

Big data and Hadoop – Assignment 2

Task 3: Data ingestion using SQOOP tool

1. Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table

Step 1: Creating database (simplifiedb) and table inside the database (Person)

```
mysql> create database simplifiedb;
Query OK, 1 row affected (0.01 sec)

mysql> use simplifiedb;
Database changed
mysql> Create table Person(
    -> person_id INT NOT NULL,
    -> lname varchar(20),
    -> fname varchar(20),
    -> area varchar(20),
    -> city varchar(20),
    -> PRIMARY KEY(person_id)
    -> );
Query OK, 0 rows affected (0.02 sec)

mysql> describe Person;
+-----+-----+-----+-----+-----+-----+
| Field | Type | Null | Key | Default | Extra |
+-----+-----+-----+-----+-----+-----+
| person_id | int(11) | NO | PRI | NULL | |
| lname | varchar(20) | YES | | NULL | |
| fname | varchar(20) | YES | | NULL | |
| area | varchar(20) | YES | | NULL | |
| city | varchar(20) | YES | | NULL | |
+-----+-----+-----+-----+-----+-----+
5 rows in set (0.22 sec)
```

Step 2: Inserting data into table Person

```
mysql> insert into Person values(1,'Abhishek','Ram','Patna','Bihar');
Query OK, 1 row affected (0.09 sec)

mysql> insert into Person values(2,'Tanya','Priya','Whitefield','Bangalore');
Query OK, 1 row affected (0.02 sec)

mysql> insert into Person values(3,'James','Brown','New York','United States');
Query OK, 1 row affected (0.00 sec)

mysql> insert into Person values(4,'John','Miller','Los Angeles','United States'
);
Query OK, 1 row affected (0.01 sec)
```

```
mysql> select * from Person;
+-----+-----+-----+-----+-----+-----+
| person_id | lname | fname | area | city |
+-----+-----+-----+-----+-----+-----+
| 1 | Abhishek | Ram | Patna | Bihar |
| 2 | Tanya | Priya | Whitefield | Bangalore |
| 3 | James | Brown | New York | United States |
| 4 | John | Miller | Los Angeles | United States |
+-----+-----+-----+-----+-----+-----+
4 rows in set (0.00 sec)
```

Step 3: Listing databases available and ensuring the connectivity

Big data and Hadoop – Assignment 2

```
[acadgild@localhost ~]$ sqoop list-databases --connect jdbc:mysql://localhost/ --username root --password Root@123
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.
18/07/28 20:36:38 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
18/07/28 20:36:38 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
18/07/28 20:36:39 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
Sat Jul 28 20:36:39 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.
mysql
information_schema
performance_schema
sys
metastore
oozie
simplicdb
simplicdb
db3
```

Step 4: Importing data from table Person into directory /sqoopout

Sqoop import --connect jdbc: mysql://localhost/simplicdb --table Person

--username root --P --target-dir /sqoopout

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/simplicdb -
-table Person --username root -P --target-dir /sqoopout
Warning: /home/acadgild/install/sqoop/sqoop-1.4.6.bin__hadoop-2.0.4-alpha/./hca
talog 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/./acc
umulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
18/07/28 19:02:46 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/07/28 19:02:54 INFO manager.MySQLManager: Preparing to use a MySQL streaming
resultset.
```

```
seconds (2.1838 bytes/sec)
18/07/28 19:04:06 INFO mapreduce.ImportJobBase: Retrieved 4 records.
```

Step 5: listing the parent directory to ensure the /sqoopout directory present

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/07/28 19:04:27 WARN util.NativeCodeLoader: Unable to load native-hadoop libr
ary for your platform... using builtin-java classes where applicable
Found 4 items
drwxr-xr-x - acadgild supergroup 0 2018-07-28 19:04 /sqoopout
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
drwxrwx--- - acadgild supergroup 0 2018-02-09 11:35 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-02-09 14:50 /user
```

Big data and Hadoop – Assignment 2

Step 6: Listing the file present in /sqoopout in which 4 part files has been created for each record in the Person table

```
[acadgild@localhost ~]$ hadoop fs -ls /sqoopout
18/07/28 19:04:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 5 items
-rw-r--r--  1 acadgild supergroup          0 2018-07-28 19:04 /sqoopout/_SUCCESS
-rw-r--r--  1 acadgild supergroup    27 2018-07-28 19:04 /sqoopout/part-m-00000
-rw-r--r--  1 acadgild supergroup    35 2018-07-28 19:04 /sqoopout/part-m-00001
-rw-r--r--  1 acadgild supergroup    37 2018-07-28 19:04 /sqoopout/part-m-00002
-rw-r--r--  1 acadgild supergroup    40 2018-07-28 19:04 /sqoopout/part-m-00003
```

Step 7: To display the content of each part file using “cat” command

```
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout/part-m-00000
18/07/28 19:05:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
1,Abhishek,Ram,Patna,Bihar
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout/part-m-00001
18/07/28 19:05:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2,Tanya,Priya,Whitefield,Bangalore
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout/part-m-00002
18/07/28 19:05:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
3,James,Brown,New York,United States
```

```
[acadgild@localhost ~]$ hadoop fs -cat /sqoopout/part-m-00003
18/07/28 19:05:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
4,John,Miller,Los Angeles,United States
```

Step 8: Checking table “Person” in Mysql which should be empty

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

Big data and Hadoop – Assignment 2

Step 9: Exporting data present in SQOOPOUT folder into database

```
sqoop export --connect jdbc:mysql://localhost/simplydb --username 'root' -P --table 'Person' --  
export-dir '/sqoopout' --input-fields-terminated-by ',' -m 1 --columns  
person_id,lname,fname,area,city
```

```
[acadgild@localhost ~]$ sqoop export --connect jdbc:mysql://localhost/simplydb --username 'root' -P --table 'Person' --export-dir '/sqoopout' --input-fields-terminated-  
by ',' -m 1 --columns person_id,lname,fname,area,city  
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.  
18/07/28 21:11:10 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6  
Enter password:  
18/07/28 21:11:18 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.  
18/07/28 21:11:18 INFO tool.CodeGenTool: Beginning code generation  
Sat Jul 28 21:11:18 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 verifyServerCertif  
icate 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 ve  
rification.  
18/07/28 21:11:19 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1  
18/07/28 21:11:19 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1  
18/07/28 21:11:19 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5  
Note: /tmp/sqoop-acadgild/compile/Of5balblec8cfe07f837f25a618d1e82/Person.java uses or overrides a deprecated API.  
Note: Recompile with -Xlint:deprecation for details.  
18/07/28 21:11:23 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/Of5balblec8cfe07f837f25a618d1e82/Person.jar  
18/07/28 21:11:23 INFO mapreduce.ExportJobBase: Beginning export of Person  
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]  
18/07/28 21:11:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
18/07/28 21:11:24 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar  
18/07/28 21:11:24 INFO mapreduce.ExportJobBase: Transferred 455 bytes in 26.5268 seconds (17.1525 bytes/sec)  
18/07/28 21:11:24 INFO mapreduce.ExportJobBase: Exported 4 records.  
You have new mail in /var/spool/mail/acadgild
```

Step 10: Checking table “Person” whether data exported from “/sqoopout”

```
mysql> select *from Person;  
+-----+-----+-----+-----+-----+  
| person_id | lname   | fname | area      | city      |  
+-----+-----+-----+-----+-----+  
| 1 | Abhishek | Ram    | Patna     | Bihar     |  
| 2 | Tanya   | Priya  | Whitefield | Bangalore |  
| 3 | James   | Brown  | New York  | United States |  
| 4 | John    | Miller | Los Angeles | United States |  
+-----+-----+-----+-----+-----+  
4 rows in set (0.00 sec)
```

Big data and Hadoop – Assignment 2

2. Use Sqoop tool to import data from the same mysql table where the person_id = 3 into a new hdfs directory SQOOP_FILTER.

Step 1: Executing command to import data with person_id="3" from table "Person" into directory "/SQOOP_FILTER"

Sqoop import --connect jdbc:mysql://localhost/simplydb --table 'Person' --username root -P --target-dir '/SQOOP_FILTER' -m 1 --where person_id=3

```
[acadgild@localhost ~]$ sqoop import --connect jdbc:mysql://localhost/simplydb --table 'Person' --username root -P --target-dir '/SQOOP_FILTER' -m 1 --where person_id=3
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.
18/07/28 21:34:12 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/07/28 21:34:21 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/07/28 21:34:21 INFO tool.CodeGenTool: Beginning code generation
Sat Jul 28 21:34:22 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.
18/07/28 21:34:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1
18/07/28 21:34:23 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'Person' AS t LIMIT 1
18/07/28 21:34:23 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/3eda467055d0e26dde8b01251296081f/Person.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/07/28 21:34:26 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/3eda467055d0e26dde8b01251296081f/Person.jar
18/07/28 21:34:26 WARN manager.MySQLManager: It looks like you are importing from mysql.
18/07/28 21:34:26 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
18/07/28 21:34:26 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
18/07/28 21:34:26 INFO manager.MySQLManager: Setting zero DATETIME behavior to convertToNull (mysql)
18/07/28 21:34:26 INFO mapreduce.ImportJobBase: Beginning import of Person
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]
18/07/28 21:34:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/28 21:34:26 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/07/28 21:34:27 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/07/28 21:34:27 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
Sat Jul 28 21:34:30 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.
18/07/28 21:34:30 INFO db.DBInputFormat: Using read committed transaction isolation
18/07/28 21:34:30 INFO mapreduce.JobSubmitter: number of splits:1
18/07/28 21:34:30 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532773085741_0003
18/07/28 21:34:31 INFO impl.YarnClientImpl: Submitted application application_1532773085741_0003
18/07/28 21:34:31 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532773085741_0003/
```

Output: one record retrieved into "/SQOOP_FILTER" directory

```
Bytes Written=37
18/07/28 21:34:51 INFO mapreduce.ImportJobBase: Transferred 37 bytes in 23.814 seconds (1.5537 bytes/sec)
18/07/28 21:34:51 INFO mapreduce.ImportJobBase: Retrieved 1 records.
You have new mail in /var/spool/mail/acadgild
```

Step 2: listing the directories under parent directory "/" to ensure "/SQOOP_FILTER" directory has been created.

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/07/28 21:35:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 5 items
drwxr-xr-x - acadgild supergroup 0 2018-07-28 21:34 /SQOOP_FILTER
drwxr-xr-x - acadgild supergroup 0 2018-07-28 19:04 /sqoopout
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
drwxrwx--- - acadgild supergroup 0 2018-02-09 11:35 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-02-09 14:50 /user
```

Big data and Hadoop – Assignment 2

Step 3: Listing the files created under directory “/SQOOP_FILTER”

OUTPUT: in which two files has been created - /SQOOP_FILTER/_SUCCESS and part file /SQOOP_FILTER/part-m-00000

```
[acadgild@localhost ~]$ hadoop fs -ls /SQOOP_FILTER
18/07/28 21:35:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform.
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-07-28 21:34 /SQOOP_FILTER/_SUCCESS
-rw-r--r--  1 acadgild supergroup       37 2018-07-28 21:34 /SQOOP_FILTER/part-m-00000
```

Step 4: Displaying the content present in /SQOOP_FILTER/part-m-00000 using “cat” command.

OUTPUT: data with respect to person_id="3" of “Person” table.

```
[acadgild@localhost ~]$ hadoop fs -cat /SQOOP_FILTER/part-m-00000
18/07/28 21:35:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
3,James,Brown,New York,United States
```


Big data and Hadoop – Assignment 2

Task 2:

Dataset is sample data of songs heard by users on an online streaming platform. The Description of data set attached in musicdata.txt is as follows: - Col_1 - UserId Col_2 - TrackId Col_3 - Songs Share status (1 for shared, 0 for not shared) Col_4 - Listening Platform (Radio or Web - 0 for radio, 1 for web) Col_5 - Song Listening Status (0 for skipped, 1 for fully heard) 1. Write MapReduce program for following tasks.

Task 1 Find the number of unique listeners in the data set.

Output for Unique Listener

Step 1: Checking data in musicdata.txt

```
[acadgild@localhost ~]$ hadoop fs -cat /musicdata.txt
18/07/30 22:55:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111115|222|0|1|0
111113|225|1|0|0
111117|223|0|1|1
111115|225|1|0|0
```

Step 2: Executing the command to get output

```
[acadgild@localhost ~]$ hadoop jar mymusic.jar com.music.musicDemo /musicdata.txt /music_out.txt
18/07/30 22:56:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/30 22:56:03 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/07/30 22:56:04 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/07/30 22:56:05 INFO input.FileInputFormat: Total input paths to process : 1
18/07/30 22:56:05 INFO mapreduce.JobSubmitter: number of splits:1
18/07/30 22:56:06 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1532970769007_0004
18/07/30 22:56:07 INFO impl.YarnClientImpl: Submitted application application_1532970769007_0004
18/07/30 22:56:07 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1532970769007_0004/
18/07/30 22:56:07 INFO mapreduce.Job: Running job: job_1532970769007_0004
18/07/30 22:56:23 INFO mapreduce.Job: Job job_1532970769007_0004 running in uber
```

Step 3: Executing command , two part files has been created

```
[acadgild@localhost ~]$ hadoop fs -ls /music_out.txt
18/07/30 22:57:31 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-07-30 22:56 /music_out.txt/_SUCCESS
-rw-r--r--  1 acadgild supergroup       27 2018-07-30 22:56 /music_out.txt/pa
rt-r-00000
```

Big data and Hadoop – Assignment 2

Step 4: To get output for unique listeners

```
[acadgild@localhost ~]$ hadoop fs -cat /music_out.txt/part-r-00000
18/07/30 22:57:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111113 1
111115 2
111117 1
You have new mail in /var/spool/mail/acadgild
```

Task 2 : What are the number of times a song was heard fully.

Output for song listened fully

Step 1: Executing the command to get output

```
[acadgild@localhost ~]$ hadoop jar songcount.jar com.song.songCount /musicdata.txt /song_out.txt
18/08/01 04:05:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/01 04:05:52 INFO client.RMPProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/01 04:05:53 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/08/01 04:05:54 INFO input.FileInputFormat: Total input paths to process : 1
18/08/01 04:05:54 INFO mapreduce.JobSubmitter: number of splits:1
18/08/01 04:05:55 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533076255605_0001
18/08/01 04:05:55 INFO impl.YarnClientImpl: Submitted application application_1533076255605_0001
18/08/01 04:05:56 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533076255605_0001/
18/08/01 04:05:56 INFO mapreduce.Job: Running job: job_1533076255605_0001
18/08/01 04:06:10 INFO mapreduce.Job: Job job_1533076255605_0001 running in uber mode : false
18/08/01 04:06:10 INFO mapreduce.Job: map 0% reduce 0%
18/08/01 04:06:17 INFO mapreduce.Job: map 100% reduce 0%
18/08/01 04:06:26 INFO mapreduce.Job: map 100% reduce 100%
```

Step 2: Executing command , two part files has been created

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/08/01 04:06:59 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 2018-07-28 21:34 /SQOOP_FILTER
-rw-r--r-- 1 acadgild supergroup 72 2018-07-30 22:48 /music.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-30 22:56 /music_out.txt
-rw-r--r-- 1 acadgild supergroup 72 2018-07-30 22:47 /musicdata.txt
drwxr-xr-x - acadgild supergroup 0 2018-08-01 04:06 /song_out.txt
drwxr-xr-x - acadgild supergroup 0 2018-07-28 19:04 /sqoopout
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
drwxrwx--- - acadgild supergroup 0 2018-02-09 11:35 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-02-09 14:50 /user
```


Big data and Hadoop – Assignment 2

```
[acadgild@localhost ~]$ hadoop fs -ls /song_out.txt
18/08/01 04:12:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup          0 2018-08-01 04:06 /song_out.txt/_SUCCESS
-rw-r--r--  1 acadgild supergroup          6 2018-08-01 04:06 /song_out.txt/part-r-000000
```

Step 3: output

```
[acadgild@localhost ~]$ hadoop fs -cat /song_out.txt/part-r-000000
18/08/01 04:12:48 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
223      1
```

```
[acadgild@localhost ~]$ hadoop fs -cat /musicdata.txt
18/08/01 04:14:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111115|222|0|1|0
111113|225|1|0|0
111117|223|0|1|1
111115|225|1|0|0
```

TASK 3 What are the number of times a song was shared.

Output for song shared

Step 1: Executing the command to get output

```
[acadgild@localhost ~]$ hadoop jar sharedSongs.jar com.shared.sharedTimes /musicdata.txt /share_out.txt
18/08/01 04:48:18 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/01 04:48:19 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/08/01 04:48:20 WARN mapreduce.JobResourceUploader: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.
18/08/01 04:48:20 INFO input.FileInputFormat: Total input paths to process : 1
18/08/01 04:48:21 INFO mapreduce.JobSubmitter: number of splits:1
18/08/01 04:48:21 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1533076255605_0003
18/08/01 04:48:21 INFO impl.YarnClientImpl: Submitted application application_1533076255605_0003
18/08/01 04:48:21 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1533076255605_0003/
18/08/01 04:48:21 INFO mapreduce.Job: Running job: job_1533076255605_0003
18/08/01 04:48:30 INFO mapreduce.Job: Job job_1533076255605_0003 running in uber mode : false
18/08/01 04:48:30 INFO mapreduce.Job:  map 0% reduce 0%
18/08/01 04:48:39 INFO mapreduce.Job:  map 100% reduce 0%
18/08/01 04:48:46 INFO mapreduce.Job:  map 100% reduce 100%
18/08/01 04:48:47 INFO mapreduce.Job: Job job_1533076255605_0003 completed successfully
```

Big data and Hadoop – Assignment 2

Step 2: Executing command , two part files has been created

```
[acadgild@localhost ~]$ hadoop fs -ls /
18/08/01 04:48:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 11 items
drwxr-xr-x  - acadgild supergroup      0 2018-07-28 21:34 /SQOOP_FILTER
-rw-r--r--  1 acadgild supergroup     72 2018-07-30 22:48 /music.txt
drwxr-xr-x  - acadgild supergroup      0 2018-07-30 22:56 /music_out.txt
-rw-r--r--  1 acadgild supergroup     72 2018-07-30 22:47 /musicdata.txt
drwxr-xr-x  - acadgild supergroup      0 2018-08-01 04:48 /share_out.txt
drwxr-xr-x  - acadgild supergroup      0 2018-08-01 04:10 /song_out
drwxr-xr-x  - acadgild supergroup      0 2018-08-01 04:06 /song_out.txt
drwxr-xr-x  - acadgild supergroup      0 2018-07-28 19:04 /sqoopout
drwxr-xr-x  - acadgild supergroup      0 2018-02-02 12:49 /sqoopout111
drwxrwx---  - acadgild supergroup      0 2018-02-09 11:35 /tmp
drwxr-xr-x  - acadgild supergroup      0 2018-02-09 14:50 /user
```

```
[acadgild@localhost ~]$ hadoop fs -ls /share_out.txt
18/08/01 04:49:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup      0 2018-08-01 04:48 /share_out.txt/_SUCCESS
-rw-r--r--  1 acadgild supergroup      6 2018-08-01 04:48 /share_out.txt/part-r-000000
```

Step 3: output

```
[acadgild@localhost ~]$ hadoop fs -cat /share_out.txt/part-r-000000
18/08/01 04:49:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
225      2
```

```
[acadgild@localhost ~]$ hadoop fs -cat /musicdata.txt
18/08/01 04:50:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
111115|222|0|1|0
111113|225|1|0|0
111117|223|0|1|1
111115|225|1|0|0
```