# Assignment 11.2

Problem Statement :-

Read from MySQL Table and load it in Hive table.Create hive table if it does not exist.If it exists, perform the incremental load.

Solution:-

First,we need 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 will upload the data from MySQL to Hive table as follows:-

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/employee --username 'acadgild' -P --table 'employee_details' --target-dir '/sqoopout
_11.1' --hive-import -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.
2017-12-02 11:15:41,319 INFO  [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
2017-12-02 11:15:43,002 INFO  [main] tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
2017-12-02 11:15:43,003 INFO  [main] tool.BaseSqoopTool: delimiters with --fields-terminated-by, etc.
2017-12-02 11:15:43,634 INFO  [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-12-02 11:15:43,648 INFO  [main] tool.CodeGenTool: Beginning code generation
2017-12-02 11:15:44,682 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee_details` AS t LIMIT 1
2017-12-02 11:15:44,847 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee_details` AS t LIMIT 1
2017-12-02 11:15:44,873 INFO  [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/hadoop-2.7.2
Note: /tmp/sqoop-acadgild/compile/b62a72d25f72e972a307f780fa69f578/employee_details.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-12-02 11:15:55,512 INFO  [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/b62a72d25f72e972a307f780fa69f578/employee_detai
ls.jar
2017-12-02 11:15:55,573 WARN  [main] manager.MySQLManager: It looks like you are importing from mysql.
2017-12-02 11:15:55,573 WARN  [main] manager.MySQLManager: This transfer can be faster! Use the --direct
2017-12-02 11:15:55,573 WARN  [main] manager.MySQLManager: option to exercise a MySQL-specific fast path.
2017-12-02 11:15:55,575 INFO  [main] manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
2017-12-02 11:15:55,603 INFO  [main] mapreduce.ImportJobBase: Beginning import of employee_details
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 guar
d. 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'.
2017-12-02 11:15:56,340 WARN  [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where app
```

```
2017-12-02 11:17:01,075 INFO  [main] mapreduce.ImportJobBase: Transferred 261 bytes in 62.5532 seconds (4.1724 bytes/sec)
2017-12-02 11:17:01,087 INFO  [main] mapreduce.ImportJobBase: Retrieved 14 records.
2017-12-02 11:17:01,130 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee_details` AS t LIMIT 1
2017-12-02 11:17:01,295 INFO  [main] hive.HiveImport: Loading uploaded data into Hive
2017-12-02 11:17:31,549 INFO  [Thread-85] hive.HiveImport: SLF4J: Class path contains multiple SLF4J bindings.
2017-12-02 11:17:31,555 INFO  [Thread-85] hive.HiveImport: SLF4J: Found binding in [jar:file:/home/acadgild/apache-hive-2.1.0-bin/lib/log4j-slf4j-impl-2.4
.1.jar!/org/slf4j/impl/StaticLoggerBinder.class]
2017-12-02 11:17:31,557 INFO  [Thread-85] hive.HiveImport: SLF4J: Found binding in [jar:file:/home/acadgild/hbase-1.0.3/lib/slf4j-log4j12-1.7.7.jar!/org/s
lf4j/impl/StaticLoggerBinder.class]
2017-12-02 11:17:31,560 INFO  [Thread-85] hive.HiveImport: SLF4J: Found binding in [jar:file:/home/acadgild/hadoop-2.7.2/share/hadoop/common/lib/slf4j-log
4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
2017-12-02 11:17:31,560 INFO  [Thread-85] hive.HiveImport: SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2017-12-02 11:17:33,900 INFO  [Thread-85] hive.HiveImport:
2017-12-02 11:17:33,900 INFO  [Thread-85] hive.HiveImport: Logging initialized using configuration in jar:file:/home/acadgild/apache-hive-2.1.0-bin/lib/hi
ve-common-2.1.0.jar!/hive-log4j2.properties Async: true
2017-12-02 11:17:34,999 INFO  [Thread-85] hive.HiveImport: Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/nat
ive/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.
2017-12-02 11:17:35,000 INFO  [Thread-85] hive.HiveImport: It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with
 '-z noexecstack'.
2017-12-02 11:17:58,247 INFO  [Thread-85] hive.HiveImport: OK
2017-12-02 11:17:58,256 INFO  [Thread-85] hive.HiveImport: Time taken: 5.67 seconds
2017-12-02 11:17:59,809 INFO  [Thread-85] hive.HiveImport: Loading data to table default.employee_details
2017-12-02 11:18:00,573 INFO  [Thread-85] hive.HiveImport: OK
2017-12-02 11:18:00,582 INFO  [Thread-85] hive.HiveImport: Time taken: 2.321 seconds
2017-12-02 11:18:00,892 INFO  [main] hive.HiveImport: Hive import complete.
2017-12-02 11:18:00,958 INFO  [main] hive.HiveImport: Export directory is contains the _SUCCESS file only, removing the directory.
```

```
hive> show databases;
OK
custom
default
employee_hive
hardik
Time taken: 3.135 seconds, Fetched: 4 row(s)
hive> use default;
OK
Time taken: 0.075 seconds
hive> show tables;
OK
college
employee_details
use
Time taken: 0.223 seconds, Fetched: 3 row(s)
hive> select * from employee_details;
OK
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
Time taken: 5.09 seconds, Fetched: 14 row(s)
hive>
```

Now we will add few new records in MySQL Table as follows:-

```
mysql> insert into employee_details values(115, 'Rajeev',1010,1);
Query OK, 1 row affected (0.01 sec)

mysql> insert into employee_details values(116, 'Anil',1020,1);
Query OK, 1 row affected (0.00 sec)

mysql> insert into employee_details values(117, 'Subhash',2010,2);
Query OK, 1 row affected (0.00 sec)

mysql> commit;
Query OK, 0 rows affected (0.00 sec)

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 |
|  115 | Rajeev   |   1010 |      1 |
|  116 | Anil     |   1020 |      1 |
|  117 | Subhash  |   2010 |      2 |
+------+----------+--------+--------+
17 rows in set (0.01 sec)

mysql>
```

Transfer these added values into Hive table:-

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/employee --username 'acadgild' -P --table 'employee_details' --target-dir '/sqoopout
_11.2' --incremental append --check-column id --last-value 114 -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.
2017-12-02 20:12:51,029 INFO  [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
2017-12-02 20:12:53,473 INFO  [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-12-02 20:12:53,474 INFO  [main] tool.CodeGenTool: Beginning code generation
2017-12-02 20:12:54,505 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee_details` AS t LIMIT 1
2017-12-02 20:12:54,598 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee_details` AS t LIMIT 1
2017-12-02 20:12:54,655 INFO  [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/hadoop-2.7.2
Note: /tmp/sqoop-acadgild/compile/16984b8d14e957d06b18059ec74195be/employee_details.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-12-02 20:13:04,716 INFO  [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/16984b8d14e957d06b18059ec74195be/employee_detai
ls.jar
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 guar
d. 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'.
2017-12-02 20:13:05,425 WARN  [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where app
licable
```

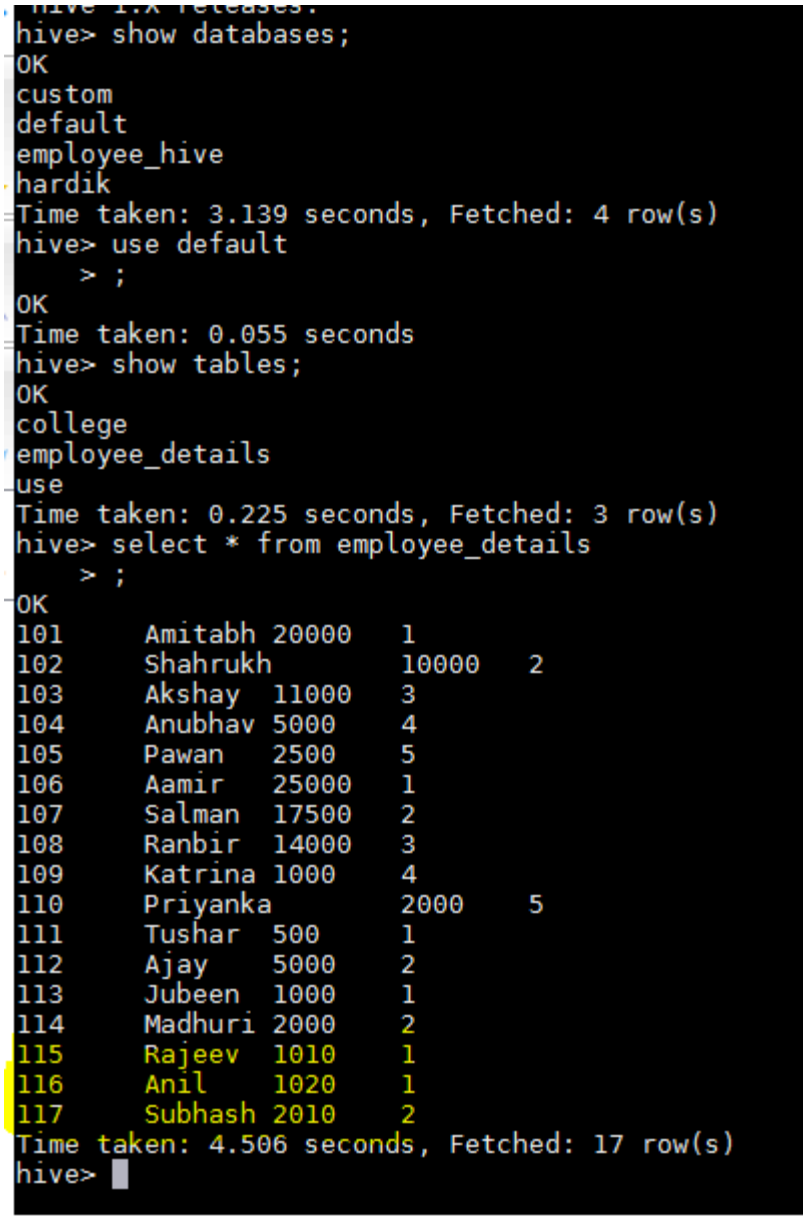
```
[acadgild@localhost ~]$ hadoop fs -ls /
Java HotSpot(TM) Client VM warning: You have loaded library /home/acadgild/hadoop-2.7.2/lib/native/libhadoop.so.1.0.0 which might have disabled stack guar
d. The VM will try to fix the stack guard now.
It's highly recommended that you fix the library with 'execstack -c <libfile>', or link it with '-z noexecstack'.
17/12/02 20:14:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 9 items
drwxr-xr-x   - acadgild supergroup          0 2017-10-09 01:18 /home
drwxr-xr-x   - acadgild supergroup          0 2017-12-02 10:43 /sample
drwxr-xr-x   - acadgild supergroup          0 2017-10-21 13:31 /sqoopout
drwxr-xr-x   - acadgild supergroup          0 2017-12-02 20:14 /sqoopout_11.2
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 2017-12-02 10:53 /tmp
drwxr-xr-x   - acadgild supergroup          0 2016-08-18 09:34 /user
[acadgild@localhost ~]$ hadoop fs -ls /sqoopout_11.2/
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 guar
d. 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'.
17/12/02 20:15:27 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         53 2017-12-02 20:14 /sqoopout_11.2/part-m-00000
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout_11.2/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 might have disabled stack guar
d. 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'.
17/12/02 20:16:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
115,Rajeev,1010,1
116,Anil,1020,1
117,Subhash,2010,2
```

In order to load this data to Hive table,we have to drop the table if already exists.

```
hive> show tables;
OK
college
employee_details
use
Time taken: 0.151 seconds, Fetched: 3 row(s)
hive> drop table employee_details;
OK
Time taken: 6.682 seconds
hive> show tables;
OK
college
use
Time taken: 0.154 seconds, Fetched: 2 row(s)
```

Now we will transfer data fro MySQL to Hive.

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/employee --username 'acadgild' -P --table 'employee_details' --target-dir '/sqoopout
_11.2_hive_import' --hive-import -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.
2017-12-02 20:46:05,401 INFO  [main] sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
2017-12-02 20:46:11,114 INFO  [main] tool.BaseSqoopTool: Using Hive-specific delimiters for output. You can override
2017-12-02 20:46:11,115 INFO  [main] tool.BaseSqoopTool: delimiters with --fields-terminated-by, etc.
2017-12-02 20:46:11,702 INFO  [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-12-02 20:46:11,716 INFO  [main] tool.CodeGenTool: Beginning code generation
2017-12-02 20:46:12,516 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee_details` AS t LIMIT 1
2017-12-02 20:46:12,633 INFO  [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM `employee_details` AS t LIMIT 1
2017-12-02 20:46:12,654 INFO  [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/hadoop-2.7.2
Note: /tmp/sqoop-acadgild/compile/1e225951b9d07d868814cf1decc6cdc8/employee_details.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-12-02 20:46:19,469 INFO  [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/1e225951b9d07d868814cf1decc6cdc8/employee_detai
ls.jar
2017-12-02 20:46:19,521 WARN  [main] manager.MySQLManager: It looks like you are importing from mysql.
```




The above shows that data changes are loaded successfully into hive table.

Submitted By:-

Hardik Kaushik