1
  Map output records=1
  Input split bytes=128
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=91
  CPU time spent (ms)=1880
  Physical memory (bytes) snapshot=119169024
  Virtual memory (bytes) snapshot=2061332480
  Total committed heap usage (bytes)=62980096
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
18/07/18 06:47:44 INFO mapreduce.ExportJobBase: Transferred 160 bytes in 27.9652 seconds (5.7214 bytes/sec)
18/07/18 06:47:44 INFO mapreduce.ExportJobBase: Exported 1 records.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$

```

Step 5: Now Check the Exported data in PERSON2 Table;

```

mysql> select * from Person2;
+-----+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+
| 3 | Manish | Anand | BTM | Bangalore |
+-----+-----+-----+-----+
1 row in set (0.00 sec)

mysql>

```

Thanks,

Shishir