

Problem Statement

- Transfer data between Mysql and HDFS (Import and Export) using Sqoop.

1. Make sure all the hadoop daemons are started,

```
[acadgild@localhost ~]$ jps
16081 Jps
5443 RunJar
2599 DataNode
2792 SecondaryNameNode
3243 NodeManager
2492 NameNode
3134 ResourceManager
```

2. Start the mysql shell,

```
[acadgild@localhost ~]$ sudo service mysqld start
[sudo] password for acadgild:
Starting mysqld: [ OK ]
[acadgild@localhost ~]$ mysql
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 2
Server version: 5.1.73 Source distribution

Copyright (c) 2000, 2013, 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> SHOW Databases;
```

Importing a Table from RDBMS to HDFS

1. Create Database db; (we are using an existing DB and Table)
2. SHOW TABLES:

```
SHOW TABLES' at line 1
mysql> SHOW TABLES;
+-----+
| Tables_in_db |
+-----+
| employee      |
+-----+
1 row in set (0.00 sec)
```

3. Data available in the table 'employee'

Select * From employee;

```
mysql> Select * From employee;
+-----+-----+-----+
| emp_id | emp_name | emp_sal |
+-----+-----+-----+
| 1      | kk       | 50000   |
| 4      | onk      | 6000    |
| 7      | mj       | 70000   |
+-----+-----+-----+
3 rows in set (0.01 sec)
```

4. DESCRIBE employee;

```
mysql>
mysql> DESCRIBE employee;
+-----+-----+-----+-----+-----+-----+
| Field      | Type          | Null | Key | Default | Extra          |
+-----+-----+-----+-----+-----+-----+
| emp_id     | int(11)       | NO   | PRI | NULL    | auto_increment |
| emp_name   | varchar(100)  | YES  |     | NULL    |                |
| emp_sal    | int(11)       | YES  |     | NULL    |                |
+-----+-----+-----+-----+-----+-----+
3 rows in set (0.01 sec)
```

5. Insert some more values into the existing table,

INSERT INTO employee(emp_id,emp_name,emp_sal) VALUES (2,'Abu',67000), (3,'Kedar',12000),(5,'Rahu',70000);

```
mysql> INSERT INTO employee(emp_id,emp_name,emp_sal) VALUES (2,'Abu',67000), (3,'Kedar',12000),(5,'Rahu',70000);
Query OK, 3 rows affected (0.00 sec)
Records: 3 Duplicates: 0 Warnings: 0

mysql>
mysql> Select * From employee;
+-----+-----+-----+
| emp_id | emp_name | emp_sal |
+-----+-----+-----+
| 1      | kk       | 50000   |
| 4      | onk      | 6000    |
| 7      | mj       | 70000   |
| 2      | Abu      | 67000   |
| 3      | Kedar    | 12000   |
| 5      | Rahu     | 70000   |
+-----+-----+-----+
6 rows in set (0.00 sec)
```

The above is the data available in the table 'employee' in the database 'db'

Sqoop import step,

sqoop import --connect jdbc:mysql://localhost/db --username root --password acadgild --table employee -m1

```
rw-r--r-- 1 acadgild supergroup 300 2017-09-24 14:16 /user/acadgild/hadoop/word-count.txt
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/db --username root --password acadgild --table employee -m1
Warning: /home/acadgild/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/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.
Warning: /home/acadgild/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-12-04 17:44:26,057 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-12-04 17:44:26,089 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2017-12-04 17:44:26,431 INFO [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-12-04 17:44:26,431 INFO [main] tool.CodeGenTool: Beginning code generation
```

Expected Output

```
2017-12-04 17:45:22,939 INFO [main] mapreduce.ImportJobBase: Retrieved 6 records.
[acadgild@localhost ~]$ hadoop fs -ls
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled
stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/12/04 17:45:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applic
able
Found 2 items
drwxr-xr-x - acadgild supergroup 0 2017-12-04 17:45 employee
drwxr-xr-x - acadgild supergroup 0 2017-11-30 18:05 hadoop
[acadgild@localhost ~]$ hadoop fs -cat /user/acadgild/employee/*
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled
stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/12/04 17:45:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applic
able
1,kk,50000
4,onk,6000
7,mj,70000
2,Abu,67000
3,Kedar,12000
5,Rahu,70000
[acadgild@localhost ~]$
```

The data from employee table in the mysql has been imported using sqoop and the output is shown above in the screen shot.

1. We are using the existing the database 'db1'

```
mysql> Show Databases;
+-----+
| Database |
+-----+
| information_schema |
| db |
| db1 |
| metastore |
| mysql |
+-----+
5 rows in set (0.02 sec)

mysql>
mysql> Use db1;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A
```

2. Create a table 'acad' in the 'db1' like below,

```
mysql>
mysql> CREATE TABLE acad
-> (emp_id int(2),
-> emp_name varchar(10),
-> emp_sal int(10),
-> date date);
Query OK, 0 rows affected (0.01 sec)

mysql>
mysql> SHOW TABLES;
+-----+
| Tables_in_db1 |
+-----+
| acad |
| company |
+-----+
2 rows in set (0.00 sec)
```

3. Table 'acad' Format,

```
mysql>
mysql> DESCRIBE acad;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| emp_id | int(2) | YES | | NULL | |
| emp_name | varchar(10) | YES | | NULL | |
| emp_sal | int(10) | YES | | NULL | |
| date | date | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+
4 rows in set (0.00 sec)
```

4. The data in this table are yet to be imported from the HDFS, hence it will be shown empty set,

```
mysql>
mysql> Select * From acad;
Empty set (0.00 sec)
```

5. We are going to use the data set 'emp_sqoop.txt', the 'emp_sqoop.txt' should be transferred from the local system to the HDFS first and read the data using cat command,

```
hadoop fs -put /home/acadgild/hadoop/emp_sqoop.txt /user/acadgild/hadoop
```

```
hadoop fs -cat /user/acadgild/hadoop/emp_sqoop.txt
```

```
[acadgild@localhost hadoop]$ hadoop fs -put /home/acadgild/hadoop/emp_sqoop.txt /user/acadgild/hadoop
p
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/11/30 18:05:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
... using builtin-java classes where applicable
[acadgild@localhost hadoop]$ hadoop fs -cat /user/acadgild/hadoop/emp_sqoop.txt
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/11/30 18:05:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform
... using builtin-java classes where applicable
1,Ashish,50000,2011-03-21
2,Abarajithan,75000,2013-03-25
3,Kedar,60000,2012-03-11
4,Sreejith,80000,2010-03-17
5,Girish,45000,2014-03-15[acadgild@localhost hadoop]$
[acadgild@localhost hadoop]$
```

Sqoop Export step

```
sqoop export --connect jdbc:mysql://localhost/db1 --username root --password acadgild --table acad --export-dir /user/acadgild/hadoop/emp_sqoop.txt
```

```
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username root --password acadgild --table acad --export-dir /user/a
cadgild/hadoop/emp_sqoop.txt
Warning: /home/acadgild/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/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.
Warning: /home/acadgild/sqoop-1.4.6.bin_hadoop-2.0.4-alpha/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-12-04 18:18:45,888 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-12-04 18:18:45,920 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
```

Expected output

```
mysql>
mysql> Select * From acad;
+-----+-----+-----+-----+
| emp_id | emp_name | emp_sal | date       |
+-----+-----+-----+-----+
| 3      | Kedar    | 60000   | 2012-03-11 |
| 1      | Ashish   | 50000   | 2011-03-21 |
| 2      | Abarajitha | 75000   | 2013-03-25 |
| 5      | Girish   | 45000   | 2014-03-15 |
| 4      | Sreejith | 80000   | 2010-03-17 |
+-----+-----+-----+-----+
5 rows in set (0.00 sec)
```

The data from the 'emp_sqoop.txt' which was in the HDFS has been exported to the table 'acad' to the database 'db1'.

A Sqoop command is used to transfer selected columns from MySQL to Hive

1. **CREATE DATABASE company;**

```
mysql> CREATE DATABASE company;
Query OK, 1 row affected (0.00 sec)

mysql>
mysql> SHOW Databases;
+-----+
| Database |
+-----+
| information_schema |
| SQOOP_SAMPLE |
| company |
| db |
| db1 |
| metastore |
| mysql |
+-----+
7 rows in set (0.00 sec)
```

2. Create a table 'company1'

CREATE TABLE company1 (EmpID int(11),EMPNAME varchar(20),Designation varchar(15),DOJ varchar(15),city varchar(15),country varchar(15));

```
mysql> CREATE TABLE company1 (EmpID int(11),EMPNAME varchar(20),Designation varchar(15),DOJ Date,city varchar(15),country varchar(15));
Query OK, 0 rows affected (0.01 sec)

mysql> DESCRIBE TABLE company1;
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 'TABLE company1' at line 1
mysql> DESCRIBE company1;
+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+
| EmpID | int(11) | YES | | NULL | |
| EMPNAME | varchar(20) | YES | | NULL | |
| Designation | varchar(15) | YES | | NULL | |
| DOJ | date | YES | | NULL | |
| city | varchar(15) | YES | | NULL | |
| country | varchar(15) | YES | | NULL | |
+-----+
6 rows in set (0.00 sec)
```

3. Add some data into the newly created table 'company1'

INSERT INTO company1(EmpID,EMPNAME,Designation,DOJ,city,country) VALUES (1,'Barack Eilam','CEO','2007-01-01','Tel Aviv','Isreal'),(2,'Barry Cooper','Vice President','2009-05-09','Denver','US'),(3,'Aditya Tiwary','Director FLS','2015-07-15','Pune','India'),(4,'Abarajithan SA','Technical Consultant','2015-05-15','Bangalore','India'),(5,'Zohar Hachoma','Technical Manager','2005-07-11','Tel Aviv','Isreal');

```
mysql> INSERT INTO company1(EmpID,EMPNAME,Designation,DOJ,city,country) VALUES (1,'Barack Eilam','CEO','2007-01-01','Tel Aviv','Isreal'),(2,'Barry Cooper','Vice President','2009-05-09','Denver','US'),(3,'Aditya Tiwary','Director FLS','2015-07-15','Pune','India'),(4,'Abarajithan SA','Technical Consultant','2015-05-15','Bangalore','India'),(5,'Zohar Hachoma','Technical Manager','2005-07-11','Tel Aviv','Isreal');
Query OK, 5 rows affected, 2 warnings (0.00 sec)
Records: 5 Duplicates: 0 Warnings: 2

mysql>
mysql> Select * From company1;
+-----+
| EmpID | EMPNAME | Designation | DOJ | city | country |
+-----+
| 1 | Barack Eilam | CEO | 2007-01-01 | Tel Aviv | Isreal |
| 2 | Barry Cooper | Vice President | 2009-05-09 | Denver | US |
| 3 | Aditya Tiwary | Director FLS | 2015-07-15 | Pune | India |
| 4 | Abarajithan SA | Technical Consu | 2015-05-15 | Bangalore | India |
| 5 | Zohar Hachoma | Technical Manag | 2005-07-11 | Tel Aviv | Isreal |
+-----+
5 rows in set (0.00 sec)
```

Sqoop import step

```
sqoop import --connect jdbc:mysql://localhost/company --username root --password acadgild --split-by EmpID --columns EmpID,EMPNAME,city --table company1 --target-dir /myhive --hive-import --create-hive-table --hive-table default.Company1Hive -m 1
```

the above command will import the data from the table 'company1' to the new table 'Company1Hive' in to the database 'default' in the HIVE,

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/company --username root --password acadgild --split-by EmpID --columns EmpID,EMPNAME,city --table company1 --target-dir /myhive --hive-import --create-hive-table --hive-table default.Company1Hive -m 1
get-dir /myhive --hive-import --create-hive-table --hive-table default.Company1Hive -m 1
Warning: /home/acadgild/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/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.
Warning: /home/acadgild/sqoop-1.4.6/bin_hadoop-2.0.4-alpha/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-12-05 11:52:00,053 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-12-05 11:52:00,097 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2017-12-05 11:52:00,104 INFO [main] tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
```

4. Go hive shell, **Select * From Company1Hive;**

Expected Output

```
Time taken: 0.143 seconds, Fetched: 4 row(s)
hive (Default)>
> Select * From company1hive;
OK
company1hive.empid    company1hive.empname    company1hive.city
1      Barack Eilam           Tel Aviv
2      Barry Cooper          Denver
3      Aditya Tiwary           Pune
4      Abarajithan SA         Bangalore
5      Zohar Hachoma          Tel Aviv
Time taken: 3.574 seconds, Fetched: 5 row(s)
hive (Default)> █
```

Hence, the data from the 'company1' table has been imported into the hive table 'Company1hive'

Export command for transferring the selected columns from Hive to MySQL

1. In order to export data from the HIVE table to mysql, we need create a table in the mysql prior,

```
CREATE TABLE company2(EmpID int(11),EMPNAME varchar(20),city varchar(20));
```

```
mysql>
mysql> CREATE TABLE company2(EmpID int(11),EMPNAME varchar(20),city varchar(20));
Query OK, 0 rows affected (0.02 sec)

mysql>
mysql> DESCRIBE company2;
+-----+-----+-----+-----+-----+-----+
| Field | Type          | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| EmpID | int(11)        | YES  |     | NULL    |       |
| EMPNAME | varchar(20)    | YES  |     | NULL    |       |
| city  | varchar(20)    | YES  |     | NULL    |       |
+-----+-----+-----+-----+-----+-----+
3 rows in set (0.01 sec)
```

2. The data in this table must be empty,

Select * From company2;

```
mysql>  
mysql> Select * From company2;  
Empty set (0.00 sec)
```


Sqoop Export step

```
sqoop export --connect jdbc:mysql://localhost/company --username root --password acadgild --columns EmpID,EMPNAME,city --table company2 --export-dir /user/hive/warehouse/company1hive --input-fields-terminated-by '\001' -m 1
```

The above command will export the data from the hive table 'company1hive' into the newly created table 'company2' in the mysql,

```
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/company --username root --password acadgild --columns EmpID,EMPNAME,city --table company2 --export-dir /user/hive/warehouse/company1hive --input-fields-terminated-by '\001' -m 1
Warning: /home/acadgild/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/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.
Warning: /home/acadgild/sqoop-1.4.6.bin_hadoop-2.0.4-alpha/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-12-05 14:28:46,597 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
2017-12-05 14:28:46,631 WARN [main] tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
2017-12-05 14:28:47,043 INFO [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-12-05 14:28:47,043 INFO [main] tool.CodeGenTool: Beginning code generation
2017-12-05 14:28:47,572 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `company2` AS t LIMIT 1
2017-12-05 14:28:47,619 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `company2` AS t LIMIT 1
2017-12-05 14:28:47,659 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/hadoop-2.7.2
Note: Recompile with -Xlint:deprecation for details.
2017-12-05 14:28:52,139 INFO [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/65275fdf892ff545f4c972a732908844/company2.jar
2017-12-05 14:28:52,172 INFO [main] mapreduce.ExportJobBase: Beginning export of company2
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/hbase-1.0.3/lib/slf4j-log4j12-1.7.7.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
2017-12-05 14:29:23,822 INFO [main] mapreduce.Job: Job job_1512453129081_0010 running in uber mode : false
2017-12-05 14:29:23,825 INFO [main] mapreduce.Job: map 0% reduce 0%
2017-12-05 14:29:35,267 INFO [main] mapreduce.Job: map 100% reduce 0%
2017-12-05 14:29:36,297 INFO [main] mapreduce.Job: Job job_1512453129081_0010 completed successfully
2017-12-05 14:29:36,574 INFO [main] mapreduce.Job: Counters: 30
File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=136343
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
```

Company1hive data,

```
Time taken: 0.143 seconds, Fetched: 4 row(s)
hive (Default)>
> Select * From company1hive;
OK
company1hive.empid    company1hive.empname    company1hive.city
1      Barack Eilam           Tel Aviv
2      Barry Cooper           Denver
3      Aditya Tiwary           Pune
4      Abarajithan SA          Bangalore
5      Zohar Hachoma           Tel Aviv
Time taken: 3.574 seconds, Fetched: 5 row(s)
hive (Default)> █
```

Expected Output

```
mysql> Select * From company2;
+-----+-----+-----+
| EmpID | EMPNAME | city |
+-----+-----+-----+
| 1     | Barack Eilam | Tel Aviv |
| 2     | Barry Cooper | Denver |
| 3     | Aditya Tiwary | Pune |
| 4     | Abarajithan SA | Bangalore |
| 5     | Zohar Hachoma | Tel Aviv |
+-----+-----+-----+
5 rows in set (0.00 sec)
```