

Bigdata Assignment – 4.3

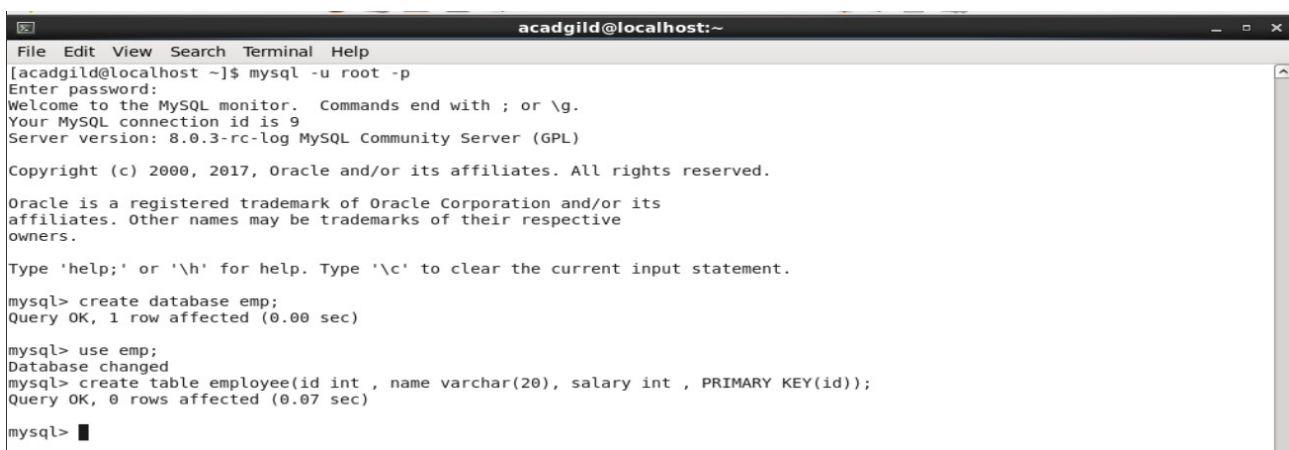
Perform and explain the code flow and the associated result for the below tasks. Candidates should create and use their own employee dataset for the same. Share the screenshot of the commands used and its associated result.

- 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.

Task 1 -

For importing data into hdfs from mysql

- Created a database and used it and it created a table.
create database emp;
use emp;
create table employee(id int, name varchar(20) , salary int , PRIMARY KEY(id));



```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ mysql -u root -p  
Enter password:  
Welcome to the MySQL monitor.  Commands end with ; or \g.  
Your MySQL connection id is 9  
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> create database emp;  
Query OK, 1 row affected (0.00 sec)  
  
mysql> use emp;  
Database changed  
mysql> create table employee(id int , name varchar(20), salary int , PRIMARY KEY(id));  
Query OK, 0 rows affected (0.07 sec)  
  
mysql> █
```

- Inserted data into the table.
Insert into employee values(1,'Ritesh',20000);

```
mysql> insert into employee values(1,'Ritesh',20000);
Query OK, 1 row affected (0.01 sec)

mysql> insert into employee values(2,'Shyam',30000);
Query OK, 1 row affected (0.01 sec)

mysql> insert into employee values(3,'Ram',10000);
Query OK, 1 row affected (0.00 sec)

mysql> select * from employee;
+----+-----+-----+
| id | name  | salary |
+----+-----+-----+
| 1  | Ritesh | 20000  |
| 2  | Shyam  | 30000  |
| 3  | Ram    | 10000  |
+----+-----+-----+
3 rows in set (0.00 sec)

mysql> █
```

- Imported the mysql data into hdfs by using sqoop
sqoop import --connect jdbc:mysql://localhost/emp --username root --password Root@123 --table employee --m 1 --target-dir /sqoopomport/emp

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/emp --username root --password Root@123 --table employee --m 1 --target-dir /sqoopomport/emp
```

- Then we checked whether the file is imported into hdfs or not, then its contents.
hadoop fs -cat /sqoopomport/emp/part-m-00000



```
acadgild@localhost:~
[acadgild@localhost ~]$ hadoop fs -ls /sqoopomport
18/06/13 21:03:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 1 items
drwxr-xr-x 1 acadgild supergroup 0 2018-06-13 21:01 /sqoopomport/emp
[acadgild@localhost ~]$ hadoop fs -ls /sqoopomport/emp
18/06/13 21:03:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Found 2 items
-rw-r--r-- 1 acadgild supergroup 0 2018-06-13 21:01 /sqoopomport/emp/_SUCCESS
-rw-r--r-- 1 acadgild supergroup 41 2018-06-13 21:01 /sqoopomport/emp/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /sqoopomport/emp/part-m-00000
18/06/13 21:04:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
1,Ritesh,20000
2,Shyam,30000
3,Ram,10000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ █
```

In the above screenshot it is visible that the same data we got in the hdfs

For exporting data into mysql from hdfs:

- Used the created database and created a table in it with same columns.
use emp;
create table sqoop_export(id int, name varchar(20) , salary int , PRIMARY KEY(id));

```
mysql> use emp;
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> create table sqoop_export(id int , name varchar(20), salary int , PRIMARY KEY(id));
Query OK, 0 rows affected (0.03 sec)
```

- Transferred the data from hdfs into mysql using sqoop
sqoop export --connect jdbc:mysql://localhost/emp --username 'root' --password [Root@123](#) --table 'sqoop_export' --export-dir '/sqoopomport/emp/part-m-00000' --input-fields-terminated-by ',' --m 1 --columns id,name,salary

```
acadgild@localhost:~$ sqoop export --connect jdbc:mysql://localhost/emp --username 'root' -password Root@123 --table 'sqoop_export' --export-dir '/sqoopomport/emp/part-m-00000' --input-fields-terminated-by ',' -m 1 --columns id,name,salary
Warning: /home/acadgild/.ssh/known_hosts:1:4:6: his hostname 'localhost' does not exist! Please add it to the list of known hosts.
```

- Checked the data of the table
select * from sqoop_export ;

```
mysql> select * from sqoop_export;
+----+-----+-----+
| id | name  | salary |
+----+-----+-----+
| 1  | Ritesh | 20000  |
| 2  | Shyam  | 30000  |
| 3  | Ram    | 10000  |
+----+-----+-----+
3 rows in set (0.00 sec)

mysql> █
```

Task 2

Transfer of data from mysql into hive:

- Transfer of data using sqoop from mysql table which was created earlier into hive.

sqoop import --connect jdbc:mysql://localhost/emp --username 'root' --password [Root@123](#) --split-by id --columns id,name,salary --table employee --target-dir 'sqoopimport/emphive' --hive-import --create-hivetable --hive-table default.mysqlemployee -m 1

```
acadgild@localhost:~$ sqoop import --connect jdbc:mysql://localhost/emp --username 'root' --password Root@123 --split-by id --columns id,name,salary --table employee --target-dir '/sqoopimport/emphive' --hive-import --create-hive-table --hive-table default.mysqlemployee -m 1
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.
18/06/13 21:25:36 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/06/13 21:25:36 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/06/13 21:25:36 INFO tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
18/06/13 21:25:36 INFO tool.BaseSqoopTool: delimiters with --fields-terminated-by, etc.
18/06/13 21:25:36 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/06/13 21:25:36 INFO tool.CodeGenTool: Beginning code generation
Wed Jun 13 21:25:37 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 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.
18/06/13 21:25:38 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
18/06/13 21:25:38 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee` AS t LIMIT 1
18/06/13 21:25:38 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/7f5a56de29c003f26f921178f78842aa/employee.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/06/13 21:25:41 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/7f5a56de29c003f26f921178f78842aa/employee.jar
18/06/13 21:25:41 WARN manager.MySQLManager: It looks like you are importing from mysql.
18/06/13 21:25:41 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
18/06/13 21:25:41 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
18/06/13 21:25:41 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
18/06/13 21:25:41 INFO mapreduce.ImportJobBase: Beginning import of employee
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]
```

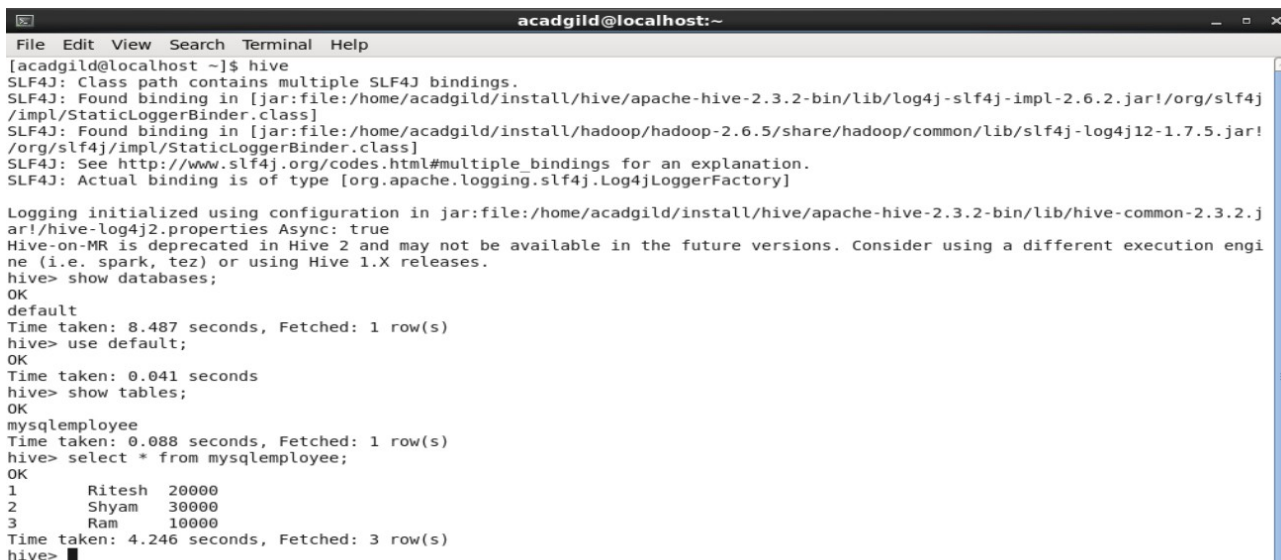
- In hive , in the default database we checked the table is present or not , then the content of the table.

show databases;

use default;

show tables;

select * from mysqlemployee;



```

acadgild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
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: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLoggerFactory]

Logging initialized using configuration in jar:file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-common-2.3.2.jar!/hive-log4j2.properties Async: true
Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
hive> show databases;
OK
default
Time taken: 8.487 seconds, Fetched: 1 row(s)
hive> use default;
OK
Time taken: 0.041 seconds
hive> show tables;
OK
mysqlemployee
Time taken: 0.088 seconds, Fetched: 1 row(s)
hive> select * from mysqlemployee;
OK
1      Ritesh  20000
2      Shyam  30000
3      Ram    10000
Time taken: 4.246 seconds, Fetched: 3 row(s)
hive>

```

In the above screenshot , we can see the contents of the table same as in mysql table.

Transfer of data from hive into mysql:

- In mysql , in the emp database we created a table with id as only columns so tha we can export only id from hive.

use emp;
create table hive_export_id(id int);

```

mysql> use emp;
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> create table hive_export_id(id int);
Query OK, 0 rows affected (0.03 sec)

mysql>

```

- We transfered the data from hive to mysql using sqoop.
sqoop export --connect jdbc:mysql://localhost/emp --username 'root' --password [Root@123](#) --table 'hive_export' --columns id --export-dir /user/hive/warehouse/mysqlemployee/part-m-00000 --input-fields-terminated-by '\001' -m 1


```
Applications Places System acadgild@localhost:~ Wed Jun 13, 9:58 PM Acadgild
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/emp --username 'root' -password Root@123 --table 'hive_export_id' --columns id --export-dir /user/hive/warehouse/mysqlemployee/part-m-00000 --input-fields-terminated-by '\001' -m 1
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.
18/06/13 21:56:28 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/06/13 21:56:28 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/06/13 21:56:28 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/06/13 21:56:28 INFO tool.CodeGenTool: Beginning code generation
Wed Jun 13 21:56:28 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 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.
18/06/13 21:56:30 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `hive_export_id` AS t LIMIT 1
18/06/13 21:56:30 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `hive_export_id` AS t LIMIT 1
18/06/13 21:56:30 INFO orm.CompilationManager: HADOOP MAPRED HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/8874b77a41f281cd5b28ac604f52a9dc/hive_export_id.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/06/13 21:56:33 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/8874b77a41f281cd5b28ac604f52a9dc/hive_export_id.jar
18/06/13 21:56:33 INFO mapreduce.ExportJobBase: Beginning export of hive_export_id
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]
18/06/13 21:56:33 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/06/13 21:56:33 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/06/13 21:56:35 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
18/06/13 21:56:35 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
```

- In mysql , we checked the contents of the table where data is exported.
select * from hive_export_id;

```
mysql> select * from hive_export_id;
+----+
| id |
+----+
| 1  |
| 2  |
| 3  |
+----+
3 rows in set (0.00 sec)

mysql> █
```

We can see we exported only id's from the hive table into mysql table. It is same as in the hive table.