

Assignment 20.1

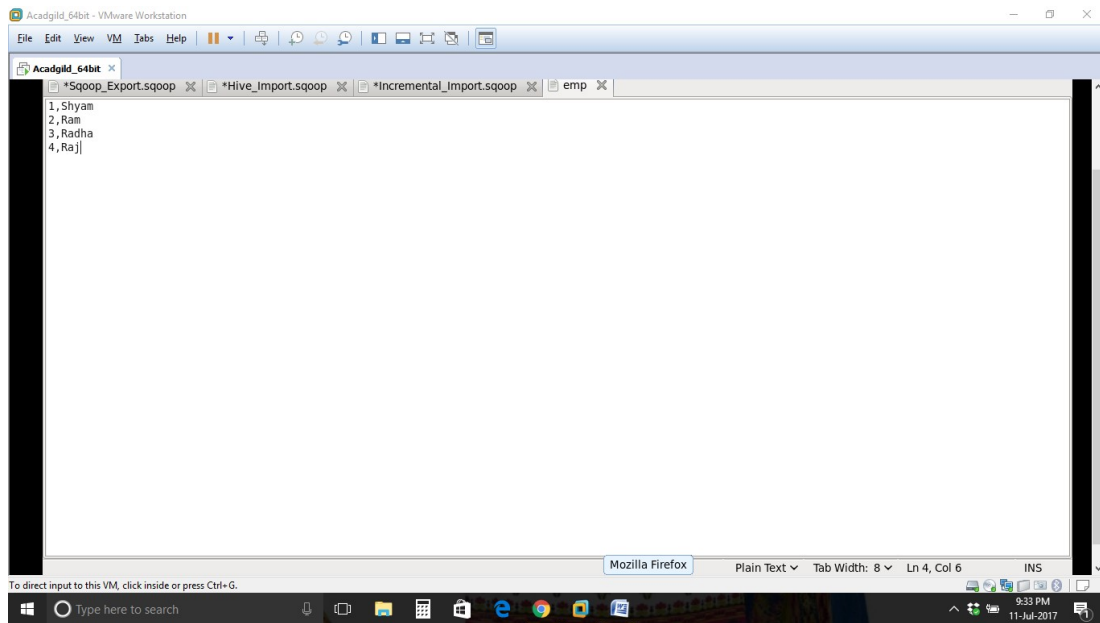
Problem Statement

Perform UPSERT in Sqoop export.

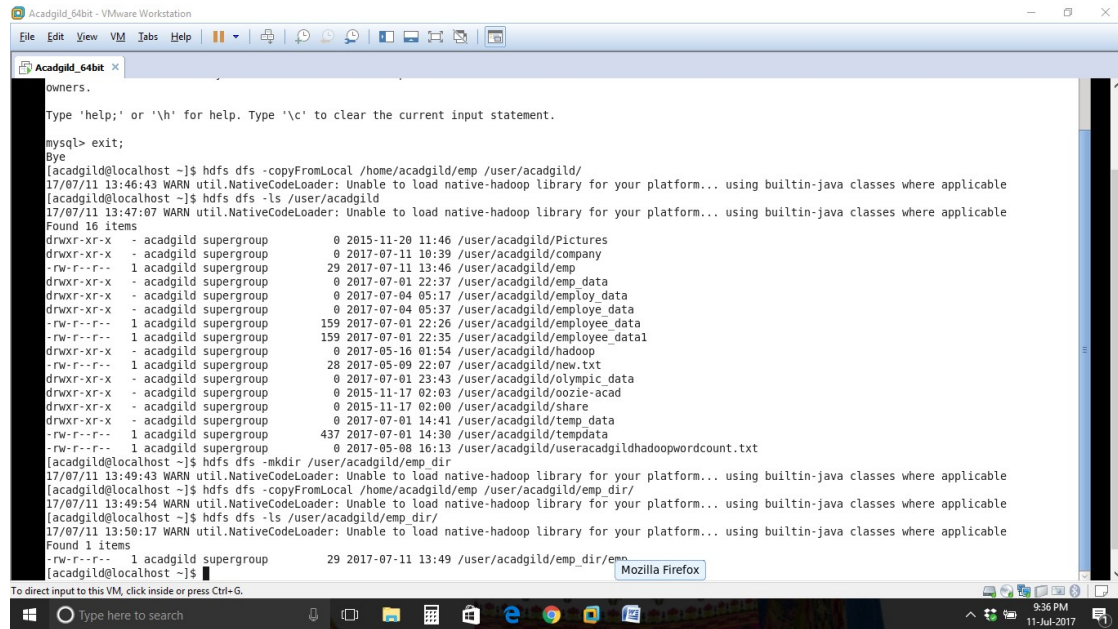
Read a file from HDFS and based on the id field, perform UPSERT in MySQL table.

In UPSERT, if the field exists, then it is updated else it is inserted.

1. Emp file in local file system

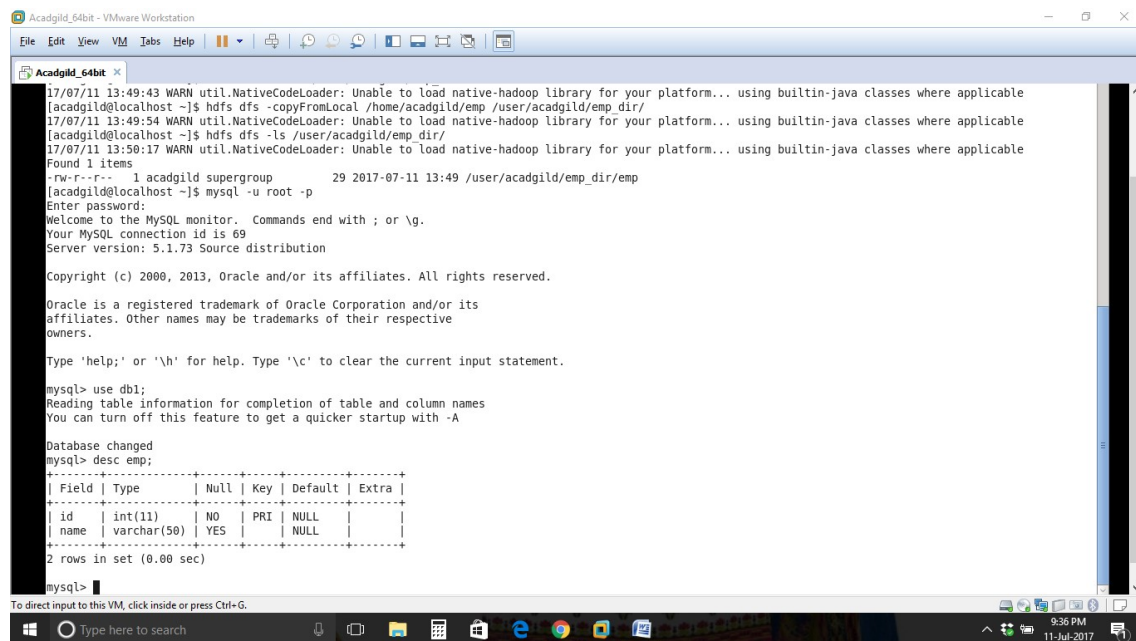


File copied to HDFS



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
owners.
Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.
mysql> exit;
Bye
[acadgild@localhost ~]$ hdfs dfs -copyFromLocal /home/acadgild/emp /user/acadgild/
17/07/11 13:46:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild
17/07/11 13:47:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 16 items
drwxr-xr-x - acadgild supergroup 0 2015-11-20 11:46 /user/acadgild/Pictures
drwxr-xr-x - acadgild supergroup 0 2017-07-11 10:39 /user/acadgild/company
-rw-r--r-- 1 acadgild supergroup 29 2017-07-11 13:46 /user/acadgild/emp
drwxr-xr-x - acadgild supergroup 0 2017-07-01 22:37 /user/acadgild/emp_data
drwxr-xr-x - acadgild supergroup 0 2017-07-04 05:17 /user/acadgild/employee_data
drwxr-xr-x - acadgild supergroup 0 2017-07-04 05:37 /user/acadgild/employee_data
-rw-r--r-- 1 acadgild supergroup 159 2017-07-01 22:26 /user/acadgild/employee_data1
-rw-r--r-- 1 acadgild supergroup 159 2017-07-01 22:35 /user/acadgild/employee_data1
drwxr-xr-x - acadgild supergroup 0 2017-05-16 01:54 /user/acadgild/hadoop
-rw-r--r-- 1 acadgild supergroup 28 2017-05-09 22:07 /user/acadgild/new.txt
drwxr-xr-x - acadgild supergroup 0 2017-07-01 23:43 /user/acadgild/olympic_data
drwxr-xr-x - acadgild supergroup 0 2015-11-17 02:03 /user/acadgild/oozie-acad
drwxr-xr-x - acadgild supergroup 0 2015-11-17 02:00 /user/acadgild/share
drwxr-xr-x - acadgild supergroup 0 2017-07-01 14:41 /user/acadgild/temp_data
-rw-r--r-- 1 acadgild supergroup 437 2017-07-01 14:30 /user/acadgild/tempdata
-rw-r--r-- 1 acadgild supergroup 0 2017-05-08 16:13 /user/acadgild/useracadgildhadoopwordcount.txt
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/emp_dir
17/07/11 13:49:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -copyFromLocal /home/acadgild/emp /user/acadgild/emp_dir/
17/07/11 13:49:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/emp_dir/
17/07/11 13:50:17 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 29 2017-07-11 13:49 /user/acadgild/emp_dir/emp
[acadgild@localhost ~]$
```

Table created in mysql



```
Acadgild_64bit - VMware Workstation
File Edit View VM Tabs Help
Acadgild_64bit x
17/07/11 13:49:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -copyFromLocal /home/acadgild/emp /user/acadgild/emp_dir/
17/07/11 13:49:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/emp_dir/
17/07/11 13:50:17 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 29 2017-07-11 13:49 /user/acadgild/emp_dir/emp
[acadgild@localhost ~]$ mysql -u root -p
Enter password:
Welcome to the MySQL monitor. Commands end with ; or \g.
Your MySQL connection id is 69
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.

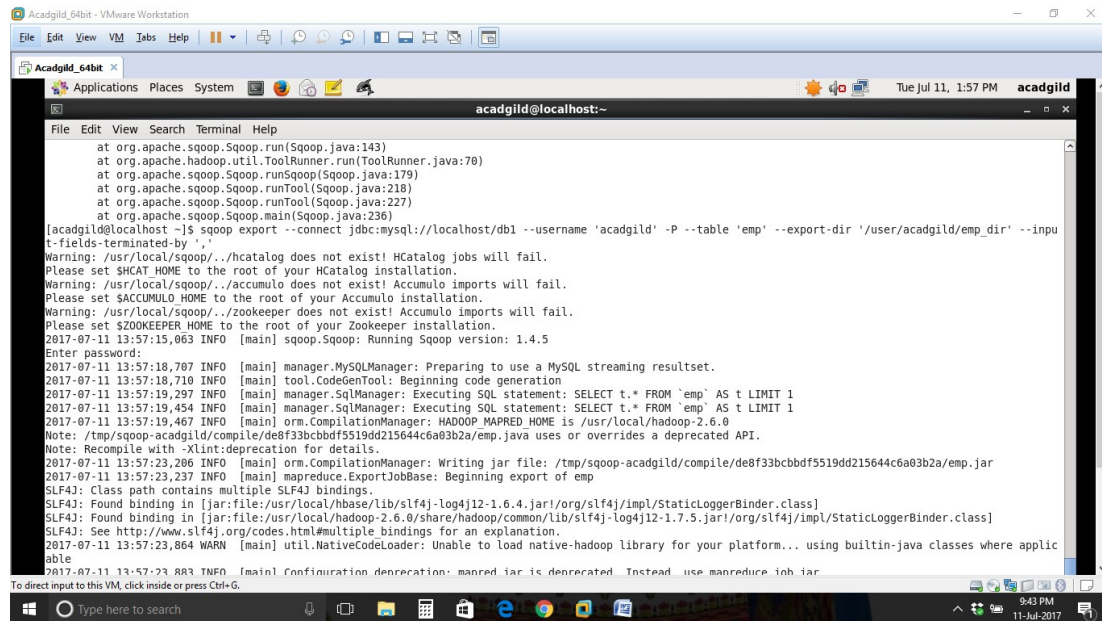
mysql> use db1;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> desc emp;
+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+
| id | int(11) | NO | PRI | NULL | |
| name | varchar(50) | YES | | NULL | |
+-----+-----+-----+-----+-----+
2 rows in set (0.00 sec)

mysql>
```

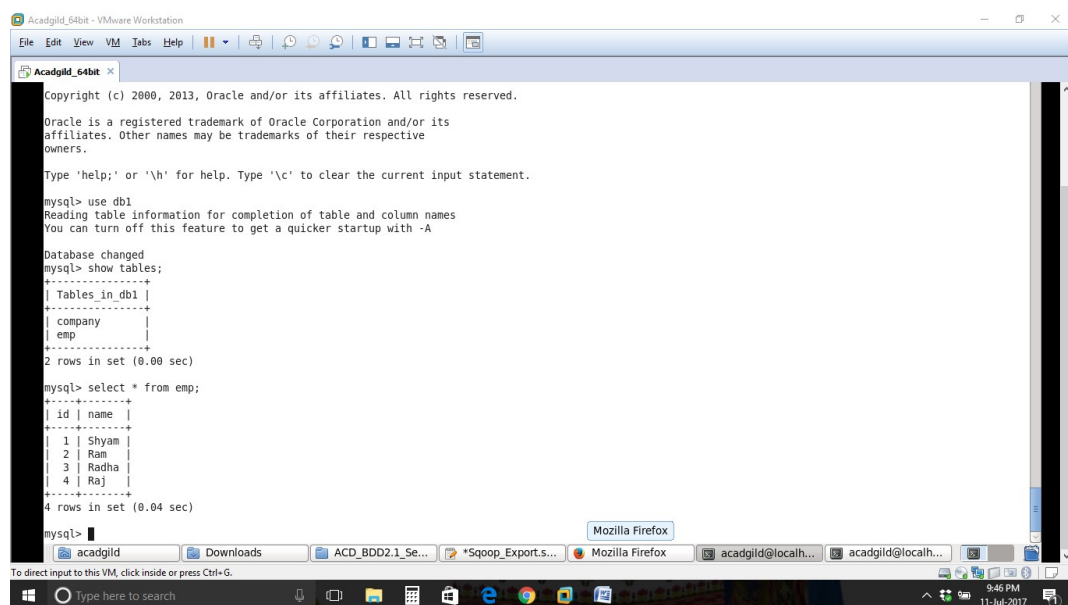
Sqoop command to export data from hdfs to mysql

Sqoop export --connect jdbc:mysql://localhost/db1 --username 'acagdild' --P --table 'emp' --export-dir '/user/acagdild/emp_dir' --input-fields-terminated-by ','



```
acagdild@localhost:~$ sqoop export --connect jdbc:mysql://localhost/db1 --username 'acagdild' --P --table 'emp' --export-dir '/user/acagdild/emp_dir' --input-fields-terminated-by ','
Warning: /usr/local/sqoop/../hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/../accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/../zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-07-11 13:57:15,063 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.5
Enter password:
2017-07-11 13:57:18,707 INFO [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-07-11 13:57:18,710 INFO [main] tool.CodeGenTool: Beginning code generation
2017-07-11 13:57:19,297 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'emp' AS t LIMIT 1
2017-07-11 13:57:19,454 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'emp' AS t LIMIT 1
2017-07-11 13:57:19,467 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop-2.6.0
Note: /tmp/sqoop-acagdild/compile/de8f33bcbbdf5519dd215644c6a03b2a/emp.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
2017-07-11 13:57:23,206 INFO [main] orm.CompilationManager: Writing jar file: /tmp/sqoop-acagdild/compile/de8f33bcbbdf5519dd215644c6a03b2a/emp.jar
2017-07-11 13:57:23,237 INFO [main] mapreduce.ExportJobBase: Beginning export of emp
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/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.
2017-07-11 13:57:23,864 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2017-07-11 13:57:23,883 INFO [main] Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
To direct input to this VM, click inside or press Ctrl-G.
```

Data exported from hdfs to mysql



```
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.

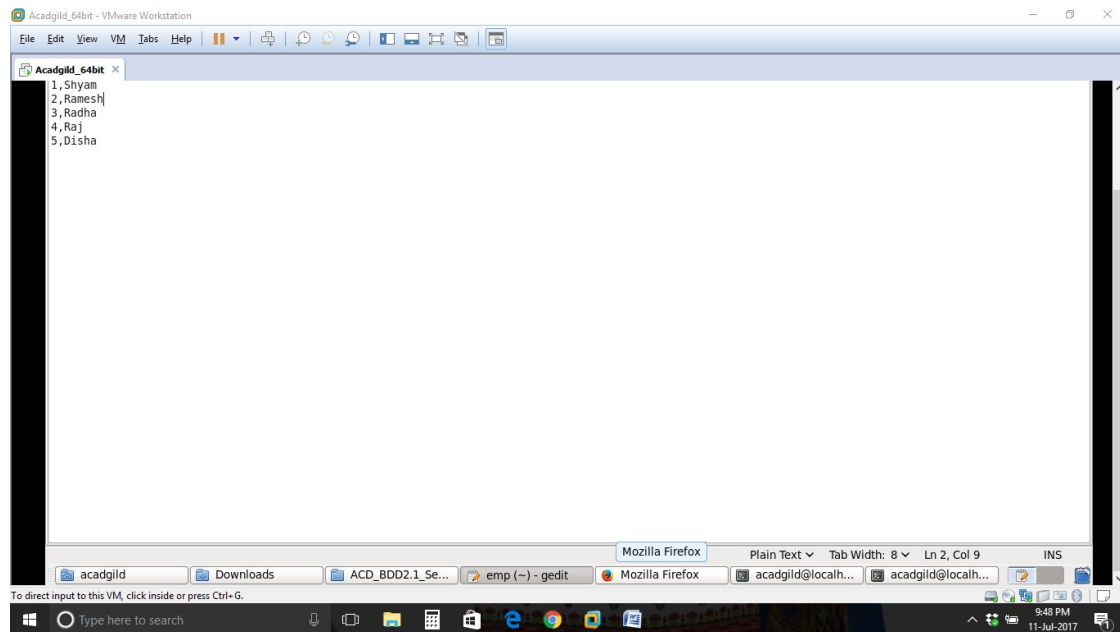
mysql> use db1
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+-----+
| Tables_in_db1 |
+-----+
| company       |
| emp           |
+-----+
2 rows in set (0.00 sec)

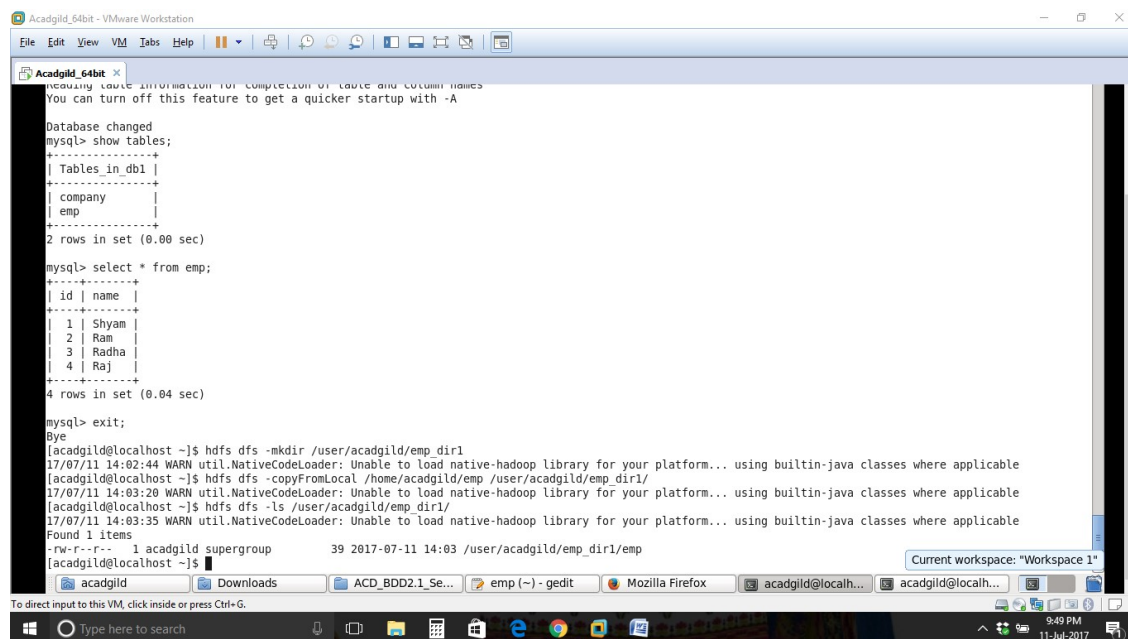
mysql> select * from emp;
+----+-----+
| id | name |
+----+-----+
| 1  | Shyam |
| 2  | Ram   |
| 3  | Radha |
| 4  | Raj   |
+----+-----+
4 rows in set (0.04 sec)

mysql>
```

Emp file in local system is changed

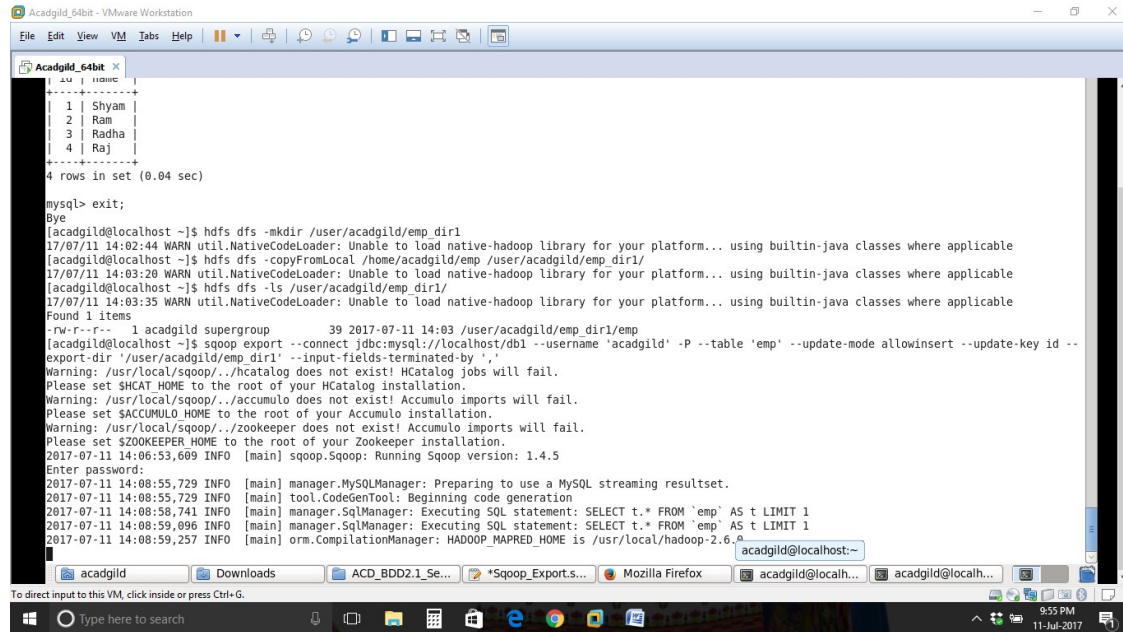


Changed emp file copied into hdfs



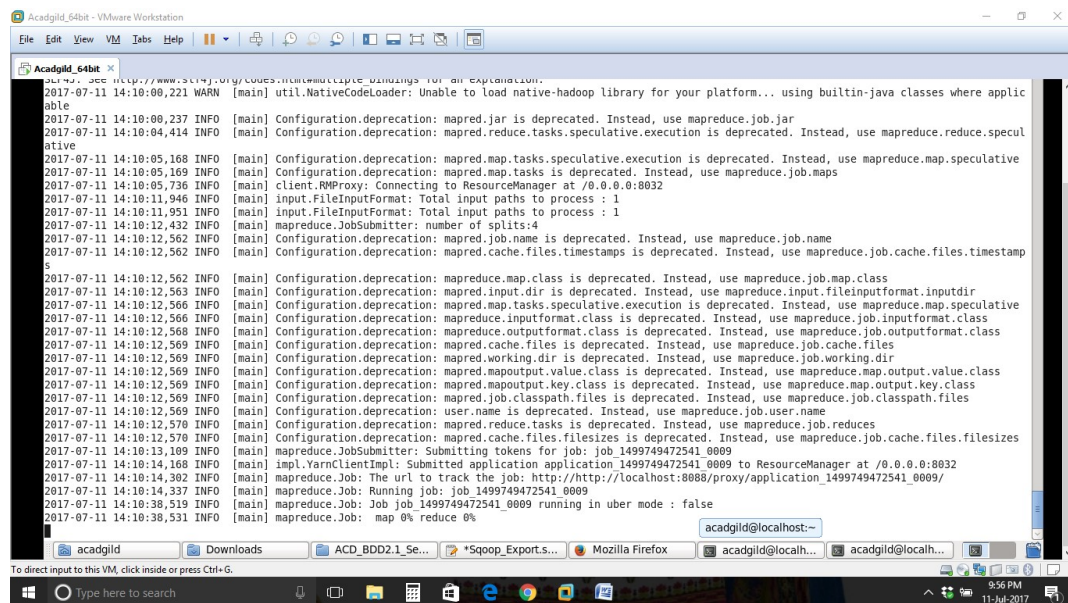
Sqoop command to do upsert (update if row exists otherwise insert)

Sqoop export --connect jdbc:mysql://localhost/db1 --username 'acadgild' --P --table 'emp' --update-mode allowinsert --update-key id --export-dir '/user/acadgild/emp_dir1' --input-fields-terminated-by ','



```
Acadgild_64bit
1 | Shyam
2 | Ram
3 | Radha
4 | Raj
4 rows in set (0.04 sec)

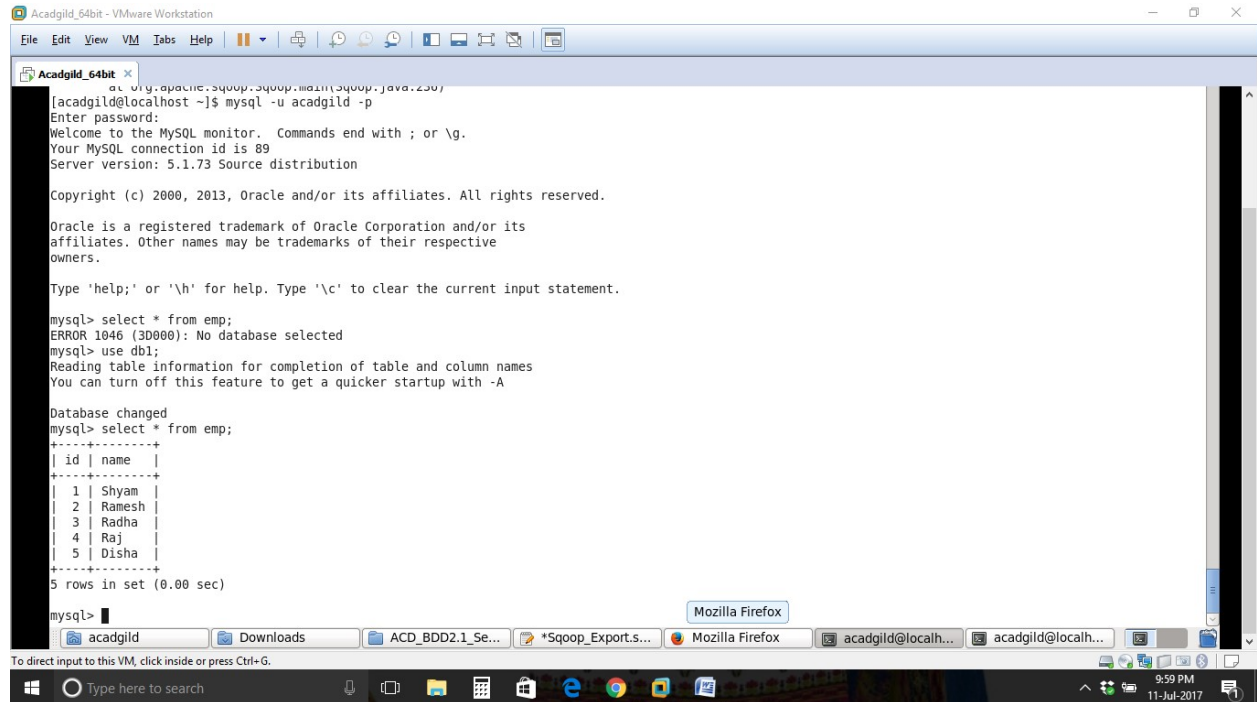
mysql> exit;
Bye
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/emp_dir1
17/07/11 14:02:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/emp_dir1/
17/07/11 14:03:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/emp_dir1/
17/07/11 14:03:35 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 39 2017-07-11 14:03 /user/acadgild/emp_dir1/emp
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username 'acadgild' --P --table 'emp' --update-mode allowinsert --update-key id --
export-dir '/user/acadgild/emp_dir1' --input-fields-terminated-by ','
Warning: /usr/local/sqoop/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-07-11 14:06:53,609 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.5
Enter password:
2017-07-11 14:08:55,729 INFO [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-07-11 14:08:55,729 INFO [main] tool.CodeGenTool: Beginning code generation
2017-07-11 14:08:58,741 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'emp' AS t LIMIT 1
2017-07-11 14:08:59,896 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'emp' AS t LIMIT 1
2017-07-11 14:08:59,257 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop-2.6.0
[acadgild@localhost ~]$
```



```
Acadgild_64bit
1 | Shyam
2 | Ram
3 | Radha
4 | Raj
4 rows in set (0.04 sec)

mysql> exit;
Bye
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/emp_dir1
17/07/11 14:02:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/emp_dir1/
17/07/11 14:03:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/emp_dir1/
17/07/11 14:03:35 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 39 2017-07-11 14:03 /user/acadgild/emp_dir1/emp
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/db1 --username 'acadgild' --P --table 'emp' --update-mode allowinsert --update-key id --
export-dir '/user/acadgild/emp_dir1' --input-fields-terminated-by ','
Warning: /usr/local/sqoop/./hcatalog does not exist! HCatalog jobs will fail.
Please set $HCAT_HOME to the root of your HCatalog installation.
Warning: /usr/local/sqoop/./accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
Warning: /usr/local/sqoop/./zookeeper does not exist! Accumulo imports will fail.
Please set $ZOOKEEPER_HOME to the root of your Zookeeper installation.
2017-07-11 14:06:53,609 INFO [main] sqoop.Sqoop: Running Sqoop version: 1.4.5
Enter password:
2017-07-11 14:08:55,729 INFO [main] manager.MySQLManager: Preparing to use a MySQL streaming resultset.
2017-07-11 14:08:55,729 INFO [main] tool.CodeGenTool: Beginning code generation
2017-07-11 14:08:58,741 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'emp' AS t LIMIT 1
2017-07-11 14:08:59,896 INFO [main] manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'emp' AS t LIMIT 1
2017-07-11 14:08:59,257 INFO [main] orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/local/hadoop-2.6.0
[acadgild@localhost ~]$
```

Select query shows existing row (row no 2) is updated and new rows (row 5) is inserted



```
acadmild@localhost ~]$ mysql -u acadmild -p
Enter password:
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 89
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> select * from emp;
ERROR 1046 (3D000): No database selected
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

Database changed
mysql> select * from emp;
+----+-----+
| id | name |
+----+-----+
| 1  | Shyam |
| 2  | Ramesh |
| 3  | Radha |
| 4  | Raj |
| 5  | Disha |
+----+-----+
5 rows in set (0.00 sec)

mysql>
```