Session 3

ADVANCE MAP REDUCE AND
INTRODUCTION TO UNIX CONCEPTS
DATA INGESTION TOOL SQOOP

Task 1.1
Find the number of unique listeners in the data set.

```
hadoop jar map-reduce-music-1.0.jar uniqueListener /music/musicdata.txt /musicout
```

Task 1.2
What are the number of times a song was heard fully.

```
hadoop jar map-reduce-music-1.0.jar songHeared /music/musicdata.txt /musicout
```

Task 1.3
What are the number of times a song was shared.

```
hadoop jar map-reduce-music-1.0.jar songShared /music/musicdata.txt /musicout
```

Task 2.1
Use Sqoop tool to export data present in SQOOPOUT folder made while demo of Import table

```
sqoop export --connect jdbc:mysql://localhost/simplidb --table Person --username root -P --export-
dir /sqoopout
```

```
[acadgild@localhost manish]$ sqoop export --connect jdbc:mysql://localhost/simplidb --table Person --username root -P --export-dir /sqoopout
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.
18/11/15 22:11:18 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/11/15 22:11:46 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/11/15 22:11:46 INFO tool.CodeGenTool: Beginning code generation
Thu Nov 15 22:11:47 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+ req
 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 ne
SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
18/11/15 22:11:49 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Person` AS t LIMIT 1
18/11/15 22:11:49 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM `Person` AS t LIMIT 1
18/11/15 22:11:49 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/d8b9fb2ba9d338e214a8f35b58fab6cf/Person.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/11/15 22:11:57 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/d8b9fb2ba9d338e214a8f35b58fab6cf/Person.jar
18/11/15 22:11:57 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]
18/11/15 22:11:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/11/15 22:11:57 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/11/15 22:12:00 INFO Configuration.deprecation: mapred.reduce.tasks.speculative.execution is deprecated. Instead, use mapreduce.reduce.speculative
18/11/15 22:12:00 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
18/11/15 22:12:00 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/11/15 22:12:01 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
18/11/15 22:12:07 INFO input.FileInputFormat: Total input paths to process : 4
18/11/15 22:12:07 INFO input.FileInputFormat: Total input paths to process : 4
18/11/15 22:12:07 INFO mapreduce.JobSubmitter: number of splits:4
18/11/15 22:12:07 INFO Configuration.deprecation: mapred.map.tasks.speculative.execution is deprecated. Instead, use mapreduce.map.speculative
18/11/15 22:12:08 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1542293665251_0006
18/11/15 22:12:08 INFO impl.YarnClientImpl: Submitted application application_1542293665251_0006
18/11/15 22:12:09 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1542293665251_0006/
18/11/15 22:12:09 INFO mapreduce.Job: Running job: job_1542293665251_0006
18/11/15 22:12:28 INFO mapreduce.Job: Job job_1542293665251_0006 running in uber mode : false
18/11/15 22:12:28 INFO mapreduce.Job:  map 0% reduce 0%
18/11/15 22:13:17 INFO mapreduce.Job:  map 25% reduce 0%
18/11/15 22:13:18 INFO mapreduce.Job:  map 100% reduce 0%
18/11/15 22:13:19 INFO mapreduce.Job: Job job_1542293665251_0006 completed successfully
18/11/15 22:13:20 INFO mapreduce.Job: Counters: 31
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=510164
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=893
                HDFS: Number of bytes written=0
                HDFS: Number of read operations=25
                HDFS: Number of large read operations=0
```

```
[acadgild@localhost manish]$ hadoop fs -ls /sqoopout
18/11/15 22:07:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 4 items
-rw-r--r--   1 acadgild supergroup         24 2018-11-15 22:06 /sqoopout/part-m-00000
-rw-r--r--   1 acadgild supergroup         35 2018-11-15 22:06 /sqoopout/part-m-00001
-rw-r--r--   1 acadgild supergroup         37 2018-11-15 22:06 /sqoopout/part-m-00002
-rw-r--r--   1 acadgild supergroup         42 2018-11-15 22:06 /sqoopout/part-m-00003
[acadgild@localhost manish]$ hadoop fs -cat /sqoopout/part*
18/11/15 22:08:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Patna,Bihar,Ram,Shyam,1
Whitefiled,Bangalore,Priya,Tanya,2
New York,United States,Brown,James,3
 Los Angeles ,United States,Miller,Jhon,4
You have new mail in /var/spool/mail/acadgild
```

```
[mysql> delete from Person;
 Query OK, 4 rows affected (2.45 sec)

[mysql> delete from Person;
 Query OK, 0 rows affected (0.00 sec)

[mysql> select * from Person;
+-----------+--------+--------+---------------+---------------+
| person_id | lname  | fname  | area          | city          |
+-----------+--------+--------+---------------+---------------+
|         1 | Shyam  | Ram    | Patna         | Bihar         |
|         2 | Tanya  | Priya  | Whitefiled    | Bangalore     |
|         3 | James  | Brown  | New York      | United States |
|         4 | Jhon   | Miller |  Los Angeles  | United States |
+-----------+--------+--------+---------------+---------------+
4 rows in set (0.01 sec)
```

Task 2.2
Use Sqoop tool to import data present in SQOOPOUT folder made while demo of Import table with parameter person_id =3.

sqoop import --connect jdbc:mysql://localhost/simplidb --username root -P --m 1 --query 'select *

from Person where $CONDITIONS and person_id = 3' --target-dir /sqoopout

```
[acadgild@localhost manish]$ sqoop import --connect jdbc:mysql://localhost/simpldb --username root -P --m 1 --query 'select * from Person where $CONDITIONS and person_id = 3' --target-dir /sqoopout
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.
18/11/15 22:32:43 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6
Enter password:
18/11/15 22:32:51 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
18/11/15 22:32:51 INFO tool.CodeGenTool: Beginning code generation
Thu Nov 15 22:32:52 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.
18/11/15 22:32:54 INFO manager.SqlManager: Executing SQL statement: select * from Person where  (1 = 0)  and person_id = 3
18/11/15 22:32:54 INFO manager.SqlManager: Executing SQL statement: select * from Person where  (1 = 0)  and person_id = 3
18/11/15 22:32:54 INFO manager.SqlManager: Executing SQL statement: select * from Person where  (1 = 0)  and person_id = 3
18/11/15 22:32:54 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /home/acadgild/install/hadoop/hadoop-2.6.5
Note: /tmp/sqoop-acadgild/compile/6d43c3ac5fd2b054544721ac6f61aef4/QueryResult.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
18/11/15 22:33:01 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-acadgild/compile/6d43c3ac5fd2b054544721ac6f61aef4/QueryResult.jar
18/11/15 22:33:01 INFO mapreduce.ImportJobBase: Beginning query import.
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]
18/11/15 22:33:01 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/11/15 22:33:02 INFO Configuration.deprecation: mapred.jar is deprecated. Instead, use mapreduce.job.jar
18/11/15 22:33:03 INFO Configuration.deprecation: mapred.map.tasks is deprecated. Instead, use mapreduce.job.maps
18/11/15 22:33:04 INFO client.RMProxy: Connecting to ResourceManager at localhost/127.0.0.1:8032
Thu Nov 15 22:33:09 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.
18/11/15 22:33:10 INFO db.DBInputFormat: Using read commited transaction isolation
18/11/15 22:33:10 INFO mapreduce.JobSubmitter: number of splits:1
18/11/15 22:33:10 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1542293665251_0007
18/11/15 22:33:11 INFO impl.YarnClientImpl: Submitted application application_1542293665251_0007
18/11/15 22:33:11 INFO mapreduce.Job: The url to track the job: http://localhost:8088/proxy/application_1542293665251_0007/
18/11/15 22:33:11 INFO mapreduce.Job: Running job: job_1542293665251_0007
18/11/15 22:33:29 INFO mapreduce.Job: Job job_1542293665251_0007 running in uber mode : false
18/11/15 22:33:29 INFO mapreduce.Job:  map 0% reduce 0%
18/11/15 22:33:44 INFO mapreduce.Job:  map 100% reduce 0%
18/11/15 22:33:44 INFO mapreduce.Job: Job job_1542293665251_0007 completed successfully
18/11/15 22:33:45 INFO mapreduce.Job: Counters: 30
        File System Counters
                FILE: Number of bytes read=0
                FILE: Number of bytes written=127557
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=87
                HDFS: Number of bytes written=37
                HDFS: Number of read operations=4
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
```

```
[acadgild@localhost manish]$ hadoop fs -ls /sqoopout
18/11/15 22:35:17 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 2018-11-15 22:33 /sqoopout/_SUCCESS
-rw-r--r--   1 acadgild supergroup         37 2018-11-15 22:33 /sqoopout/part-m-00000
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost manish]$ hadoop fs -cat /sqoopout/part*
18/11/15 22:35:39 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 manish]$
```