Sqoop Assignment 1

For this assignment, I've used Hortonworks VM and used MySQl, Hive and HDFS.

1) Starting the mySql using below syntax

```
sudo service mysqld start
```

2) Logging into mysql command line as user 'root'

```
mysql -u root
```

3) Created a database 'db1'.

```
create database db1;

use db1;
```

4) Granting previliges to user 'root'

```
grant all on *.* to 'root'@'localhost' with grant option;
```

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

- For this problem, First i created a Employee table in MySql like this.

```
CREATE TABLE employee (
   id int,
   name varchar(20),
     designation varchar(25),
   city varchar(15)
);
```

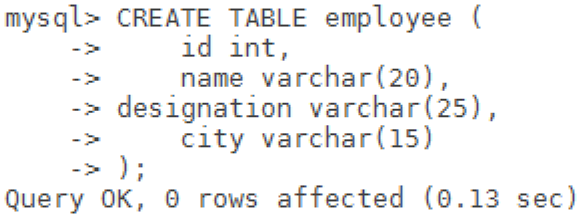
- Inserted few records in the employee table, by using the below syntax.

insert into employee values(1, 'kasthuri','Software Engg','Bangalore');

insert into employee values(2, 'malini','IT Analyst','Walnut');

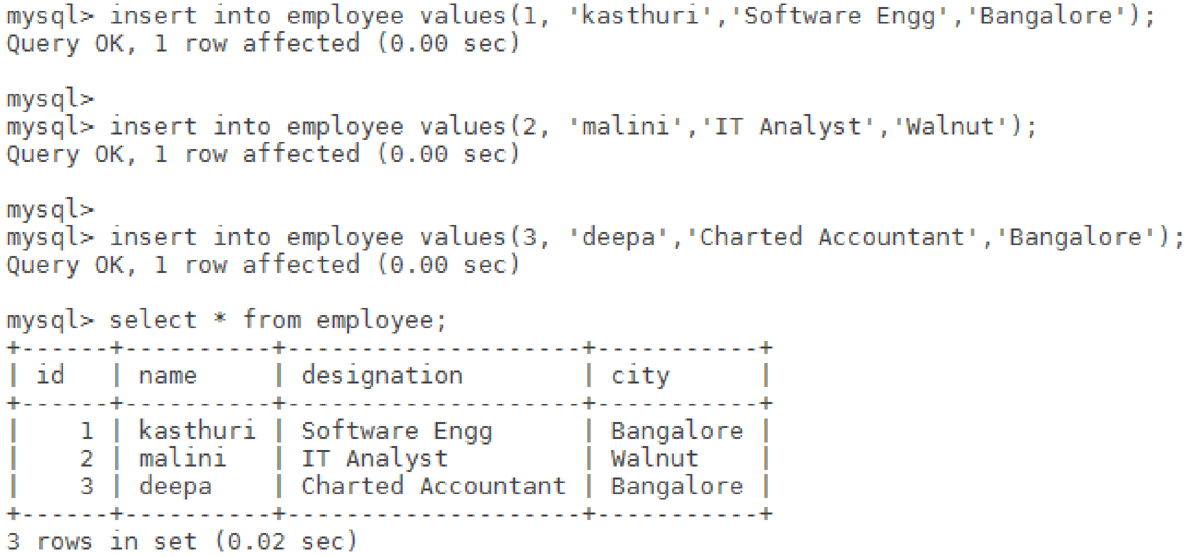
insert into employee values(3, 'deepa','Charted Accountant','Bangalore');

<u>Screenshot for creating employee table in mySql</u>



```
mysql> CREATE TABLE employee (
    ->     id int,
    ->     name varchar(20),
    -> designation varchar(25),
    ->     city varchar(15)
    -> );
Query OK, 0 rows affected (0.13 sec)
```

<u>Screenshot for inserting records in employee table in MySql</u>



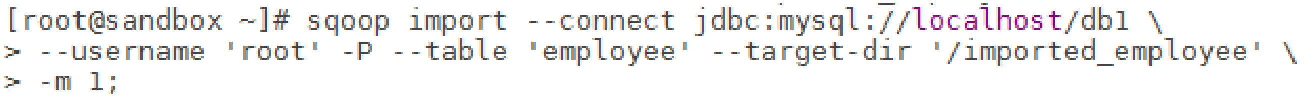
```
mysql> insert into employee values(1, 'kasthuri','Software Engg','Bangalore');
Query OK, 1 row affected (0.00 sec)

mysql>
mysql> insert into employee values(2, 'malini','IT Analyst','Walnut');
Query OK, 1 row affected (0.00 sec)

mysql>
mysql> insert into employee values(3, 'deepa','Charted Accountant','Bangalore');
Query OK, 1 row affected (0.00 sec)

mysql> select * from employee;
+------+----------+--------------------+-----------+
| id   | name     | designation        | city      |
+------+----------+--------------------+-----------+
|    1 | kasthuri | Software Engg      | Bangalore |
|    2 | malini   | IT Analyst         | Walnut    |
|    3 | deepa    | Charted Accountant | Bangalore |
+------+----------+--------------------+-----------+
3 rows in set (0.02 sec)
```

Importing the data from MySql to HDFS:

Then on the command line, executed following command to run Sqoop to import content of employee table in HDFS

```
sqoop import --connect jdbc:mysql://localhost/db1 \
--username 'root' -P --table 'employee' --target-dir '/imported_employee' \
-m 1;
```

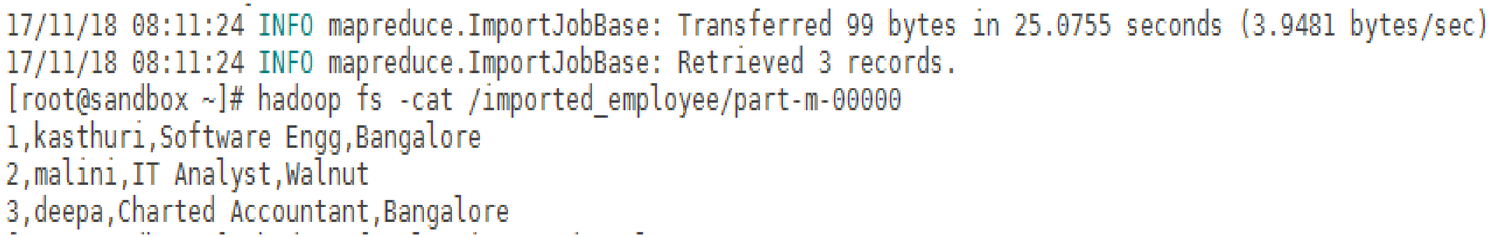
Screenshot of Mobaxterm for importing employee data in HDFS:



```
[root@sandbox ~]# sqoop import --connect jdbc:mysql://localhost/db1 \
> --username 'root' -P --table 'employee' --target-dir '/imported_employee' \
> -m 1;
```

In the above syntax, the target directory is /imported_employee. We can verify the contents of /imported_employee directory in hdfs, to check if the employee table contents are stored in HDFS

Screenshot of Mobaxterm after importing employee data in HDFS:



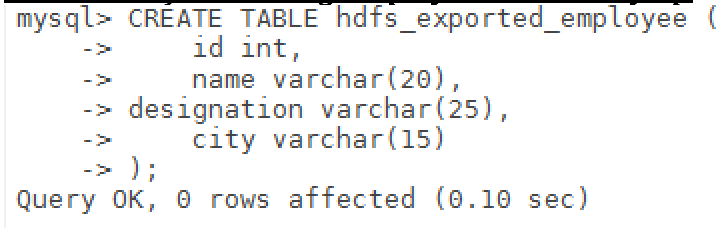
```
17/11/18 08:11:24 INFO mapreduce.ImportJobBase: Transferred 99 bytes in 25.0755 seconds (3.9481 bytes/sec)
17/11/18 08:11:24 INFO mapreduce.ImportJobBase: Retrieved 3 records.
[root@sandbox ~]# hadoop fs -cat /imported_employee/part-m-00000
1,kasthuri,Software Engg,Bangalore
2,malini,IT Analyst,Walnut
3,deepa,Charted Accountant,Bangalore
```

Exporting the data from HDFS to MySql:

For exporting employee data from HDFS to MySql, I've created a table "hdfs_exported_employee" using the below syntax in MySql.

```
CREATE TABLE hdfs_exported_employee (
   id int,
   name varchar(20),
  designation varchar(25),
   city varchar(15)
);
```

Screenshot for creating employee table in mySql

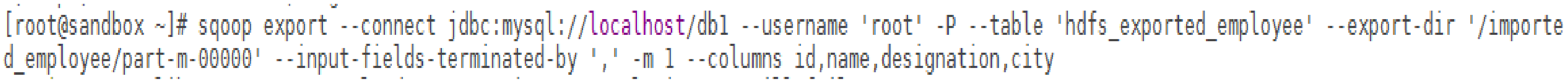


```
mysql> CREATE TABLE hdfs_exported_employee (
    ->     id int,
    ->     name varchar(20),
    -> designation varchar(25),
    ->     city varchar(15)
    -> );
Query OK, 0 rows affected (0.10 sec)
```

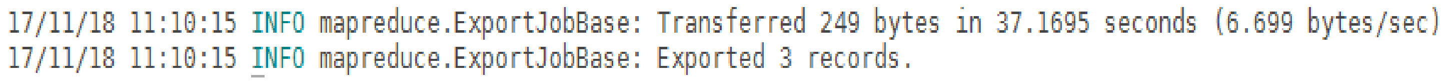
Then on the command line, executed following command to run Sqoop to export content of employee table from HDFS to "hdfs_exported_employee"

```
sqoop export --connect jdbc:mysql://localhost/db1 --username 'root' -P --table 'hdfs_exported_employee' --export-dir '/imported_employee/part-m-00000' --input-fields-terminated-by ',' -m 1 --columns id,name,designation,city
```

Screenshot of Mobaxterm for exporting employee data from HDFS to MySql:



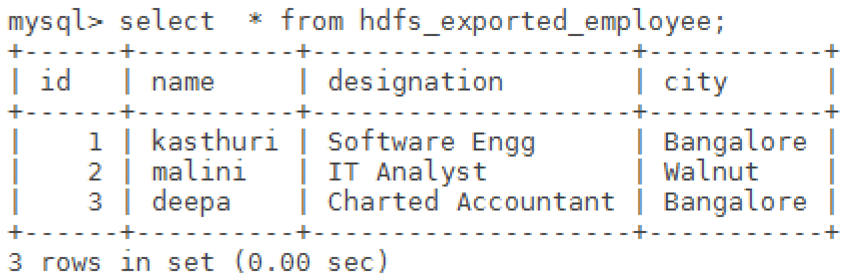
```
[root@sandbox ~]# sqoop export --connect jdbc:mysql://localhost/db1 --username 'root' -P --table 'hdfs_exported_employee' --export-dir '/importe
d_employee/part-m-00000' --input-fields-terminated-by ',' -m 1 --columns id,name,designation,city
```



```
17/11/18 11:10:15 INFO mapreduce.ExportJobBase: Transferred 249 bytes in 37.1695 seconds (6.699 bytes/sec)
17/11/18 11:10:15 INFO mapreduce.ExportJobBase: Exported 3 records.
```

Data is exported. As shown in the below screen, verified the contents HDFS exported employee table

Screenshot of Mobaxterm "hdfs exported employee"



```
mysql> select  * from hdfs_exported_employee;
+------+----------+-------------------+-----------+
| id   | name     | designation       | city      |
+------+----------+-------------------+-----------+
|    1 | kasthuri | Software Engg     | Bangalore |
|    2 | malini   | IT Analyst        | Walnut    |
|    3 | deepa    | Charted Accountant | Bangalore |
+------+----------+-------------------+-----------+
3 rows in set (0.00 sec)
```

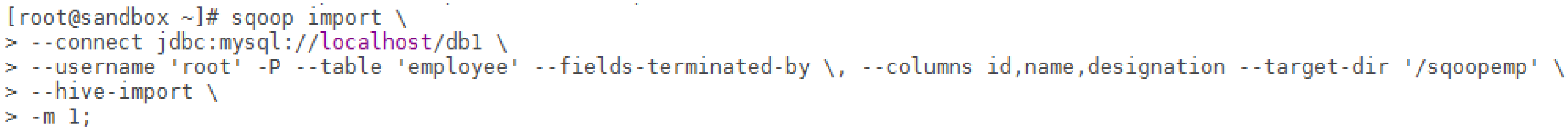
Transfer data between Mysql and Hive (Import and Export only selected columns) using Sqoop.

Importing the data from MySql to Hive:

On the command line, executed following command to run Sqoop to import content of employee table in Hive. We are not importing all the columns of employee table. We are importing id, name and designation of employee MySql table into the Hive employee table, and so specified the column that need to be to imported in the syntax.

```
sqoop import \
--connect jdbc:mysql://localhost/db1 \
--username 'root' -P --table 'employee' --fields-terminated-by \, --columns id,name,designation --target-dir '/sqoopemp' \
--hive-import \
-m 1;
```

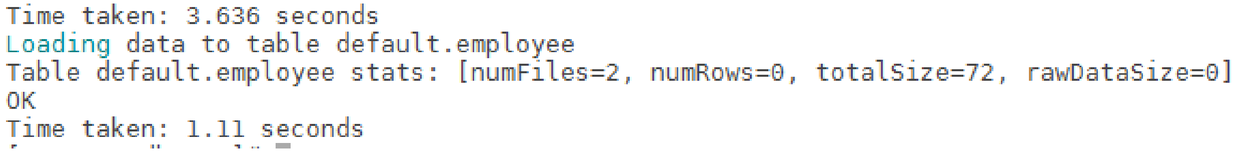
Screenshot of Mobaxterm for importing employee data in Hive:



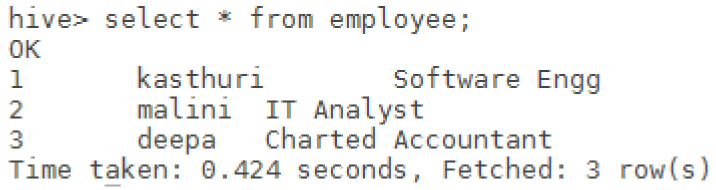
```
[root@sandbox ~]# sqoop import \
> --connect jdbc:mysql://localhost/db1 \
> --username 'root' -P --table 'employee' --fields-terminated-by \, --columns id,name,designation --target-dir '/sqoopemp' \
> --hive-import \
> -m 1;
```

In the above syntax, the target directory is /sqoopemp. The import is successful and so verified the contents of employee table in hive as below.

Screenshot of Mobaxterm after importing employee data in Hive:



```
Time taken: 3.636 seconds
Loading data to table default.employee
Table default.employee stats: [numFiles=2, numRows=0, totalSize=72, rawDataSize=0]
OK
Time taken: 1.11 seconds
```



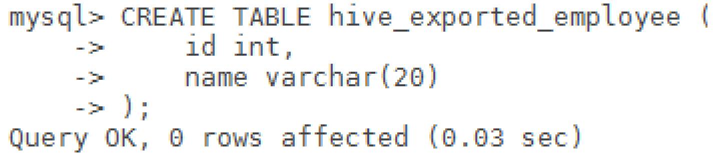
```
hive> select * from employee;
OK
1       kasthuri        Software Engg
2       malini  IT Analyst
3       deepa   Charted Accountant
Time taken: 0.424 seconds, Fetched: 3 row(s)
```

Exporting the data from Hive to MySql:

For exporting employee data from Hive to MySql, I've created a table "hive_exported_employee" using the below syntax in MySql. We are not exporting all the columns of employee hive table..

```
CREATE TABLE hive_exported_employee (
   id int,
   name varchar(20)
);
```

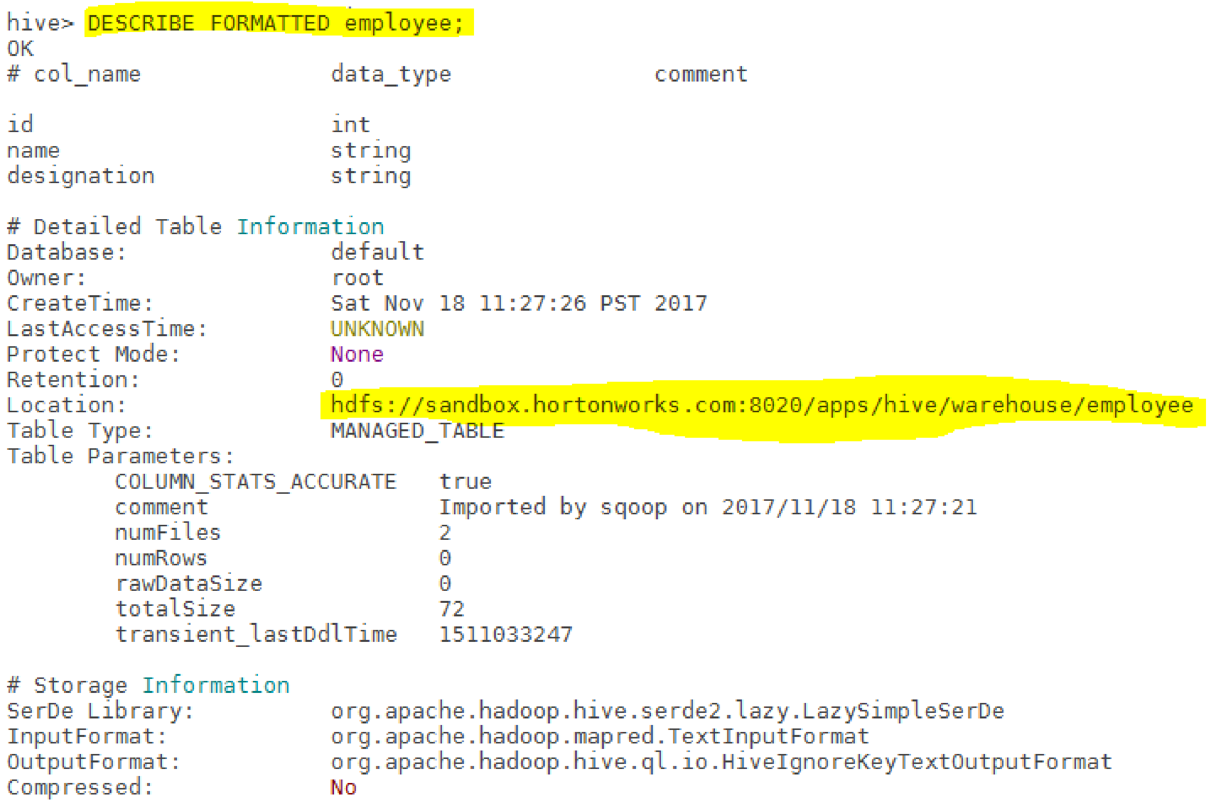
Screenshot for creating employee table in mySql



```
mysql> CREATE TABLE hive_exported_employee (
    ->     id int,
    ->     name varchar(20)
    -> );
Query OK, 0 rows affected (0.03 sec)
```

Steps followed to export from Hive:

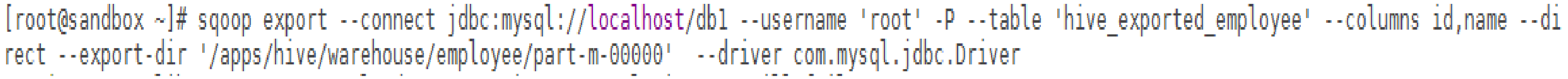
- When a hive table is created, the contents are stored in HDFS location . The location of employee hive table is found using DESCRIBE FORMATTED command. The hdfs location is used to export the data to MySql.



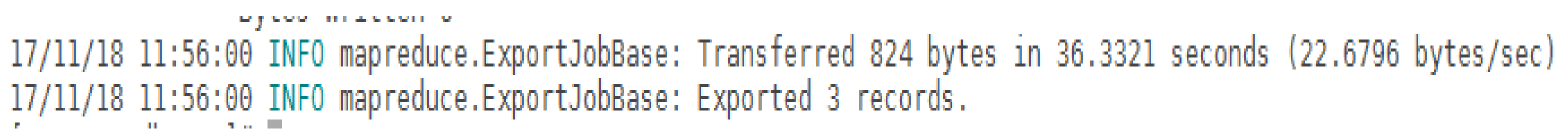
```
hive> DESCRIBE FORMATTED employee;
OK
# col_name              data_type               comment

id                      int
name                    string
designation             string

# Detailed Table Information
Database:               default
Owner:                  root
CreateTime:             Sat Nov 18 11:27:26 PST 2017
LastAccessTime:         UNKNOWN
Protect Mode:           None
Retention:              0
Location:               hdfs://sandbox.hortonworks.com:8020/apps/hive/warehouse/employee
Table Type:             MANAGED_TABLE
Table Parameters:
        COLUMN_STATS_ACCURATE   true
        comment                 Imported by sqoop on 2017/11/18 11:27:21
        numFiles                2
        numRows                 0
        rawDataSize             0
        totalSize               72
        transient_lastDdlTime   1511033247

# Storage Information
SerDe Library:          org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe
InputFormat:            org.apache.hadoop.mapred.TextInputFormat
OutputFormat:           org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat
Compressed:             No
```

- Then on the command line, executed following command to run Sqoop to export selected columns of employee hive table to “hive_exported_employee”. We are importing id and name of employee Hive table into the “hive_exported_employee”, MySql table. So specified the column that need to be to exported in the syntax sqoop export --connect jdbc:mysql://localhost/db1 --username 'root' -P --table 'hive_exported_employee' --columns id,name --direct --export-dir '/apps/hive/warehouse/employee/part-m-00000' --driver com.mysql.jdbc.Driver

<u>Screenshot of Mobaxterm for exporting employee data from Hive to MySql:</u>



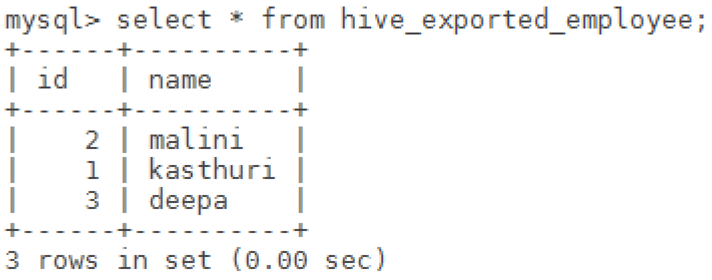
```
[root@sandbox ~]# sqoop export --connect jdbc:mysql://localhost/db1 --username 'root' -P --table 'hive_exported_employee' --columns id,name --di
rect --export-dir '/apps/hive/warehouse/employee/part-m-00000'  --driver com.mysql.jdbc.Driver
```



```
17/11/18 11:56:00 INFO mapreduce.ExportJobBase: Transferred 824 bytes in 36.3321 seconds (22.6796 bytes/sec)
17/11/18 11:56:00 INFO mapreduce.ExportJobBase: Exported 3 records.
```

Data is exported. As shown in the below screen, verified the contents Hive exported employee table

<u>Screenshot of Mobaxterm "hive_exported_employee"</u>



```
mysql> select * from hive_exported_employee;
+------+----------+
| id   | name     |
+------+----------+
|    2 | malini   |
|    1 | kasthuri |
|    3 | deepa    |
+------+----------+
3 rows in set (0.00 sec)
```