

## ASSIGNMENT SQOOP

### Problem Statement

#### Task 1

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

#### Task2

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

### SOLUTION:

TASK1: Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table

EXPLANATION: LOGIN TO MYSQL CREDENTIALS AND CREATE TABLE PERSON AS MENTIONED IN THE GIVEN EXAMPLE DEMO AND THEN IMPORT THE DATA INTO HDFS USING SQOOP COMMANDS AFTER THAT WE NEED TO INSERT VALUES AS GIVEN IN EXAMPLE DEMO. AFTER DOING ALL THESE INSTRUCTIONS NOW WE HAVE TO USE EXPORT COMMAND TO TRANSFER FILES INSIDE THE HDFS INTO MYSQL DATABASE. FOR THIS WE NEED TO FOLLOW THE BELOW STEPS.

A) FIRST LOGIN INTO MYSQL-->

COMMAND: sudo services mysqld start

EXPLANATION: AFTER TYPING THE ABOVE COMMAND. ENTER PASSWORD FOR ACADGILD AS "acadgild".

COMMAND: mysql -u root -p

EXPLANATION: AFTER TYPING THE ABOVE COMMAND WE HAVE TO ENTER PASSWORD FOR MYSQL AS "Root@123". BY DOING THIS U WILL BE LOGGED INTO MYSQL.

### SOLUTION:

```
[acadgild@localhost ~]$ sudo service mysqld start
```

```
[sudo] password for acadgild:
```

```
Starting mysqld: [ OK ]
```

```
[acadgild@localhost ~]$ mysql -u root -p
```

```
Enter password:
```

```
Welcome to the MySQL monitor. Commands end with ; or \g.
```

```
Your MySQL connection id is 19
```

```
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.

B) NOW CREATE A DATABASE FOR THE NEW EXPORTED DATA AND USE THE SAME CREATED DATABASE:-

COMMAND: create database export

EXPLANATION: THIS WILL CREATE A DATABASE INSIDE MYSQL

SOLUTION:

```
mysql> create database export;
Query OK, 1 row affected (0.01 sec)
```

COMMAND: use export

SOLUTION:

```
mysql> use export;
Database changed
```

EXPLANATION: THIS COMMAND WILL USE THE CREATED DATABASE.

C) CREATE A TABLE FOR EXPORTING THE DATA FROM HDFS TO MYSQL WHICH WAS IMPORTED USING SQOOP COMMAND INTO HDFS FOR THE CREATED TABLE "Person" BEFORE.

COMMAND: Create table Exportedfile( person\_id INT NOT NULL, lname varchar(20), fname varchar(20), area varchar(20), city varchar(20), PRIMARY KEY(person\_id));

EXPLANATION: WE HAD CREATED A TABLE "Exportedfile" WITH THE SAME SCHEMA FOR THE IMPORTED DATA INSIDE HDFS.

SOLUTION:

```
mysql> Create table Exportedfile( person_id INT NOT NULL, lname
varchar(20), fname varchar(20), area varchar(20), city varchar(20) );
Query OK, 0 rows affected (0.14 sec)
```

EXPLANATION: WE CAN SEE THE TABLE CREATED BY USING BELOW QUERY AND ALSO CAN SEE THE SCHEMA INFORMATION.

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

```
mysql> describe Exportedfile;
```

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)
```

D) NOW GO TO HDFS, AS THE IMPORTED TABLE FROM MYSQL IS STORED IN LOCAL DIRECTORY BUT NOT THE HDFS. WE NEED TO COPY THE SAME INSIDE THE HDFS SO FOR DOING THAT WE NEED TO CREATE A DIRECTORY AND THE USING COPY COMMAND WE HAVE TO COPY THE IMPORTED TABLE FROM MYSQL IN LOCAL DIRECTORY INTO HDFS:-

COMMAND: `hadoop fs -mkdir /user/acadgild/mysqldata`

SOLUTION:

```
acadgild@localhost ~]$ hadoop fs -mkdir /user/acadgild/mysqldata
19/01/07 16:05:37 WARN util.NativeCodeLoader: Unable to load native-
hadoop library for your platform... using builtin-java classes where
applicable
```

COMMAND: `hadoop fs -cp /sqoopout/part-m-00000 /user/acadgild/mysqldata`

EXPLANATION: AS FOR THE IMPORTED TABLE, THE TABLE IS TRANSFERED TO THE LOCATION "/sqoopout". SO HERE WE ARE COPYING THE DATA FROM THE SAME LOCATION INTO THE DIRECTORY CREATED IN HDFS "/user/acadgild/mysqldata". AFTER WE CAN SEE THAT THE FILE IS BEEN COPIED INTO THE HDFS LOCATION

SOLUTION:

```
[acadgild@localhost ~]$ hadoop fs -cp /sqoopout/part-m-00000
/user/acadgild/mysqldata
19/01/07 16:06:22 WARN util.NativeCodeLoader: Unable to load native-
hadoop library for your platform... using builtin-java classes where
applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -ls /user/acadgild/mysqldata
19/01/07 16:07:07 WARN util.NativeCodeLoader: Unable to load native-
hadoop library for your platform... using builtin-java classes where
applicable
Found 1 items
-rw-r--r--    1 acadgild supergroup          24 2019-01-07 15:49
/user/acadgild/mysqldata/part-m-00000
```

E) NOW USING THE SQOOP EXPORT COMMAND WE NEED TO EXPORT THE DATA INSIDE HDFS INTO MYSQL DATABASE:-

COMMAND: `sqoop export --connect jdbc:mysql://localhost/export --username root -P --table Exportedfile --export-dir /user/acadgild/mysqldata -m1;`

EXPLANATION: SQOOP CALLS THE JDBC DRIVER WRITTEN IN THE "-connect" STATEMENT FROM THE LOCATION WHERE SQOOP IS INSTALLED. THE "-username" and "-p" PASSWORD OPTIONS USED TO AUTHENTICATE THE USER AND SQOOP. THE "-table" ARGUMENT DEFINES THE MYSQL TABLE NAME WHICH WILL RECEIVE DATA FROM HDFS. THIS DATA WILL BE FETCHED FROM THE DIRECTORY WHERE THE FILE IS LOCATED INSIDE THE HDFS. THUS WHEN THE "export" STATEMENT IS EXECUTED, THEN INITIATES AND CREATES INSERT STATEMENTS IN MYSQL. THUS IT WILL READ THE DATA PRESENT INSIDE THE DIRECTORY TILL END LINE AND EXPORTS THE DATA TO THE TABLE CREATED IN MYSQL

SOLUTION:

```
[acadgild@localhost ~]$ sqoop export --connect
jdbc:mysql://localhost/export --username root -P --table Exportedfile --
export-dir /user/acadgild/mysqlldata -ml;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.
19/01/07 15:58:40 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
19/01/07 15:58:47 INFO manager.MySQLManager: Preparing to use a MySQL
streaming resultset.
19/01/07 15:58:47 INFO tool.CodeGenTool: Beginning code generation
Mon Jan 07 15:58:48 IST 2019 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 explicitly disable SSL by setting
useSSL=false, or set useSSL=true and provide truststore for server
certificate verification.
19/01/07 15:58:51 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM `Exportedfile` AS t LIMIT 1
19/01/07 15:58:51 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM `Exportedfile` AS t LIMIT 1
19/01/07 15:58:51 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is
/home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-
acadgild/compile/9c5a05744031b5c0f2a1022a7fe502b4/Exportedfile.java uses
or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
19/01/07 15:58:59 INFO orm.CompilationManager: Writing jar file:
/tmp/sqoop-
acadgild/compile/9c5a05744031b5c0f2a1022a7fe502b4/Exportedfile.jar
19/01/07 15:58:59 INFO mapreduce.ExportJobBase: Beginning export of
Exportedfile
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/slf4j/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/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]
19/01/07 15:59:00 WARN util.NativeCodeLoader: Unable to load native-
hadoop library for your platform... using builtin-java classes where
applicable
19/01/07 15:59:00 INFO Configuration.deprecation: mapred.jar is
deprecated. Instead, use mapreduce.job.jar
```

19/01/07 15:59:04 INFO Configuration.deprecation:  
mapred.reduce.tasks.speculative.execution is deprecated. Instead, use  
mapreduce.reduce.speculative  
19/01/07 15:59:04 INFO Configuration.deprecation:  
mapred.map.tasks.speculative.execution is deprecated. Instead, use  
mapreduce.map.speculative  
19/01/07 15:59:04 INFO Configuration.deprecation: mapred.map.tasks is  
deprecated. Instead, use mapreduce.job.maps  
19/01/07 15:59:05 INFO client.RMPProxy: Connecting to ResourceManager at  
localhost/127.0.0.1:8032  
19/01/07 15:59:10 INFO input.FileInputFormat: Total input paths to  
process : 1  
19/01/07 15:59:10 INFO input.FileInputFormat: Total input paths to  
process : 1  
19/01/07 15:59:10 INFO mapreduce.JobSubmitter: number of splits:1  
19/01/07 15:59:10 INFO Configuration.deprecation:  
mapred.map.tasks.speculative.execution is deprecated. Instead, use  
mapreduce.map.speculative  
19/01/07 15:59:11 INFO mapreduce.JobSubmitter: Submitting tokens for job:  
job\_1546852017617\_0004  
19/01/07 15:59:12 INFO impl.YarnClientImpl: Submitted application  
application\_1546852017617\_0004  
19/01/07 15:59:13 INFO mapreduce.Job: The url to track the job:  
[http://localhost:8088/proxy/application\\_1546852017617\\_0004/](http://localhost:8088/proxy/application_1546852017617_0004/)  
19/01/07 15:59:13 INFO mapreduce.Job: Running job: job\_1546852017617\_0004  
19/01/07 15:59:39 INFO mapreduce.Job: Job job\_1546852017617\_0004 running  
in uber mode : false  
19/01/07 15:59:40 INFO mapreduce.Job: map 0% reduce 0%  
19/01/07 16:00:06 INFO mapreduce.Job: map 100% reduce 0%  
19/01/07 16:00:07 INFO mapreduce.Job: Job job\_1546852017617\_0004  
completed successfully  
19/01/07 16:00:08 INFO mapreduce.Job: Counters: 30

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=127572  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=166  
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)=22487  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=22487  
Total vcore-milliseconds taken by all map tasks=22487  
Total megabyte-milliseconds taken by all map tasks=23026688

Map-Reduce Framework

Map input records=1  
Map output records=1

```

    Input split bytes=139
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=423
    CPU time spent (ms)=3220
    Physical memory (bytes) snapshot=97763328
    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
19/01/07 16:00:08 INFO mapreduce.ExportJobBase: Transferred 166 bytes in
63.5013 seconds (2.6141 bytes/sec)
19/01/07 16:00:08 INFO mapreduce.ExportJobBase: Exported 1 records.
You have new mail in /var/spool/mail/acadgild

```

E) CHECK INSIDE THE MYSQL DATABASE THAT WHETHER THE RECORDS ARE EXPORTED OR NOT

```
COMMAND: mysql> select * from Exportedfile;
```

EXPLANATION: BY USING THE ABOVE CODE WE CAN SEE THE RECORDS ARE EXPORTED INTO THE TABLE "Exportedfile".

SOLUTION:

```
mysql> select * from Exportedfile;
+-----+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+-----+
|          1 | Shyam | Ram   | Patna | Bihar |
+-----+-----+-----+-----+-----+
1 row in set (0.01 sec)
```

SOLUTION:

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

EXPLANATION: WE USE THE SAME DATABASE 'export' FOR THIS TASK TOO AND ALSO THE SAME TABLE "Exportedfile" TO EXPORT DATA INTO. I HAVE USED SAME STEPS FOR IMPORT COMMAND IN SQOOP AND SAVED THE IMPORTED DATA WHICH INTO THE LOCATION "/sqoopoutput/myquery". THEN I USED THE SAME STEP AS IN STEP(D) AND COPIED THE FILE INTO HDFS LOCATION IN "/user/acadgild/filemq" AND THEN CONTINUED TO EXPORT THE DATA INTO MYSQL.

```
COMMAND: sqoop export --connect jdbc:mysql://localhost/export --username
root -P --table Exportedfile --export-dir /user/acadgild/filemq -m1;
```

## SOLUTION:

```
[acadgild@localhost ~]$ sqoop export --connect
jdbc:mysql://localhost/export --username root -P --table Exportedfile --
export-dir /user/acadgild/filemq -m1;
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.
19/01/07 20:20:26 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
19/01/07 20:20:31 INFO manager.MySQLManager: Preparing to use a MySQL
streaming resultset.
19/01/07 20:20:31 INFO tool.CodeGenTool: Beginning code generation
Mon Jan 07 20:20:32 IST 2019 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 explicitly disable SSL by setting
useSSL=false, or set useSSL=true and provide truststore for server
certificate verification.
19/01/07 20:20:36 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM `Exportedfile` AS t LIMIT 1
19/01/07 20:20:36 INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM `Exportedfile` AS t LIMIT 1
19/01/07 20:20:36 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is
/home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-
acadgild/compile/3fe4e81da1d970c5bde83bf4dda7f11d/Exportedfile.java uses
or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
19/01/07 20:20:45 INFO orm.CompilationManager: Writing jar file:
/tmp/sqoop-
acadgild/compile/3fe4e81da1d970c5bde83bf4dda7f11d/Exportedfile.jar
19/01/07 20:20:45 INFO mapreduce.ExportJobBase: Beginning export of
Exportedfile
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/slf4j/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/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]
19/01/07 20:20:46 WARN util.NativeCodeLoader: Unable to load native-
hadoop library for your platform... using builtin-java classes where
applicable
19/01/07 20:20:46 INFO Configuration.deprecation: mapred.jar is
deprecated. Instead, use mapreduce.job.jar
```

19/01/07 20:20:49 INFO Configuration.deprecation:  
mapred.reduce.tasks.speculative.execution is deprecated. Instead, use  
mapreduce.reduce.speculative  
19/01/07 20:20:49 INFO Configuration.deprecation:  
mapred.map.tasks.speculative.execution is deprecated. Instead, use  
mapreduce.map.speculative  
19/01/07 20:20:49 INFO Configuration.deprecation: mapred.map.tasks is  
deprecated. Instead, use mapreduce.job.maps  
19/01/07 20:20:49 INFO client.RMPProxy: Connecting to ResourceManager at  
localhost/127.0.0.1:8032  
19/01/07 20:20:54 INFO input.FileInputFormat: Total input paths to  
process : 1  
19/01/07 20:20:54 INFO input.FileInputFormat: Total input paths to  
process : 1  
19/01/07 20:20:54 INFO mapreduce.JobSubmitter: number of splits:1  
19/01/07 20:20:54 INFO Configuration.deprecation:  
mapred.map.tasks.speculative.execution is deprecated. Instead, use  
mapreduce.map.speculative  
19/01/07 20:20:55 INFO mapreduce.JobSubmitter: Submitting tokens for job:  
job\_1546852017617\_0011  
19/01/07 20:20:56 INFO impl.YarnClientImpl: Submitted application  
application\_1546852017617\_0011  
19/01/07 20:20:56 INFO mapreduce.Job: The url to track the job:  
[http://localhost:8088/proxy/application\\_1546852017617\\_0011/](http://localhost:8088/proxy/application_1546852017617_0011/)  
19/01/07 20:20:56 INFO mapreduce.Job: Running job: job\_1546852017617\_0011  
19/01/07 20:21:18 INFO mapreduce.Job: Job job\_1546852017617\_0011 running  
in uber mode : false  
19/01/07 20:21:18 INFO mapreduce.Job: map 0% reduce 0%  
19/01/07 20:21:35 INFO mapreduce.Job: map 100% reduce 0%  
19/01/07 20:21:36 INFO mapreduce.Job: Job job\_1546852017617\_0011  
completed successfully  
19/01/07 20:21:36 INFO mapreduce.Job: Counters: 30

File System Counters

FILE: Number of bytes read=0  
FILE: Number of bytes written=127569  
FILE: Number of read operations=0  
FILE: Number of large read operations=0  
FILE: Number of write operations=0  
HDFS: Number of bytes read=176  
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)=13585  
Total time spent by all reduces in occupied slots (ms)=0  
Total time spent by all map tasks (ms)=13585  
Total vcore-milliseconds taken by all map tasks=13585  
Total megabyte-milliseconds taken by all map tasks=13911040

Map-Reduce Framework

Map input records=1  
Map output records=1



```

Input split bytes=136
Spilled Records=0
Failed Shuffles=0
Merged Map outputs=0
GC time elapsed (ms)=181
CPU time spent (ms)=3360
Physical memory (bytes) snapshot=105709568
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
19/01/07 20:21:36 INFO mapreduce.ExportJobBase: Transferred 176 bytes in
47.5355 seconds (3.7025 bytes/sec)
19/01/07 20:21:36 INFO mapreduce.ExportJobBase: Exported 1 records.
You have new mail in /var/spool/mail/acadgild

```

```

COMMAND: select * from Exportedfile;

```

EXPLANATION: USING THE ABOVE COMMAND WE CAN SEE THE RECORD HAS BEEN EXPORTED TO THE TABLE.

SOLUTION:

```

mysql> select * from Exportedfile;

```

```

+-----+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+-----+
|          1 | Shyam | Ram   | Patna | Bihar |
|          3 | James | Brown | New York | United states |
+-----+-----+-----+-----+-----+
2 rows in set (0.00 sec)

```