Assignment 11.1

Problem Statement :-

Perform and explain the code flow and the associated result for the below tasks. Candidates should create and use their own employee dataset for the same. Share the screenshot of the commands used and its associated result.

- Transfer data between Mysql and HDFS (Import and Export) using Sqoop
- Transfer data between Mysql and Hive (Import and Export only selected columns) using Sqoop.

Solution:-

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

Following is the contents of employee dataset used:-

```
[acadgild@localhost hadoop]$ cat employee_details.txt
101,Amitabh,20000,1
102,Shahrukh,10000,2
103,Akshay,11000,3
104,Anubhav,5000,4
105,Pawan,2500,5
106,Aamir,25000,1
107,Salman,17500,2
108,Ranbir,14000,3
109,Katrina,1000,4
110,Priyanka,2000,5
111,Tushar,500,1
112,Ajay,5000,2
113,Jubeen,1000,1
114,Madhuri,2000,2[acadgild@localhost hadoop]$
```

First,we nned to upload this dataset in MySql Table by the following steps:-

```
mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| information_schema |
| db                 |
| db1                |
| employee           |
| metastore          |
| mysql              |
+--------------------+
6 rows in set (0.06 sec)

mysql> use employee;
Database changed
mysql> show tables;
Empty set (0.00 sec)

mysql> create table employee_details
    -> (
    -> id int,
    -> name varchar(20),
    -> salary int,
    -> rating int
    -> );
Query OK, 0 rows affected (0.06 sec)

mysql> show tables;
+--------------------+
| Tables_in_employee |
+--------------------+
| employee_details   |
+--------------------+
1 row in set (0.00 sec)
```

```
mysql> desc employee_details
    -> ;
+--------+-------------+------+-----+---------+-------+
| Field  | Type        | Null | Key | Default | Extra |
+--------+-------------+------+-----+---------+-------+
| id     | int(11)     | YES  |     | NULL    |       |
| name   | varchar(20) | YES  |     | NULL    |       |
| salary | int(11)     | YES  |     | NULL    |       |
| rating | int(11)     | YES  |     | NULL    |       |
+--------+-------------+------+-----+---------+-------+
4 rows in set (0.01 sec)

mysql>
```

```
mysql> LOAD DATA LOCAL INFILE '/home/acadgild/hadoop/employee_details.txt' INTO TABLE employee_details COLUMNS TERMINATED BY ',';
Query OK, 14 rows affected (0.01 sec)
Records: 14  Deleted: 0  Skipped: 0  Warnings: 0

mysql> select * from employee_details;
+------+----------+--------+--------+
| id   | name     | salary | rating |
+------+----------+--------+--------+
|  101 | Amitabh  |  20000 |      1 |
|  102 | Shahrukh |  10000 |      2 |
|  103 | Akshay   |  11000 |      3 |
|  104 | Anubhav  |   5000 |      4 |
|  105 | Pawan    |   2500 |      5 |
|  106 | Aamir    |  25000 |      1 |
|  107 | Salman   |  17500 |      2 |
|  108 | Ranbir   |  14000 |      3 |
|  109 | Katrina  |   1000 |      4 |
|  110 | Priyanka |   2000 |      5 |
|  111 | Tushar   |    500 |      1 |
|  112 | Ajay     |   5000 |      2 |
|  113 | Jubeen   |   1000 |      1 |
|  114 | Madhuri  |   2000 |      2 |
+------+----------+--------+--------+
14 rows in set (0.00 sec)

mysql>
```

Now we transfer data from MySql to HDFS using sqoop import as follows:-

sqoop import --connect jdbc:mysql://localhost/employee --username 'acadgild' -P --table 'employee_details' --target-dir '/sqoopout_11.1' -m 1;

```
[acadgild@localhost ~]$
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/employee --username 'acadgild' -P --table 'employee_detai
ls' --target-dir '/sqoopout_11.1' -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.
2017-12-01 00:10:07,111 INFO  [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
```

```
[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 mig
ht 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'.
17/12/01 00:12:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java clas
ses where applicable
Found 9 items
drwxr-xr-x   - acadgild supergroup          0 2017-10-09 01:18 /home
drwxr-xr-x   - acadgild supergroup          0 2017-11-29 22:33 /sample
drwxr-xr-x   - acadgild supergroup          0 2017-10-21 13:31 /sqoopout
drwxr-xr-x   - acadgild supergroup          0 2017-12-01 00:11 /sqoopout_11.1
drwxr-xr-x   - acadgild supergroup          0 2017-10-21 14:19 /sqoopout_incremential_import
drwxr-xr-x   - acadgild supergroup          0 2017-10-22 11:59 /sqoopout_job_import
drwxr-xr-x   - acadgild supergroup          0 2017-10-21 13:37 /sqoopout_split
drwxrwx---   - acadgild supergroup          0 2017-11-01 00:54 /tmp
drwxr-xr-x   - acadgild supergroup          0 2016-08-18 09:34 /user
[acadgild@localhost ~]$
```

```
[acadgild@localhost ~]$ hadoop fs -ls /sqoopout_11.1/
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which mig
ht 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'.
17/12/01 00:13:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java clas
ses where applicable
Found 2 items
-rw-r--r--   1 acadgild supergroup          0 2017-12-01 00:11 /sqoopout_11.1/_SUCCESS
-rw-r--r--   1 acadgild supergroup        261 2017-12-01 00:11 /sqoopout_11.1/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout_11.1/part-m-00000
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which mig
ht 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'.
17/12/01 00:14:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java clas
ses where applicable
101,Amitabh,20000,1
102,Shahrukh,10000,2
103,Akshay,11000,3
104,Anubhav,5000,4
105,Pawan,2500,5
106,Aamir,25000,1
107,Salman,17500,2
108,Ranbir,14000,3
109,Katrina,1000,4
110,Priyanka,2000,5
111,Tushar,500,1
112,Ajay,5000,2
113,Jubeen,1000,1
114,Madhuri,2000,2
[acadgild@localhost ~]$
```

Now we will drop this MySql Table and export the contents to the table using sqoop export

```
mysql> delete from employee_details;
Query OK, 14 rows affected (0.01 sec)

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

mysql>
```

sqoop export --connect jdbc:mysql://localhost/employee --username 'acadgild' -P --table 'employee_details' --export-dir '/sqoopout_11.1' --input-fields-terminated-by ',' -m 1 --columns id,name,salary,rating

```
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/employee --username 'acadgild' -P --table 'employee_details' --export-dir '/sqoopout
_11.1' --input-fields-terminated-by ',' -m 1 --columns id,name,salary,rating
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.
2017-12-01 00:28:45,622 INFO  [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
```

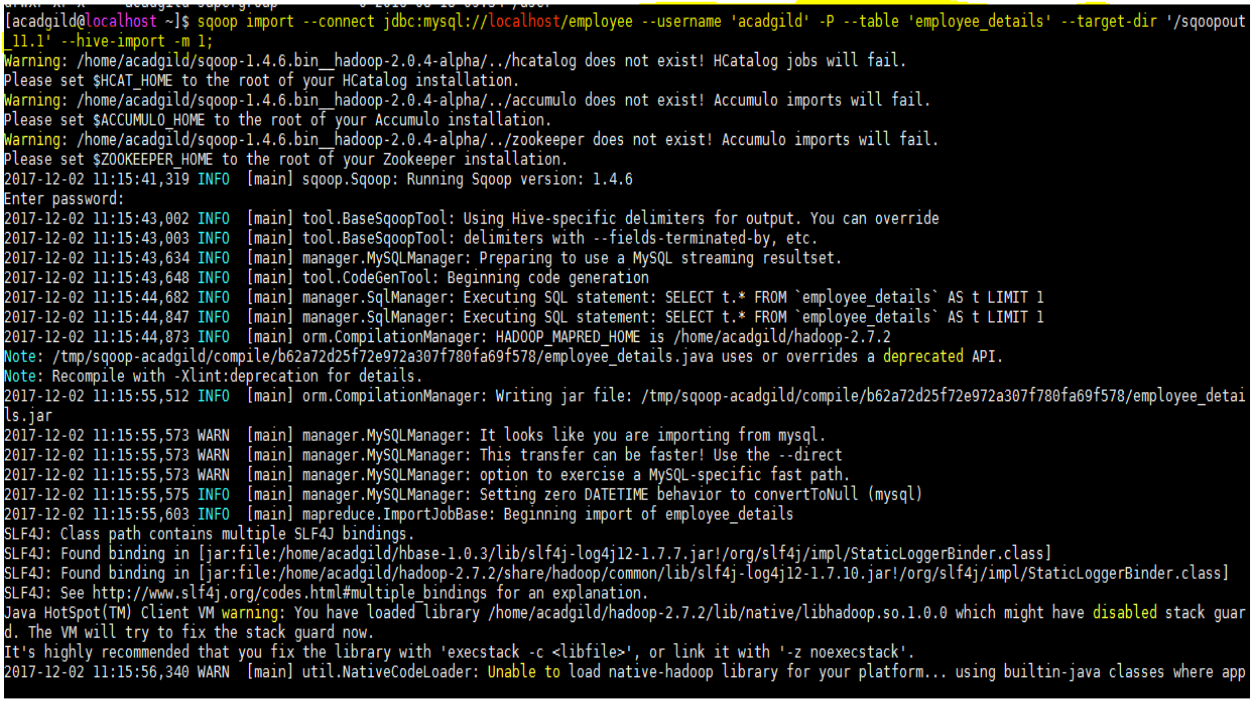
```
2017-12-01 00:30:57,864 INFO  [main] mapreduce.Job:  map 100% reduce 0%
2017-12-01 00:30:58,925 INFO  [main] mapreduce.Job: Job job_1512066217556_0003 completed successfully
2017-12-01 00:30:59,360 INFO  [main] mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=136356
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=393
                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)=15344
                Total time spent by all reduces in occupied slots (ms)=0
                Total time spent by all map tasks (ms)=15344
                Total vcore-seconds taken by all map tasks=15344
                Total megabyte-seconds taken by all map tasks=15712256
        Map-Reduce Framework
                Map input records=14
                Map output records=14
                Input split bytes=129
                Spilled Records=0
                Failed Shuffles=0
                Merged Map outputs=0
                GC time elapsed (ms)=280
                CPU time spent (ms)=1610
                Physical memory (bytes) snapshot=68091904
                Virtual memory (bytes) snapshot=323260416
                Total committed heap usage (bytes)=16318464
        File Input Format Counters
                Bytes Read=0
        File Output Format Counters
                Bytes Written=0
2017-12-01 00:30:59,390 INFO  [main] mapreduce.ExportJobBase: Transferred 393 bytes in 66.4735 seconds (5.9121 bytes/sec)
2017-12-01 00:30:59,401 INFO  [main] mapreduce.ExportJobBase: Exported 14 records.
[acadgild@localhost ~]$
```

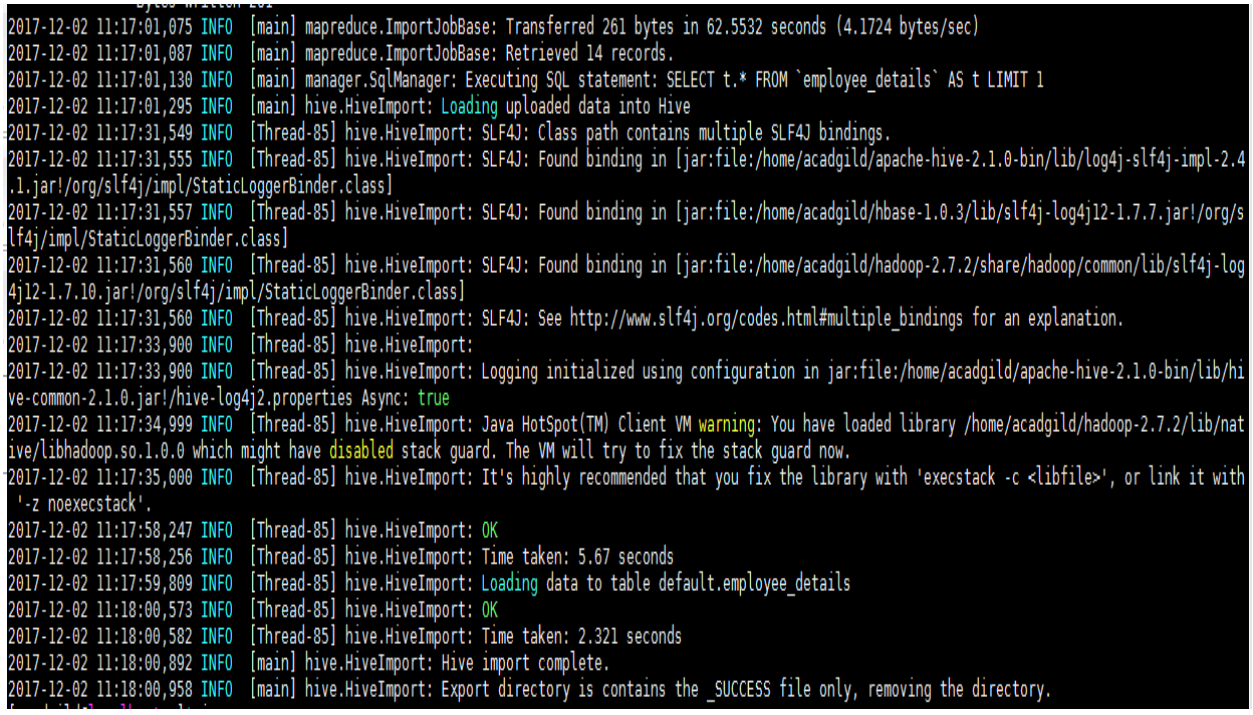
```
mysql> select * from employee_details;
+------+----------+--------+--------+
| id   | name     | salary | rating |
+------+----------+--------+--------+
|  101 | Amitabh  |  20000 |      1 |
|  102 | Shahrukh |  10000 |      2 |
|  103 | Akshay   |  11000 |      3 |
|  104 | Anubhav  |   5000 |      4 |
|  105 | Pawan    |   2500 |      5 |
|  106 | Aamir    |  25000 |      1 |
|  107 | Salman   |  17500 |      2 |
|  108 | Ranbir   |  14000 |      3 |
|  109 | Katrina  |   1000 |      4 |
|  110 | Priyanka |   2000 |      5 |
|  111 | Tushar   |    500 |      1 |
|  112 | Ajay     |   5000 |      2 |
|  113 | Jubeen   |   1000 |      1 |
|  114 | Madhuri  |   2000 |      2 |
+------+----------+--------+--------+
14 rows in set (0.00 sec)

mysql>
```

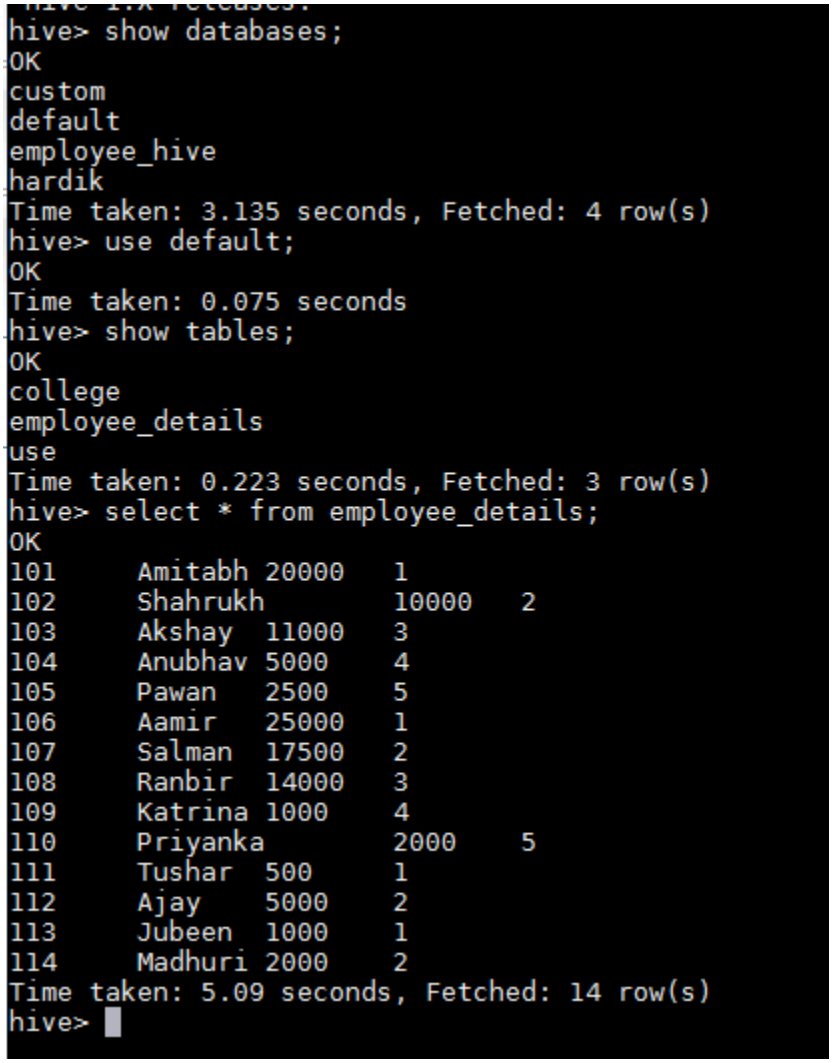
Sqoop export transfer the data fro HDFS to MySql as shown in above figure.

- **Transfer data between Mysql and Hive (Import and Export only selected columns) using Sqoop.**

As the data is already transfer back to MySql,we will transfer this data from MySql to Hive.







The above figure shows that data is transferred from MySQL to hive table in hive default database.

Submitted By:-

Hardik Kaushik