

Music Data Analysis

1. We have created a batch file “start-daemon.sh” which starts the daemons such as hive, hbase, Mysql and rest of the all hadoop daemons.

Batch file script:



2. As per the batch file script all the hadoop daemons and the Hive, MySQL and Hive daemons are started shown in the below screenshot:

```
[acadgild@localhost music]$  
[acadgild@localhost music]$ ./music_project_master.sh  
Preparing to execute python scripts to generate data...  
Data Generated Successfully !  
Starting the daemons....  
6388 SecondaryNameNode  
6628 NodeManager  
7205 HMaster  
8983 Main  
30856 Jps  
28296 RunJar  
6185 DataNode  
6090 NameNode  
7307 HRegionServer  
7116 HQuorumPeer  
6526 ResourceManager  
7423 JobHistoryServer  
All hadoop daemons started !
```

3. By using the “populate-lookup.sh” script we will create lookup tables in Hbase. These tables have to be used in Data formatting, Data enrichment and analysis stage. The lookup tables are as follows:

- station-geomap

Contains mapping of a geo_cd with station_id . Related file: stn-geocd.txt

- **subscribedusers**

Contains user_id, subscription_start_date and subscription_end_date. Contains details only for subscribed users. Related file: user-subscn.txt

- **song-artistmap**

Contains mapping of song_id with artist_id Along with royalty associated with each play of the song. Related file: song-artist.txt

- **user-artistmap**

Contains an array of artist_id(s) followed by a user_id. Related file: user-artist.txt

“populate-lookup.sh” script



populate-lookup.sh

The “populate-lookup.sh” shell script creates the above 4 lookup tables in the Hbase and populate the data into the lookup tables from the dataset files.

In the below screen shots, we can see the create-lookup.sh scripts and the following screen shots shows the tables creation and population of the data in the Hbase. Also, the values loaded into the Hbase Tables are also shown, please see the below screen shots.

```
[acadgild@localhost music]$ ./music_project_master.sh
Preparing to execute python scripts to generate data...
Data Generated Successfully !
Starting the daemons...
6388 SecondaryNameNode
6628 NodeManager
7205 HMaster
8983 Main
8985 Jps
2008 DataNode
6185 DataNode
6090 NameNode
7307 HRegionServer
7116 HQuorumPeer
6526 ResourceManager
7423 JobHistoryServer
All hadoop daemons started !
Upload the look up tables now in Hbase...
2018-06-17 22:42:30,562 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

disable 'station-geo-map'
0 row(s) in 4.3540 seconds

2018-06-17 22:43:06,633 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
```

```
0 row(s) in 4.3540 seconds

2018-06-17 22:43:06,633 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

drop 'station-geo-map'
0 row(s) in 3.5510 seconds

2018-06-17 22:43:40,444 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

disable 'subscribed-users'
0 row(s) in 4.5890 seconds

2018-06-17 22:44:17,766 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```
2.127.0.0.1(acadgild) x 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)
2018-06-17 22:44:17.766 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

drop 'subscribed-users'
0 row(s) in 4.9630 seconds

2018-06-17 22:44:54.862 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

disable 'song-artist-map'
0 row(s) in 4.3200 seconds

2018-06-17 22:45:30.236 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.class]
```

```
2.127.0.0.1(acadgild) x 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

drop 'song-artist-map'
0 row(s) in 3.3700 seconds

2018-06-17 22:46:02.560 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

create 'station-geo-map', 'geo'
0 row(s) in 3.8890 seconds

Hbase::Table - station-geo-map
2018-06-17 22:46:40.772 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
```

```
2.127.0.0.1(acadgild) x 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)
create 'subscribed-users', 'subscn'
0 row(s) in 2.4250 seconds

Hbase::Table - subscribed-users
2018-06-17 22:47:12.033 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

create 'song-artist-map', 'artist'
0 row(s) in 3.8090 seconds

Hbase::Table - song-artist-map
2018-06-17 22:47:46.194 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'station-geo-map', 'ST400', 'geo:geo_cd', 'A'
0 row(s) in 1.6150 seconds
```

```

put 'station-geo-map', 'ST400', 'geo:geo_cd', 'A'
0 row(s) in 1.6150 seconds

2018-06-17 22:48:16,098 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'station-geo-map', 'ST401', 'geo:geo_cd', 'AU'
0 row(s) in 2.3080 seconds

2018-06-17 22:48:54,522 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'station-geo-map', 'ST402', 'geo:geo_cd', 'AP'
0 row(s) in 1.8710 seconds

2018-06-17 22:49:26,320 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable

```

```

put 'station-geo-map', 'ST403', 'geo:geo_cd', 'J'
0 row(s) in 3.2650 seconds

2018-06-17 22:49:26,320 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'station-geo-map', 'ST404', 'geo:geo_cd', 'E'
0 row(s) in 2.8010 seconds

2018-06-17 22:50:39,396 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.class]

```

```

put 'station-geo-map', 'ST405', 'geo:geo_cd', 'A'
0 row(s) in 1.8410 seconds

2018-06-17 22:51:11,266 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'station-geo-map', 'ST406', 'geo:geo_cd', 'AU'
0 row(s) in 1.7670 seconds

2018-06-17 22:51:55,686 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'station-geo-map', 'ST407', 'geo:geo_cd', 'AP'
0 row(s) in 1.4210 seconds

2018-06-17 22:52:40,974 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable

```

```
View Split MultExec Tunneling Packages Settings Help Xserver Exit
[2.127.0.0.1(acadgild)] [3.127.0.0.1(acadgild)] [4.127.0.0.1(acadgild)]
2018-06-17 22:52:40.974 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/:org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help->RETURN-' for list of supported commands.
Type "exit->RETURN-" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017
put 'station-geo-map', 'ST408', 'geo:geo_cd', 'E'
0 row(s) in 2.3180 seconds

2018-06-17 22:53:25.017 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/:org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help->RETURN-' for list of supported commands.
Type "exit->RETURN-" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017
put 'station-geo-map', 'ST409', 'geo:geo_cd', 'E'
0 row(s) in 4.2270 seconds

2018-06-17 22:54:07.086 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/:org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/impl/StaticLoggerBinder.class]
```

```
View Split MultExec Tunneling Packages Settings Help Xserver Exit
[2.127.0.0.1(acadgild)] [3.127.0.0.1(acadgild)] [4.127.0.0.1(acadgild)]
put 'station-geo-map', 'ST410', 'geo:geo_cd', 'A'
0 row(s) in 1.6840 seconds

2018-06-17 22:54:36.550 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/:org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help->RETURN-' for list of supported commands.
Type "exit->RETURN-" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017
put 'station-geo-map', 'ST411', 'geo:geo_cd', 'A'
0 row(s) in 1.9080 seconds

2018-06-17 22:55:22.168 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/:org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help->RETURN-' for list of supported commands.
Type "exit->RETURN-" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017
put 'station-geo-map', 'ST412', 'geo:geo_cd', 'AP'
0 row(s) in 2.9890 seconds

2018-06-17 22:55:54.398 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```
View Split MultExec Tunneling Packages Settings Help Xserver Exit
[2.127.0.0.1(acadgild)] [3.127.0.0.1(acadgild)] [4.127.0.0.1(acadgild)]
put 'station-geo-map', 'ST413', 'geo:geo_cd', 'J'
0 row(s) in 1.9200 seconds

2018-06-17 22:56:22.409 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/:org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help->RETURN-' for list of supported commands.
Type "exit->RETURN-" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017
put 'station-geo-map', 'ST414', 'geo:geo_cd', 'E'
0 row(s) in 1.4370 seconds

2018-06-17 22:56:50.222 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/:org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help->RETURN-' for list of supported commands.
Type "exit->RETURN-" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017
put 'song-artist-map', 'S200', 'artist:artistid', 'A300'
0 row(s) in 1.6570 seconds

2018-06-17 22:57:18.647 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```
View Split MultExec Tunneling Padages Settings Help Xserver Exit
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)

put 'song-artist-map', 'S201', 'artist:artistid', 'A301'
0 row(s) in 2.0590 seconds

2018-06-17 22:58:02,385 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help:<RETURN>' for list of supported commands.
Type "exit:<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'song-artist-map', 'S202', 'artist:artistid', 'A302'
0 row(s) in 3.5110 seconds

2018-06-17 22:59:25,409 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help:<RETURN>' for list of supported commands.
Type "exit:<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'song-artist-map', 'S203', 'artist:artistid', 'A303'
0 row(s) in 3.1560 seconds

2018-06-17 23:00:25,405 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```
View Split MultExec Tunneling Padages Settings Help Xserver Exit
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)

put 'song-artist-map', 'S204', 'artist:artistid', 'A304'
0 row(s) in 3.2960 seconds

2018-06-17 23:01:24,063 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help:<RETURN>' for list of supported commands.
Type "exit:<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'song-artist-map', 'S205', 'artist:artistid', 'A301'
0 row(s) in 3.4500 seconds

2018-06-17 23:03:33,346 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help:<RETURN>' for list of supported commands.
Type "exit:<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'song-artist-map', 'S206', 'artist:artistid', 'A302'
0 row(s) in 11.4570 seconds

2018-06-17 23:05:30,932 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```
View Split MultExec Tunneling Padages Settings Help Xserver Exit
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)

put 'song-artist-map', 'S207', 'artist:artistid', 'A303'
0 row(s) in 7.6560 seconds

2018-06-17 23:07:44,948 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help:<RETURN>' for list of supported commands.
Type "exit:<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'song-artist-map', 'S208', 'artist:artistid', 'A304'
0 row(s) in 5.0130 seconds

2018-06-17 23:08:47,685 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help:<RETURN>' for list of supported commands.
Type "exit:<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'song-artist-map', 'S209', 'artist:artistid', 'A305'
0 row(s) in 7.6670 seconds

2018-06-17 23:10:27,180 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```
put 'subscribed-users', 'U100', 'subscn:startdt', '1465230523'
0 row(s) in 8.3270 seconds

2018-06-17 23:12:08,885 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'subscribed-users', 'U100', 'subscn:enddt', '1465130523'
0 row(s) in 9.5240 seconds

2018-06-17 23:13:48,996 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'subscribed-users', 'U101', 'subscn:startdt', '1465230523'
0 row(s) in 2.0560 seconds

2018-06-17 23:14:35,781 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```
put 'subscribed-users', 'U101', 'subscn:enddt', '1475130523'
0 row(s) in 4.1590 seconds

2018-06-17 23:15:29,523 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'subscribed-users', 'U102', 'subscn:startdt', '1465230523'
0 row(s) in 1.8520 seconds

2018-06-17 23:16:19,851 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'subscribed-users', 'U102', 'subscn:enddt', '1475130523'
0 row(s) in 4.1000 seconds

2018-06-17 23:16:59,183 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```
put 'subscribed-users', 'U103', 'subscn:startdt', '1465230523'
0 row(s) in 4.9160 seconds

2018-06-17 23:18:05,858 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'subscribed-users', 'U103', 'subscn:enddt', '1475130523'
0 row(s) in 3.5980 seconds

2018-06-17 23:19:23,844 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'subscribed-users', 'U104', 'subscn:startdt', '1465230523'
0 row(s) in 3.9260 seconds

2018-06-17 23:20:26,259 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
```

```

put 'subscribed-users', 'U104', 'subscn:enddt', '1475130523'
0 row(s) in 2.7190 seconds

2018-06-17 23:21:04,485 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'subscribed-users', 'U105', 'subscn:startdt', '1465230523'
0 row(s) in 1.8000 seconds

2018-06-17 23:21:52,313 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/home/acadgild/install/hbase/hbase-1.2.6/lib/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBi
nder.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/slf
4j/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]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'subscribed-users', 'U105', 'subscn:enddt', '1475130523'
0 row(s) in 2.0730 seconds

2018-06-17 23:22:23,364 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-j
ava classes where applicable

```



```

View Split MultiExec Tunneling Packages Settings Help Xserver Exit

SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.6, rUnknown, Mon May 29 02:25:32 CDT 2017

put 'subscribed-users', 'U114', 'subscn:enddt', '1468130523'
0 row(s) in 1.5940 seconds

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/St
aticLoggerBinder.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/slf
4j/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
-log4j.properties Async: true
OK
Time taken: 35.626 seconds
OK
Time taken: 0.077 seconds
OK
Time taken: 2.741 seconds
Loading data to table project.users_artists
OK
Time taken: 5.673 seconds
Done with data population in look up tables !
Lets do some data formatting now...
Data formatting complete !
Creating hive tables on top of hbase tables for data enrichment and filtering...
Hive table with Hbase Mapping Complete !
Let us do data enrichment as per the requirement...
Data Enrichment Complete
Lets run some use cases now...
USE CASES COMPLETE !!
USE have new mail in /var/spool/mail/acadgild
[acadgild@localhost music]$ 

```

Creation of hbase tables

```

hbase(main):003:> list
TABLE
SparkHBaseTable
SparkHBaseTable_new
bulktable
clicks
employee
hbase
hbase
song-artist-map
station-geo-map
student
subscribed-users
transactions
transactions_counts
12 row(s) in 4.0930 seconds
=> [SparkHBaseTable, "SparkHBaseTable_new", "bulktable", "clicks", "employee", "hbase", "song-artist-map", "station-geo-map", "stud
ent", "subscribed-users", "transactions", "transactions_counts"]
hbase(main):004:> 

```

The populate-lookup.sh also creates a lookup table “users_artists” in the HIVE, loading the data from the user-artist.txt, the below screen shot shows that the table has been created in the HIVE.

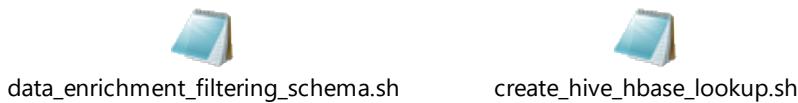
```

hive> show databases;
OK
custom
default
project
simplidb
test
Time taken: 2.146 seconds, Fetched: 5 row(s)
hive> use project;
OK
Time taken: 0.371 seconds
hive> show tables;
OK
users_artists
Time taken: 0.513 seconds, Fetched: 1 row(s)
hive> 

```

Now we need to link these lookup tables in hive using the Hbase Storage Handler.

4. With the help of “data_enrichment_filtering_schema.sh” file we will create hive tables on the top of Hbase tables using “create_hive_hbase_lookup.hql”.



Data Formatting

In this stage we are merging the data coming from both web applications and mobile applications and create a common table for analyzing purpose and create partitioned data based on batchid, since we are running this scripts for every 3 hours.

Run the script: ./dataformatting.sh



```

you have new mail in /var/spool/mail/acadgild
[acadgild@localhost MusicDataAnalysis]$ sbt -v package
[process_args] java_version = '1.8'
# Executing command line:
java
-Xms1024m
-Xmx1024m
-XX:ReservedCodeCacheSize=128m
-XX:MaxMetaspaceSize=256m
-jar
/usr/share/sbt/bin/sbt-launch.jar
package

[info] Loading project definition from /home/acadgild/examples/music/MusicDataAnalysis/project
[info] Loading settings from build.sbt ...
[info] Set current project to MusicDataAnalysis (in build file:/home/acadgild/examples/music/MusicDataAnalysis/)
[info] Compiling 3 Scala sources to /home/acadgild/examples/music/MusicDataAnalysis/target/scala-2.11/classes ...
[warn] there were three deprecation warnings; re-run with -deprecation for details
[warn] one warning found
[info] Done compiling.
[warn] Multiple main classes detected. Run 'show discoveredMainClasses' to see the list
[info] Packaging /home/acadgild/examples/music/MusicDataAnalysis/target/scala-2.11/musicdataanalysis_2.11-1.0.jar ...
[info] Done packaging.
[success] Total time: 40 s, completed Jul 13, 2018 4:29:56 AM
you have new mail in /var/spool/mail/acadgild
[acadgild@localhost MusicDataAnalysis]$ cd ..
[acadgild@localhost music]$ ./music_project_master.sh
Preparing to execute python scripts to generate data...
Data Generated Successfully !
Starting the daemons....
7763 Jps
5510 RunJar
4647 HMaster
3783 ResourceManager
6729 RunJar
3338 NameNode
3595 SecondaryNameNode
4748 HRegionServer

```

```

2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)
All hadoop daemons started !
upload the look up tables now in Hbase...
done with data population in look up tables !
lets do some data formatting now...
live table with Hbase Mapping Complete !
lets us do data enrichment as per the requirement...
ivy Default Cache set to: /home/acadgild/.ivy2/cache
the jars for the packages stored in: /home/acadgild/.ivy2/jars
:: loading settings :: url = jar:file:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/ivy-2.4.0.jar!/org/apache/ivy/core/settings
ivysettings.xml
com.databricks#spark-xml_2.10 added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent;1.0
  confs: [default]
    found com.databricks#spark-xml_2.10;0.4.1 in central
  resolution report :: resolve 71ms :: artifacts dl 18ms
  :: modules in use:
  com.databricks#spark-xml_2.10;0.4.1 from central in [default]
  -----
  |      modules      ||   artifacts   | | | | | |
  |      number| search|dwnlded|evicted|| number|dwnlded|
  | default | 1  | 0  | 0  | 0  || 1  | 0  |
  -----
  :: retrieving :: org.apache.spark#spark-submit-parent
  confs: [default]
  0 artifacts copied, 1 already retrieved (0kB/47ms)
18/07/13 04:30:23 INFO spark.SparkContext: Running Spark version 2.2.1
18/07/13 04:30:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/13 04:30:24 WARN util.Utils: Your hostname, localhost.localdomain resolves to a loopback address: 127.0.0.1; using 192.168.5.111 instead (on interface eth10)
18/07/13 04:30:25 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address
18/07/13 04:30:25 INFO spark.SparkContext: Submitted application: Data Formatting
18/07/13 04:30:25 INFO spark.SecurityManager: Changing view acls to: acadgild
18/07/13 04:30:25 INFO spark.SecurityManager: Changing modify acls to: acadgild
18/07/13 04:30:25 INFO spark.SecurityManager: Changing view acls groups to:
18/07/13 04:30:25 INFO spark.SecurityManager: Changing modify acls groups to:
18/07/13 04:30:25 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(a

```

```

2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)
18/07/13 04:30:25 INFO spark.SecurityManager: Changing modify acls groups to:
18/07/13 04:30:25 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(a
18/07/13 04:30:26 INFO util.Utils: Successfully started service 'sparkDriver' on port 46127.
18/07/13 04:30:26 INFO spark.SparkEnv: Registering MapOutputTracker
18/07/13 04:30:27 INFO spark.SparkEnv: Registering BlockManagerMaster
18/07/13 04:30:27 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
18/07/13 04:30:27 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
18/07/13 04:30:27 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-4133a8c4-9a3b-4f17-8c8f-3d3fa4c4bfae
18/07/13 04:30:27 INFO memory.MemoryStore: MemoryStore started with capacity 413.9 MB
18/07/13 04:30:27 INFO spark.SparkEnv: Registering OutputCommitCoordinator
18/07/13 04:30:28 INFO util.log: Logging initialized @9638ms
18/07/13 04:30:28 INFO server.Server: jetty-9.3.2-SNAPSHOT
18/07/13 04:30:28 INFO server.Server: Started @10175ms
18/07/13 04:30:28 INFO server.AbstractConnector: Started ServerConnector@30249771{HTTP/1.1,[http://1.1]}{0.0.0.0:4040}
18/07/13 04:30:28 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3887cf88{/jobs,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@241a0c3a{/jobs/json,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@30e92cb9{/jobs/job,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3f4f9acd{/jobs/job/json,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4bf324f9{/stages,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@ca66933{/stages/json,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4fa06f52{/stages/stage,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3c443976{/stages/stage/json,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3bd55d8{/stages/pool,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@7e0b9178{/stages/pool/json,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6f63c44f{/storage,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@62a8fd44{/storage/json,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@f5b5ca4f{/storage/rdd,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6b04acb2{/storage/rdd/json,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1d035be3{/environment,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4cfbaf4f{/environment/json,null,AVAILABLE,@spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@f212d84{/executors,null,AVAILABLE,@spark}
18/07/13 04:30:29 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6b9c69a9{/executors/json,null,AVAILABLE,@spark}
18/07/13 04:30:29 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@30b9ead3{/executors/threaddump,null,AVAILABLE,@spark}
18/07/13 04:30:29 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@412c995df{/executors/threaddump/json,null,AVAILABLE,@spark}
}

```

```
2. 127.0.0.1(acadgild) 3. 127.0.0.1(acadgild) 4. 127.0.0.1(acadgild)
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3c443976{/stages/stage/json,null,AVAILABLE,@Spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3bds5d8{/stages/pool,null,AVAILABLE,@Spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@7e0b9178{/stages/pool/json,null,AVAILABLE,@Spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6f63c44f{/storage,null,AVAILABLE,@Spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@62a0fd44{/storage/json,null,AVAILABLE,@Spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@f5b5ca4f{/storage/rdd,null,AVAILABLE,@Spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6b04acb2{/storage/rdd/json,null,AVAILABLE,@Spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@ld035be3{/environment,null,AVAILABLE,@Spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4cfbaef4{/environment/json,null,AVAILABLE,@Spark}
18/07/13 04:30:28 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@2424686b{/api,null,AVAILABLE,@Spark}
18/07/13 04:30:29 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@56febdcf{/jobs/job/kill,null,AVAILABLE,@Spark}
18/07/13 04:30:29 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@7d151af{/stages/stage/kill,null,AVAILABLE,@Spark}
18/07/13 04:30:29 INFO ui.SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.5.111:4040
18/07/13 04:30:29 INFO spark.SparkContext: Added JAR file:/home/acadgild/.ivy2/jars/com.databricks_spark-xm_2.10-0.4.1.jar at spark://192.168.5.111:46127/jars/com.databricks.spark.xml_2.10-0.4.1.jar with timestamp 1531436429285
18/07/13 04:30:29 INFO spark.SparkContext: Added file:/home/acadgild/examples/music/MusicDataAnalysis/target/scala-2.11/musicdataanalysis_2.11-1.0.jar at spark://192.168.5.111:46127/jars/musicdataanalysis_2.11-1.0.jar with timestamp 1531436429301
18/07/13 04:30:29 INFO executor.Executor: Starting executor ID driver on host localhost
18/07/13 04:30:29 INFO util.Utils: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port 46817.
18/07/13 04:30:29 INFO netty.NettyBlockTransferService: Server created on 192.168.5.111:46817
18/07/13 04:30:29 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
18/07/13 04:30:29 INFO storage.BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.5.111, 46817, None)
18/07/13 04:30:30 INFO storage.BlockManagerMasterEndpoint: Registering block manager 192.168.5.111:46817 with 413.9 MB RAM, BlockManagerId(driver, 192.168.5.111, 46817, None)
18/07/13 04:30:30 INFO storage.BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 192.168.5.111, 46817, None)
18/07/13 04:30:30 INFO storage.BlockManager: Initialized BlockManager: BlockManagerId(driver, 192.168.5.111, 46817, None)
18/07/13 04:30:30 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4d4fbabef{/metrics/json,null,AVAILABLE,@Spark}
18/07/13 04:30:31 INFO internal.SharedState: loading hive config file: file:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/conf/hive-site.xml
18/07/13 04:30:31 INFO internal.SharedState: Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('file:/home/acadgild/examples/music/spark-warehouse').
```

```
2. 127.0.0.1(acadgild) 3. 127.0.0.1(acadgild) 4. 127.0.0.1(acadgild)
site.xml
18/07/13 04:30:31 INFO internal.SharedState: Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('file:/home/acadgild/examples/music/spark-warehouse').
18/07/13 04:30:31 INFO internal.SharedState: Warehouse path is 'file:/home/acadgild/examples/music/spark-warehouse'.
18/07/13 04:30:31 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@31133b6e{/SQL,null,AVAILABLE,@Spark}
18/07/13 04:30:31 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@7ea08277{/SQL/json,null,AVAILABLE,@Spark}
18/07/13 04:30:31 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@61f80d55{/SQL/execution,null,AVAILABLE,@Spark}
18/07/13 04:30:31 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@280e81af{/SQL/execution/json,null,AVAILABLE,@Spark}
18/07/13 04:30:31 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4cb0a000{/static/sql,null,AVAILABLE,@Spark}
18/07/13 04:30:35 INFO hive.HiveUtils: Initializing HiveMetastoreConnection version 1.2.1 using Spark classes.
18/07/13 04:30:37 INFO hive.metastore: Trying to connect to metastore with URI thrift://localhost:9083
18/07/13 04:30:38 INFO hive.metastore: Connected to metastore.
18/07/13 04:30:42 INFO session.SessionState: Created local directory: /tmp/7a3b6f4e-f240-478f-a11f-1267a4cb402c_resources
18/07/13 04:30:42 INFO session.SessionState: Created HDFS directory: /tmp/hive/acadgild/7a3b6f4e-f240-478f-a11f-1267a4cb402c
18/07/13 04:30:42 INFO session.SessionState: Created local directory: /tmp/acadgild/7a3b6f4e-f240-478f-a11f-1267a4cb402c
18/07/13 04:30:42 INFO session.SessionState: Created HDFS directory: /tmp/hive/acadgild/7a3b6f4e-f240-478f-a11f-1267a4cb402c/_tmp_space.db
18/07/13 04:30:42 INFO client.HiveClientImpl: Warehouse location for Hive client (version 1.2.1) is file:/home/acadgild/examples/music/spark-warehouse
18/07/13 04:30:43 INFO session.SessionState: Created local directory: /tmp/llc37b94-e7d0-471e-a002-d0cefbc51efd_resources
18/07/13 04:30:43 INFO session.SessionState: Created HDFS directory: /tmp/hive/acadgild/llc37b94-e7d0-471e-a002-d0cefbc51efd
18/07/13 04:30:43 INFO session.SessionState: Created local directory: /tmp/acadgild/llc37b94-e7d0-471e-a002-d0cefbc51efd
18/07/13 04:30:43 INFO session.SessionState: Created HDFS directory: /tmp/hive/acadgild/llc37b94-e7d0-471e-a002-d0cefbc51efd/_tmp_space.db
18/07/13 04:30:43 INFO client.HiveClientImpl: Warehouse location for Hive client (version 1.2.1) is file:/home/acadgild/examples/music/spark-warehouse
18/07/13 04:30:43 INFO state.StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint
18/07/13 04:30:45 INFO memory.MemoryStore: Block broadcast_0 stored as values in memory (estimated size 281.1 KB, free 413.7 MB)
18/07/13 04:30:45 INFO memory.MemoryStore: Block broadcast_0_piece0 stored as bytes in memory (estimated size 24.1 KB, free 413.6 MB)
18/07/13 04:30:45 INFO storage.BlockManagerInfo: Added broadcast_0_piece0 in memory on 192.168.5.111:46817 (size: 24.1 KB, free: 413.9 MB)
18/07/13 04:30:45 INFO spark.SparkContext: Created broadcast_0 from newAPIHadoopFile at XmlFile.scala:46
18/07/13 04:30:45 INFO input.FileInputFormat: Total input paths to process : 1
18/07/13 04:30:46 INFO spark.SparkContext: Starting job: treeAggregate at InferSchema.scala:109
18/07/13 04:30:46 INFO scheduler.DAGScheduler: Got job 0 (treeAggregate at InferSchema.scala:109) with 1 output partitions
18/07/13 04:30:46 INFO scheduler.DAGScheduler: Final stage: ResultStage 0 (treeAggregate at InferSchema.scala:109)
18/07/13 04:30:46 INFO scheduler.DAGScheduler: Parents of final stage: List()
18/07/13 04:30:46 INFO scheduler.DAGScheduler: Missing parents: List()
18/07/13 04:30:46 INFO scheduler.DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[3] at treeAggregate at InferSchema.scala:109), which has no missing parents
18/07/13 04:30:47 INFO memory.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 4.6 KB, free 413.6 MB)
```

```
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)
18/07/13 04:30:46 INFO scheduler.DAGScheduler: Submitting ResultStage 0 (MapPartitionsRDD[3] at treeAggregate at InferSchema.scala:109), which has no missing parents
18/07/13 04:30:47 INFO memory.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 4.6 KB, free 413.6 MB)
18/07/13 04:30:47 INFO memory.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 2.7 KB, free 413.6 MB)
18/07/13 04:30:47 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on 192.168.5.111:46817 (size: 2.7 KB, free: 413.9 MB)
18/07/13 04:30:47 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1006
18/07/13 04:30:47 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 0 (MapPartitionsRDD[3] at treeAggregate at InferSchema.scala:109) (first 15 tasks are for partitions Vector(0))
18/07/13 04:30:47 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks
18/07/13 04:30:47 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition 0, PROCESS_LOCAL, 4913 bytes)
18/07/13 04:30:48 INFO executor.Executor: Running task 0.0 in stage 0.0 (TID 0)
18/07/13 04:30:48 INFO executor.Executor: Fetching spark://192.168.5.111:46127/jars/com.databricks_spark_xml_2.10-0.4.1.jar with timestamp 15314364285
18/07/13 04:30:48 INFO client.TransportClientFactory: Successfully created connection to /192.168.5.111:46127 after 164 ms (0 ms spent in bootstrapstraps)
18/07/13 04:30:48 INFO util.Utils: Fetching spark://192.168.5.111:46127/jars/com.databricks_spark_xml_2.10-0.4.1.jar to /tmp/spark-4ad63551-8628-4a58-849c-2e17ff5e86c5/userFiles-9b4925d8-2beb-4740-a4ba-e46b090d5192/fetchFileTemp2657189890019272814.tmp
18/07/13 04:30:48 INFO executor.Executor: Adding file:/tmp/spark-4ad63551-8628-4a58-849c-2e17ff5e86c5/userFiles-9b4925d8-2beb-4740-a4ba-e46b090d5192/com.databricks_spark_xml_2.10-0.4.1.jar to class loader
18/07/13 04:30:48 INFO executor.Executor: Fetching spark://192.168.5.111:46127/jars/musicdataanalysis_2.11-1.0.jar with timestamp 15314364293
01
18/07/13 04:30:48 INFO util.Utils: Fetching spark://192.168.5.111:46127/jars/musicdataanalysis_2.11-1.0.jar to /tmp/spark-4ad63551-8628-4a58-849c-2e17ff5e86c5/userFiles-9b4925d8-2beb-4740-a4ba-e46b090d5192/fetchFileTemp1858263179928555292.tmp
18/07/13 04:30:48 INFO executor.Executor: Adding file:/tmp/spark-4ad63551-8628-4a58-849c-2e17ff5e86c5/userFiles-9b4925d8-2beb-4740-a4ba-e46b090d5192/musicdataanalysis_2.11-1.0.jar to class loader
18/07/13 04:30:49 INFO rdd.NewHadoopRDD: Input split: file:/home/acadgild/examples/music/data/web/file.xml:0+6717
18/07/13 04:30:50 INFO executor.Executor: Finished task 0.0 in stage 0.0 (TID 0). 2033 bytes result sent to driver
18/07/13 04:30:50 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 2644 ms on localhost (executor driver) (1/1)
18/07/13 04:30:50 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
18/07/13 04:30:50 INFO scheduler.DAGScheduler: ResultStage 0 (treeAggregate at InferSchema.scala:109) finished in 2.765 s
18/07/13 04:30:50 INFO storage.BlockManagerInfo: Job 0 finished: treeAggregate at InferSchema.scala:109, took 4.141249 s
18/07/13 04:30:57 INFO storage.BlockManagerInfo: Removed broadcast_1_piece0 on 192.168.5.111:46817 in memory (size: 2.7 KB, free: 413.9 MB)
18/07/13 04:30:57 INFO storage.BlockManagerInfo: Removed broadcast_0_piece0 on 192.168.5.111:46817 in memory (size: 24.1 KB, free: 413.9 MB)
18/07/13 04:30:58 INFO execution.SparkSqlParser: Parsing command: web_data
18/07/13 04:31:00 INFO execution.SparkSqlParser: Parsing command: CREATE TABLE IF NOT EXISTS project.formatted_input
(
    User_id STRING,
```

```
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild)
18/07/13 04:30:58 INFO execution.SparkSqlParser: Parsing command: web_data
18/07/13 04:31:00 INFO execution.SparkSqlParser: Parsing command: CREATE TABLE IF NOT EXISTS project.formatted_input
(
    User_id STRING,
    Song_id STRING,
    Artist_id STRING,
    Timestamp STRING,
    Start_ts STRING,
    End_ts STRING,
    Geo_cd STRING,
    Station_id STRING,
    Song_end_type INT,
    Like INT,
    Dislike INT
)
PARTITIONED BY
    (batchid INT)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','

18/07/13 04:31:01 INFO parser.CatalystSqlParser: Parsing command: array<string>
18/07/13 04:31:08 INFO execution.SparkSqlParser: Parsing command: LOAD DATA LOCAL INPATH 'file:///home/acadgild/examples/music/data/mob/file.txt'
INTO TABLE project.formatted_input PARTITION (batchid='1')
18/07/13 04:31:09 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:09 INFO parser.CatalystSqlParser: Parsing command: string
18/07/13 04:31:09 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:10 INFO parser.CatalystSqlParser: Parsing command: string
18/07/13 04:31:10 INFO parser.CatalystSqlParser: Parsing command: string
```

```
18/07/13 04:31:09 INFO parser.CatalystSqlParser: Parsing command: string
18/07/13 04:31:09 INFO parser.CatalystSqlParser: Parsing command: string
18/07/13 04:31:09 INFO parser.CatalystSqlParser: Parsing command: string
18/07/13 04:31:09 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:10 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:10 INFO parser.CatalystSqlParser: Parsing command: string
18/07/13 04:31:10 INFO common.FileUtils: Creating directory if it doesn't exist: hdfs://localhost:8020/user/hive/warehouse/project.db/formatted_input/batchid=1
18/07/13 04:31:11 ERROR hdfs.KeyProviderCache: Could not find uri with key [dfs.encryption.key.provider.uri] to create a keyProvider !!
18/07/13 04:31:13 INFO metadata.Hive: Renaming src: file:/home/acadgild/examples/music/data/mob/file.txt, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/formatted_input/batchid=1/file.txt, Status:true
18/07/13 04:31:14 INFO execution.SparkSqlParser: Parsing command: INSERT INTO project.formatted_input
    PARTITION(batchid='1')
    SELECT user_id,
    song_id,
    artist_id,
    unix_timestamp(timestamp,'yyyy-MM-dd HH:mm:ss') AS timestamp,
    unix_timestamp(start_ts,'yyyy-MM-dd HH:mm:ss') AS start_ts,
    unix_timestamp(end_ts,'yyyy-MM-dd HH:mm:ss') AS end_ts,
    geo_cd,
    station_id,
    song_end_type,
    like,
    dislike
    FROM web_data
```

```
ed_input/batchid=1
18/07/13 04:31:11 ERROR hdfs.KeyProviderCache: Could not find uri with key [dfs.encryption.key.provider.url] to create a keyProvider !!
18/07/13 04:31:13 INFO metadata.Hive: Renaming src: file:/home/acadgild/examples/music/data/mob/file.txt, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/formatted_input/batchid=1/file.txt, Status:true
18/07/13 04:31:14 INFO execution.SparkSqlParser: Parsing command: INSERT INTO project.formatted_input
    PARTITION(batchid='1')
        SELECT user_id,
        song_id,
        artist_id,
        unix_timestamp(timestamp, 'yyyy-MM-dd HH:mm:ss') AS timestamp,
        unix_timestamp(start_ts, 'yyyy-MM-dd HH:mm:ss') AS start_ts,
        unix_timestamp(end_ts, 'yyyy-MM-dd HH:mm:ss') AS end_ts,
        geo_cd,
        station_id,
        song_end_type,
        like,
        dislike
    FROM web_data

18/07/13 04:31:15 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:15 INFO parser.CatalystSqlParser: Parsing command: string
18/07/13 04:31:15 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:15 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:15 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:15 INFO common.FileUtils: Creating directory if it doesn't exist: hdfs://localhost:8020/user/hive/warehouse/project.db/formatted_input/.hive-staging_hive_2018-07-13_04-31-15_855_2099640197049581634-1
18/07/13 04:31:16 INFO memory.MemoryStore: Block broadcast_2 stored as values in memory (estimated size 281.1 KB, free 413.7 MB)
18/07/13 04:31:17 INFO memory.MemoryStore: Block broadcast_2_piece0 stored as bytes in memory (estimated size 24.1 KB, free 413.6 MB)
18/07/13 04:31:17 INFO storage.BlockManagerInfo: Added broadcast_2_piece0 in memory on 192.168.5.111:46817 (size: 24.1 KB, free: 413.9 MB)
18/07/13 04:31:17 INFO spark.SparkContext: Created broadcast 2 from newAPIHadoopFile at XmlInputFormat.scala:46
18/07/13 04:31:17 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
```

```
18/07/13 04:31:17 INFO spark.SparkContext: Created broadcast 2 from newAPIHadoopFile at XmlFile.scala:46
18/07/13 04:31:17 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/07/13 04:31:17 INFO datasources.SQLHadoopMapReduceCommitProtocol: Using output committer class org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
18/07/13 04:31:19 INFO codegen.CodeGenerator: Code generated in 1038.065135 ms
18/07/13 04:31:19 INFO input.FileInputFormat: Total input paths to process : 1
18/07/13 04:31:20 INFO spark.SparkContext: Starting job: sql at DataFormatting.scala:56
18/07/13 04:31:20 INFO scheduler.DAGScheduler: Got job 1 (sql at DataFormatting.scala:56) with 1 output partitions
18/07/13 04:31:20 INFO scheduler.DAGScheduler: Final stage: ResultStage 1 (sql at DataFormatting.scala:56)
18/07/13 04:31:20 INFO scheduler.DAGScheduler: Parents of final stage: List()
18/07/13 04:31:20 INFO scheduler.DAGScheduler: Missing parents: List()
18/07/13 04:31:20 INFO scheduler.DAGScheduler: Submitting ResultStage 1 (MapPartitionsRDD[8] at sql at DataFormatting.scala:56), which has no missing parents
18/07/13 04:31:20 INFO memory.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 160.7 KB, free 413.5 MB)
18/07/13 04:31:20 INFO memory.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 59.1 KB, free 413.4 MB)
18/07/13 04:31:20 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on 192.168.5.11:46817 (size: 59.1 KB, free: 413.8 MB)
18/07/13 04:31:20 INFO spark.SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1096
18/07/13 04:31:20 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ResultStage 1 (MapPartitionsRDD[8] at sql at DataFormatting.scala:56) (first 15 tasks are for partitions Vector(0))
18/07/13 04:31:20 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 1 tasks
18/07/13 04:31:20 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 1, localhost, executor driver, partition 0, PROCESS_LOCAL, 4931 bytes)
18/07/13 04:31:20 INFO executor.Executor: Running task 0.0 in stage 1.0 (TID 1)
18/07/13 04:31:20 INFO rdd.NewHadoopRDD: Input split: file:/home/acgadgil/examples/music/data/web/file.xml+0+6717
18/07/13 04:31:20 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/07/13 04:31:20 INFO datasources.SQLHadoopMapReduceCommitProtocol: Using output committer class org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
18/07/13 04:31:21 INFO output.FileOutputCommitter: Saved output of task 'attempt_20180713043120_0001_m_000000_0' to hdfs://localhost:8020/use/r/hive/warehouse/project.db/formatted_input/.hive-staging_hive_2018-07-13_04-31-15_855_2099640197049581634-1-ext-10000/_temporary/0/task_20180713043120_0001_m_000000
18/07/13 04:31:21 INFO mapred.SparkHadoopMapReduceUtil: attempt_20180713043120_0001_m_000000_0: Committed
18/07/13 04:31:21 INFO executor.Executor: Finished task 0.0 in stage 1.0 (TID 1). 1651 bytes sent to driver
18/07/13 04:31:21 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 1.0 (TID 1) in 1466 ms on localhost (executor driver) (1/1)
18/07/13 04:31:21 INFO scheduler.DAGScheduler: ResultStage 1 (sql at DataFormatting.scala:56) finished in 1.467 s
18/07/13 04:31:21 INFO scheduler.DAGScheduler: Job 1 finished: sql at DataFormatting.scala:56, took 1.919243 s
18/07/13 04:31:21 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 1.0, whose tasks have all completed, from pool
18/07/13 04:31:22 INFO datasources.FileFormatWriter: Job null committed.
18/07/13 04:31:22 INFO parser.CatalystSqlParser: Parsing command: int
```

```

18/07/13 04:31:23 INFO parser.CatalystSqlParser: Parsing command: string
18/07/13 04:31:23 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:23 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:23 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:23 INFO metadata.Hive: Renaming src: hdfs://localhost:8020/user/hive/warehouse/project.db/formatted_input/.hive-staging_hive_2018-07-13-04-31-15_855_2099640197049581634-1-ext-10000/part-00000-377da395-9118-45f9-854e-592d01a038f1-c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/formatted_input/batchid=1/part-00000-377da395-9118-45f9-854e-592d01a038f1-c000, Status:true
18/07/13 04:31:24 INFO execution.SparkSqlParser: Parsing command: 'project'.'formatted_input'
18/07/13 04:31:24 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:24 INFO parser.CatalystSqlParser: Parsing command: string
18/07/13 04:31:24 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:24 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:24 INFO parser.CatalystSqlParser: Parsing command: int
18/07/13 04:31:24 INFO spark.SparkContext: Invoking stop() from shutdown hook
18/07/13 04:31:25 INFO server.AbstractConnector: Stopped Spark@30249771{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
18/07/13 04:31:25 INFO ui.SparkUI: Stopped Spark web UI at http://192.168.5.111:4040
18/07/13 04:31:25 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/07/13 04:31:25 INFO memory.MemoryStore: MemoryStore cleared
18/07/13 04:31:25 INFO storage.BlockManager: BlockManager stopped
18/07/13 04:31:25 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
18/07/13 04:31:25 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/07/13 04:31:25 INFO spark.SparkContext: Successfully stopped SparkContext
18/07/13 04:31:25 INFO util.ShutdownHookManager: Shutdown hook called
18/07/13 04:31:25 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-4ad63551-8628-4a58-849c-2e17ff5e86c5
data formatting complete !
Creating hive tables on top of hbase tables for data enrichment and filtering...
Data Enrichment Complete
Lets run some use cases now...
USE CASES COMPLETE !!
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost music]$
```

Data Enrichment & Filtering

In this stage, we will enrich the data coming from web and mobile applications using the lookup table stored in Hbase and divide the records based on the enrichment rules into ‘pass’ and ‘fail’ records.

Rules for data enrichment:

1. If any of like or dislike is NULL or absent, consider it as 0.
2. If fields like Geo_cd and Artist_id are NULL or absent, consult the lookup tables for fields Station_id and Song_id respectively to get the values of Geo_cd and Artist_id.
3. If corresponding lookup entry is not found, consider that record to be invalid

So, based on the enrichment rules, we will fill the null geo_cd and artist_id values with the help of corresponding lookup values in song-artist-map and station-geo-map tables in Hive-Hbase tables.



data_enrichment.sh

```
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild) 5.127.0.0.1(acadgild)
acadgild@localhost music]$ ./music_project_master.sh
preparing to execute python scripts to generate data...
ata Generated Successfully !
starting the daemons...
227 HMaster
284 SecondaryNameNode
494 ResourceManager
503 Jps
856 RunJar
329 HRegionServer
049 NameNode
146 DataNode
579 RunJar
596 NodeManager
140 HQuorumPeer
11 hadoop daemons started !
load the look up tables now in Hbase...
one with data population in look up tables !
lets do some data formatting now...
ive table with Hbase Mapping Complete !
et us do data enrichment as per the requirement...
ata formatting complete !
reating hive tables on top of hbase tables for data enrichment and filtering...
8/07/19 04:29:33 INFO spark.SparkContext: Running Spark version 2.2.1
8/07/19 04:29:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
8/07/19 04:29:35 WARN util.Utils: Your hostname, localhost.localdomain resolves to a loopback address: 127.0.0.1; using 192.168.5.119 instead
(on interface eth16)
8/07/19 04:29:35 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address
8/07/19 04:29:35 INFO spark.SparkContext: Submitted application: Data Enrichment
8/07/19 04:29:35 INFO spark.SecurityManager: Changing view acls to: acadgild
8/07/19 04:29:35 INFO spark.SecurityManager: Changing modify acls to: acadgild
8/07/19 04:29:35 INFO spark.SecurityManager: Changing view acls.groups to:
8/07/19 04:29:35 INFO spark.SecurityManager: Changing modify acls.groups to:
8/07/19 04:29:35 INFO spark.SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(acadgild); groups with view permissions: Set(); users with modify permissions: Set(acadgild); groups with modify permissions: Set()
8/07/19 04:29:37 INFO util.Utils: Successfully started service 'sparkDriver' on port 37739.
8/07/19 04:29:37 INFO spark.SparkEnv: Registering MapOutputTracker
```

```
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild) 5.127.0.0.1(acadgild)
18/07/19 04:29:35 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address
18/07/19 04:29:35 INFO spark.SparkContext: Submitted application: Data Enrichment
18/07/19 04:29:35 INFO spark.SecurityManager: Changing view acls to: acadgild
18/07/19 04:29:35 INFO spark.SecurityManager: Changing modify acls to: acadgild
18/07/19 04:29:35 INFO spark.SecurityManager: Changing view acls.groups to:
18/07/19 04:29:35 INFO spark.SecurityManager: Changing modify acls.groups to:
18/07/19 04:29:35 INFO spark.SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(acadgild); groups with view permissions: Set(); users with modify permissions: Set(acadgild); groups with modify permissions: Set()
18/07/19 04:29:37 INFO util.Utils: Successfully started service 'sparkDriver' on port 37739.
18/07/19 04:29:37 INFO spark.SparkEnv: Registering MapOutputTracker
18/07/19 04:29:37 INFO spark.SparkEnv: Registering BlockManagerMaster
18/07/19 04:29:37 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
18/07/19 04:29:37 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
18/07/19 04:29:37 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgr-95fc18e0-b505-4c31-b011-b027d6c59441
18/07/19 04:29:37 INFO memory.MemoryStore: Memorystore started with capacity 413.9 MB
18/07/19 04:29:38 INFO spark.SparkEnv: Registering OutputCommitCoordinator
18/07/19 04:29:38 INFO util.log: Logging initialized @9054ms
18/07/19 04:29:39 INFO server.Server: jetty-9.3.z-SNAPSHOT
18/07/19 04:29:39 INFO server.Server: Started @9712ms
18/07/19 04:29:39 INFO server.AbstractConnector: Started ServerConnector@35365c6f{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
18/07/19 04:29:39 INFO util.Utils: Successfully started service 'SparkUT' on port 4040.
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@42c67247{/jobs,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1827a871{/jobs/json,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@7249dadf{/jobs/job,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1c25b8a7{/jobs/job/json,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@750fe12e{/stages,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3e587920{/stages/stage,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@24f43aa3{/stages/stage,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@70ee0acd{/stages/stage/json,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6ab72419{/stages/pool,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4fdfa676{/storage/pool/json,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5be82d43{/storage,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@345e5a17{/storage/json,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@443dbe42{/storage/rd, null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1734f68{/storage/rdd/json,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5ed190be{/environment,null,AVAILABLE,@Spark}
18/07/19 04:29:39 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5bbc9f97{/environment/json,null,AVAILABLE,@Spark}
```

```
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild) 5.127.0.0.1(acadgild)
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hive-hbase-handler-2.3.2.jar at spark://192.168.5.119:37739/jars/hive-hbase-handler-2.3.2.jar with timestamp 1531954780035
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hbase-client-1.1.1.jar at spark://192.168.5.119:37739/jars/hbase-client-1.1.1.jar with timestamp 1531954780040
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hbase-common-1.1.1.jar at spark://192.168.5.119:37739/jars/hbase-common-1.1.1.jar with timestamp 1531954780045
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hbase-hadoop-compat-1.1.1.jar at spark://192.168.5.119:37739/jars/hbase-hadoop-compat-1.1.1.jar with timestamp 1531954780048
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hbase-server-1.1.1.jar at spark://192.168.5.119:37739/jars/hbase-server-1.1.1.jar with timestamp 1531954780048
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/hbase-protocol-1.1.1.jar at spark://192.168.5.119:37739/jars/hbase-protocol-1.1.1.jar with timestamp 1531954780051
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/zookeeper-3.4.6.jar at spark://192.168.5.119:37739/jars/zookeeper-3.4.6.jar with timestamp 1531954780052
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/guava-14.0.1.jar at spark://192.168.5.119:37739/jars/guava-14.0.1.jar with timestamp 1531954780052
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/htrace-core-3.1.0-incubating.jar at spark://192.168.5.119:37739/jars/htrace-core-3.1.0-incubating.jar with timestamp 1531954780058
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/examples/music/MusicDataAnalysis/target-scala-2.11/musicdataanalysis_2.11-1.0.jar at spark://192.168.5.119:37739/jars/musicdataanalysis_2.11-1.0.jar with timestamp 1531954780061
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/zookeeper-3.4.6.jar at spark://192.168.5.119:37739/jars/zookeeper-3.4.6.jar with timestamp 1531954780052
18/07/19 04:29:40 INFO spark.SparkContext: Added JAR file:/home/acadgild/install/hive/apache-hive-2.3.2-bin/lib/guava-14.0.1.jar at spark://192.168.5.119:37739/jars/guava-14.0.1.jar with timestamp 1531954780052
18/07/19 04:29:40 INFO util.Utils: Successfully started service 'org.apache.spark.netty.NettyBlockTransferService' on port 45174.
18/07/19 04:29:40 INFO netty.NettyBlockTransferService: Server created on 192.168.5.119:45174
18/07/19 04:29:40 INFO storage.BlockManager: Using org.apache.spark.storage.RandomBlockReplicationPolicy for block replication policy
18/07/19 04:29:40 INFO storage.BlockManagerMaster: Registering BlockManager BlockManagerId(driver, 192.168.5.119, 45174, None)
18/07/19 04:29:40 INFO storage.BlockManagerMasterEndpoint: Registering block manager 192.168.5.119:45174 with 413.9 MB RAM, BlockManagerId(driver, 192.168.5.119, 45174, None)
18/07/19 04:29:40 INFO storage.BlockManagerMaster: Registered BlockManager BlockManagerId(driver, 192.168.5.119, 45174, None)
18/07/19 04:29:40 INFO storage.BlockManager: Initialized BlockManager: BlockManagerId(driver, 192.168.5.119, 45174, None)
18/07/19 04:29:41 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3d7fa3ae{/metrics/json,null,AVAILABLE,@Spark}
18/07/19 04:29:42 INFO internal.SharedState: loading hive config file: /file:/home/acadgild/install/spark/spark-2.1.1-bin-hadoop2.7/conf/hive-site.xml
18/07/19 04:29:42 INFO internal.SharedState: Setting hive.metastore.warehouse.dir ('null') to the value of spark.sql.warehouse.dir ('file:/home/acadgild/examples/music/spark-warehouse').
18/07/19 04:29:42 INFO internal.SharedState: Warehouse path is 'file:/home/acadgild/examples/music/spark-warehouse'.
18/07/19 04:29:43 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3eed0f5{/SQL,null,AVAILABLE,@Spark}
18/07/19 04:29:43 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@2032e725{/SQL/json,null,AVAILABLE,@Spark}
18/07/19 04:29:43 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@26dcdb8c0{/SQL/execution,null,AVAILABLE,@Spark}
```

```
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild) 5.127.0.0.1(acadgild)
18/07/19 04:29:55 INFO session.SessionState: Created HDFS directory: /tmp/hive/acadgild/97e3daaf-b5fa-48e3-86f6-15b691d97567/_tmp_space.db
18/07/19 04:29:55 INFO client.HiveClientImpl: Warehouse location for Hive client (version 1.2.1) is file:/home/acadgild/examples/music/spark-warehouse
18/07/19 04:29:55 INFO state.StateStoreCoordinatorRef: Registered StateStoreCoordinator endpoint
18/07/19 04:29:55 INFO execution.SparkSqlParser: Parsing command: SET hive.auto.convert.join=false
18/07/19 04:30:01 INFO execution.SparkSqlParser: Parsing command: SET hive.exec.dynamic.partition.mode=nonstrict
18/07/19 04:30:01 INFO execution.SparkSqlParser: Parsing command: USE project
18/07/19 04:30:01 INFO execution.SparkSqlParser: Parsing command: CREATE TABLE IF NOT EXISTS enriched_data
(
    user_id STRING,
    Song_id STRING,
    Artist_id STRING,
    Timestamp STRING,
    Start_ts STRING,
    End_ts STRING,
    End_cd STRING,
    Station_id STRING,
    Song_end_type INT,
    Like INT,
    Dislike INT
)
PARTITIONED BY
(
batchid INT,
status STRING)
STORED AS ORC
18/07/19 04:30:03 INFO parser.CatalystSqlParser: Parsing command: array<string>
18/07/19 04:30:03 INFO execution.SparkSqlParser: Parsing command: INSERT OVERWRITE TABLE enriched_data
    PARTITION (batchid, status)
    SELECT
        i.user_id,
        i.song_id,
        sa.artist_id,
        i.timestamp,
        i.start_ts,
        i.end_ts,
        sg.geo_cd,
        i.station_id,
```

```
18/07/19 04:30:03 INFO parser.CatalystsSqlParser: Parsing command: array<string>
18/07/19 04:30:03 INFO execution.SparkSqlParser: Parsing command: INSERT OVERWRITE TABLE enrich_
PARTITION (batchid, status)
SELECT
  i.user_id,
  i.song_id,
  sa.artist_id,
  i.timestamp,
  i.start_ts,
  i.end_ts,
  sg.geo_cd,
  i.station_id,
  IF(i.song_end_type IS NULL, 3, i.song_end_type) AS song_end_type,
  IF(i.like IS NULL, 0, i.like) AS like,
  IF(i.dislike IS NULL, 0, i.dislike) AS dislike,
  i.batchid,
  IF((i.like=1 AND i.dislike=1)
  OR i.user_id IS NULL
  OR i.song_id IS NULL
  OR i.timestamp IS NULL
  OR i.start_ts IS NULL
  OR i.end_ts IS NULL
  OR i.geo_cd IS NULL
  OR i.user_id=''
  OR i.song_id=''
  OR i.timestamp=''
  OR i.start_ts=''
  OR i.end_ts=''
  OR i.geo_cd=''
  OR sg.geo_cd IS NULL
  OR sg.geo_cd=''
  OR sa.artist_id IS NULL
  OR sa.artist_id='', 'fail', 'pass') AS status
FROM formatted_input i LEFT OUTER JOIN station_geo_map sg ON i.station_
LEFT OUTER JOIN song_artist_map sa ON i.song_id = sa.song_id
WHERE i.batchid=1
```

```
2.127.0.0.1 (acadgild) 3.127.0.0.1 (acadgild) 4.127.0.0.1 (acadgild) 5.127.0.0.1 (acadgild)
WHERE i.batchid=1

18/07/19 04:30:04 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:30:04 INFO parser.CatalystSqlParser: Parsing command: string
18/07/19 04:30:04 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:30:04 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:30:04 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:30:05 INFO parser.CatalystSqlParser: Parsing command: string
18/07/19 04:30:05 INFO parser.CatalystSqlParser: Parsing command: string
18/07/19 04:30:06 INFO parser.CatalystSqlParser: Parsing command: string
18/07/19 04:30:06 INFO parser.CatalystSqlParser: Parsing command: string
18/07/19 04:30:07 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:30:07 INFO parser.CatalystSqlParser: Parsing command: string
18/07/19 04:30:08 INFO common.FileUtils: Creating directory if it doesn't exist: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data.hive-staging_hive_2018-07-19-04-30-08_317_6744492802765411400-1
18/07/19 04:30:11 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/07/19 04:30:11 INFO datastores.SQLHadoopMapReduceCommitProtocol: Using output committer class org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
18/07/19 04:30:11 INFO spark.ContextCleaner: Cleaned accumulator 2
18/07/19 04:30:15 INFO codegen.CodeGenerator: Code generated in 1643.219041 ms
```

```
18/07/19 04:30:24 INFO memory.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 24.9 KB, free 413.3 MB)
18/07/19 04:30:24 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on 192.168.5.119:45174 (size: 24.9 KB, free: 413.9 MB)
18/07/19 04:30:24 INFO spark.SparkContext: Created broadcast 1 from 
18/07/19 04:30:24 INFO hbbase.HBaseStorageHandler: Configuring input job properties
18/07/19 04:30:25 INFO zookeeper.RecoverableZooKeeper: Process identifier=hconnection-0x6da8e98 connecting to ZooKeeper ensemble=localhost:2181
18/07/19 04:30:26 INFO zookeeper.ZooKeeper: Client environment:zookeeper.version=3.4.6-1569965, built on 02/20/2014 09:09 GMT
18/07/19 04:30:26 INFO zookeeper.ZooKeeper: Client environment:host.name=localhost
18/07/19 04:30:26 INFO zookeeper.ZooKeeper: Client environment:java.version=1.8.0_151
18/07/19 04:30:26 INFO zookeeper.ZooKeeper: Client environment:java.vendor=Oracle Corporation
18/07/19 04:30:26 INFO zookeeper.ZooKeeper: Client environment:java.home=/usr/java/jdk1.8.0_151/jre
18/07/19 04:30:26 INFO zookeeper.ZooKeeper: Client environment:java.class.path=/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/conf/:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/JavaEWAH-0.3.2.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/hadoop-mapreduce-client-app-2.7.3.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/spark-mesos-2.11-2.2.1.jar:/home/acadgild/install/all/spark/spark-2.2.1-bin-hadoop2.7/jars/datanucleus-rdbms-3.2.9.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/hbase64-2.3.8.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jarhill-java-0.8.0.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/oro-2.0.8.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jdktrace-core-3.1.0-incubating.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/kryo-shaded-3.0.3.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jdk/api-3.0.1.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jackson-core-2.6.5.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jackson-databind-2.9.1.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/xercesImpl-2.9.1.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jsons-jackson-2.11-3.2.11.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/commons-lang-2.6.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/libthrift-0.9.3.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/commons-logging-1.3.0.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jetty-6.1.26.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/minlog-1.3.0.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/janino-3.0.0.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jersey-client-2.22.2.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/httpclient-4.5.2.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jsp-api-2.1.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/scalap-2.11.8.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/hadoop-yarn-common-2.7.3.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/protobuf-java-2.5.0.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jersey-container-servlet-core-2.22.2.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jcl-over-slf4j-1.7.16.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/py4j-0.10.4.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/jcl-over-slf4j-1.7.16.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/validation-api-1.0.1.Final.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/avro-1.7.7.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/core-1.1.2.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/metrics-jvm-3.1.2.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/spark-graphx-2.11-2.2.1.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/macromedia-comps-2.11-1.1.1.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/hadoop-kfs-2.4.0-b34.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/calcite-core-1.2.0-incubating.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/apolliance-1.0.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/commons-compress-1.4.1.jar:/home/acadgild/install/spark/spark-2.2.1-bin-hadoop2.7/jars/xbe
```

```
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild) 5.127.0.0.1(acadgild)
18/07/19 04:30:32 INFO zookeeper.ClientCnxn: Socket connection established to localhost/0:0:0:0:0:0:1:2181, initiating session
18/07/19 04:30:32 INFO zookeeper.ClientCnxn: Session establishment complete on server localhost/0:0:0:0:0:0:1:2181, sessionid = 0x164af47df
e50007, negotiated timeout = 90000
18/07/19 04:30:32 INFO util.RegionSizeCalculator: Calculating region sizes for table "song-artist-map".
18/07/19 04:30:33 INFO client.ConnectionManager$ConnectionImplementation: Closing master protocol: MasterService
18/07/19 04:30:33 INFO client.ConnectionManager$ConnectionImplementation: Closing zookeeper sessionid=0x164af47dfe50007
18/07/19 04:30:33 INFO zookeeper.ClientCnxn: EventThread shut down
18/07/19 04:30:33 INFO zookeeper.ZooKeeper: Session: 0x164af47dfe50007 closed
18/07/19 04:30:34 INFO spark.SparkContext: Starting job: sql at DataEnrichment.scala:73
18/07/19 04:30:34 INFO scheduler.DAGScheduler: Registering RDD 12 (sql at DataEnrichment.scala:73)
18/07/19 04:30:34 INFO scheduler.DAGScheduler: Registering RDD 5 (sql at DataEnrichment.scala:73)
18/07/19 04:30:34 INFO scheduler.DAGScheduler: Registering RDD 17 (sql at DataEnrichment.scala:73)
18/07/19 04:30:34 INFO scheduler.DAGScheduler: Registering RDD 24 (sql at DataEnrichment.scala:73)
18/07/19 04:30:34 INFO scheduler.DAGScheduler: Got job 0 (sql at DataEnrichment.scala:73) with 200 output partitions
18/07/19 04:30:34 INFO scheduler.DAGScheduler: Final stage: ResultStage 4 (sql at DataEnrichment.scala:73)
18/07/19 04:30:34 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 2, ShuffleMapStage 3)
18/07/19 04:30:34 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 0 (MapPartitionsRDD[12] at sql at DataEnrichment.scala:73), which h
as no missing parents
18/07/19 04:30:34 INFO memory.MemoryStore: Block broadcast_3 stored as values in memory (estimated size 16.6 KB, free 413.0 MB)
18/07/19 04:30:34 INFO memory.MemoryStore: Block broadcast_3_piece0 stored as bytes in memory (estimated size 9.1 KB, free 413.0 MB)
18/07/19 04:30:34 INFO storage.BlockManagerInfo: Added broadcast_3_piece0 in memory on 192.168.5.119:45174 (size: 9.1 KB, free: 413.8 MB)
18/07/19 04:30:35 INFO spark.SparkContext: Created broadcast 3 from broadcast at DAGScheduler.scala:1006
18/07/19 04:30:35 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 0 (MapPartitionsRDD[12] at sql at DataEnrichme
nt.scala:73) (first 15 tasks are for partitions Vector(0, 1))
18/07/19 04:30:35 INFO scheduler.TaskSchedulerImpl: Adding task set 0.0 with 1 tasks
18/07/19 04:30:35 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 1 (MapPartitionsRDD[5] at sql at DataEnrichment.scala:73), which ha
s no missing parents
18/07/19 04:30:35 INFO memory.MemoryStore: Block broadcast_4 stored as values in memory (estimated size 11.9 KB, free 413.0 MB)
18/07/19 04:30:35 INFO memory.MemoryStore: Block broadcast_4_piece0 stored as bytes in memory (estimated size 6.0 KB, free 412.9 MB)
18/07/19 04:30:35 INFO storage.BlockManagerInfo: Added broadcast_4_piece0 in memory on 192.168.5.119:45174 (size: 6.0 KB, free: 413.8 MB)
18/07/19 04:30:35 INFO scheduler.DAGScheduler: Submitting 2 missing tasks from ShuffleMapStage 1 (MapPartitionsRDD[5] at sql at DataEnrichmen
t.scala:73) (first 15 tasks are for partitions Vector(0, 1))
18/07/19 04:30:35 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 2 tasks
18/07/19 04:30:35 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 0.0 (TID 0, localhost, executor driver, partition 0, PROCESS_L
OCAL, 4918 bytes)
18/07/19 04:30:35 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 3 (MapPartitionsRDD[24] at sql at DataEnrichment.scala:73), which h
```

```
2.127.0.0.1(acadgild) 3.127.0.0.1(acadgild) 4.127.0.0.1(acadgild) 5.127.0.0.1(acadgild)
18/07/19 04:30:45 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 2 blocks
18/07/19 04:30:45 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 2 blocks
18/07/19 04:30:45 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/07/19 04:30:45 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 12 ms
18/07/19 04:30:45 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/07/19 04:30:45 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/07/19 04:30:45 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/07/19 04:30:45 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 2 ms
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 2 ms
18/07/19 04:30:46 INFO executor.Executor: Finished task 5.0 in stage 2.0 (TID 7), 2800 bytes result sent to driver
18/07/19 04:30:46 INFO scheduler.TaskSetManager: Starting task 6.0 in stage 2.0 (TID 8, localhost, executor driver, partition 6, PROCESS_L
OCAL, 5043 bytes)
18/07/19 04:30:46 INFO executor.Executor: Running task 6.0 in stage 2.0 (TID 8)
18/07/19 04:30:46 INFO scheduler.TaskSetManager: Finished task 5.0 in stage 2.0 (TID 7) in 268 ms on localhost (executor driver) (3/200)
18/07/19 04:30:46 INFO executor.Executor: Finished task 3.0 in stage 2.0 (TID 6), 2800 bytes result sent to driver
18/07/19 04:30:46 INFO scheduler.TaskSetManager: Starting task 7.0 in stage 2.0 (TID 9, localhost, executor driver, partition 7, PROCESS_L
OCAL, 5043 bytes)
18/07/19 04:30:46 INFO scheduler.TaskSetManager: Finished task 3.0 in stage 2.0 (TID 6) in 316 ms on localhost (executor driver) (4/200)
18/07/19 04:30:46 INFO executor.Executor: Running task 7.0 in stage 2.0 (TID 9)
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 2 blocks
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 10 ms
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 2 blocks
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 14 ms
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 7 ms
18/07/19 04:30:46 INFO executor.Executor: Finished task 6.0 in stage 2.0 (TID 8), 2757 bytes result sent to driver
18/07/19 04:30:46 INFO scheduler.TaskSetManager: Starting task 8.0 in stage 2.0 (TID 10, localhost, executor driver, partition 8, PROCESS_L
OCAL, 5043 bytes)
18/07/19 04:30:46 INFO executor.Executor: Running task 8.0 in stage 2.0 (TID 10)
18/07/19 04:30:46 INFO scheduler.TaskSetManager: Finished task 6.0 in stage 2.0 (TID 8) in 233 ms on localhost (executor driver) (5/200)
18/07/19 04:30:46 INFO executor.Executor: Finished task 7.0 in stage 2.0 (TID 9), 2757 bytes result sent to driver
18/07/19 04:30:46 INFO scheduler.TaskSetManager: Starting task 9.0 in stage 2.0 (TID 11, localhost, executor driver, partition 9, PROCESS_L
OCAL, 5043 bytes)
18/07/19 04:30:46 INFO executor.Executor: Running task 9.0 in stage 2.0 (TID 11)
18/07/19 04:30:46 INFO scheduler.TaskSetManager: Finished task 7.0 in stage 2.0 (TID 9) in 265 ms on localhost (executor driver) (6/200)
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 2 blocks
18/07/19 04:30:46 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 5 ms
```

```
2: 127.0.0.1(acadgild) 3: 127.0.0.1(acadgild) 4: 127.0.0.1(acadgild) 5: 127.0.0.1(acadgild)
18/07/19 04:31:02 INFO scheduler.TaskSetManager: Starting task 192.0 in stage 2.0 (TID 202, localhost, executor driver, partition 192, ANY, 5 043 bytes)
18/07/19 04:31:02 INFO executor.Executor: Running task 192.0 in stage 2.0 (TID 202)
18/07/19 04:31:02 INFO scheduler.TaskSetManager: Finished task 157.0 in stage 2.0 (TID 201) in 141 ms on localhost (executor driver) (197/200)
)
18/07/19 04:31:02 INFO executor.Executor: Finished task 118.0 in stage 2.0 (TID 200). 2886 bytes result sent to driver
18/07/19 04:31:02 INFO scheduler.TaskSetManager: Starting task 199.0 in stage 2.0 (TID 203, localhost, executor driver, partition 199, ANY, 5 043 bytes)
18/07/19 04:31:02 INFO scheduler.TaskSetManager: Finished task 118.0 in stage 2.0 (TID 200) in 175 ms on localhost (executor driver) (198/200)
)
18/07/19 04:31:02 INFO executor.Executor: Running task 199.0 in stage 2.0 (TID 203)
18/07/19 04:31:02 INFO storage.ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 2 blocks
18/07/19 04:31:02 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/07/19 04:31:02 INFO storage.ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
18/07/19 04:31:02 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/07/19 04:31:02 INFO storage.ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 2 blocks
18/07/19 04:31:02 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 2 ms
18/07/19 04:31:02 INFO storage.ShuffleBlockFetcherIterator: Getting 1 non-empty blocks out of 1 blocks
18/07/19 04:31:02 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/07/19 04:31:02 INFO executor.Executor: Finished task 192.0 in stage 2.0 (TID 202). 2886 bytes result sent to driver
18/07/19 04:31:02 INFO scheduler.TaskSetManager: Finished task 192.0 in stage 2.0 (TID 202) in 146 ms on localhost (executor driver) (199/200)
)
18/07/19 04:31:02 INFO executor.Executor: Finished task 199.0 in stage 2.0 (TID 203). 2886 bytes result sent to driver
18/07/19 04:31:02 INFO scheduler.TaskSetManager: Finished task 199.0 in stage 2.0 (TID 203) in 167 ms on localhost (executor driver) (200/200)
)
18/07/19 04:31:02 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 2.0, whose tasks have all completed, from pool
18/07/19 04:31:02 INFO scheduler.DAGScheduler: ShuffleMapStage 2 (sql at DataEnrichment.scala:73) finished in 19.161 s
18/07/19 04:31:02 INFO scheduler.DAGScheduler: looking for newly runnable stages
18/07/19 04:31:02 INFO scheduler.DAGScheduler: running: Set()
18/07/19 04:31:02 INFO scheduler.DAGScheduler: waiting: Set(ResultStage 4)
18/07/19 04:31:02 INFO scheduler.DAGScheduler: failed: Set()
18/07/19 04:31:02 INFO scheduler.DAGScheduler: Submitting ResultStage 4 (MapPartitionsRDD[29] at sql at DataEnrichment.scala:73), which has n 0 missing parents
18/07/19 04:31:03 INFO memory.MemoryStore: Block broadcast_7 stored as values in memory (estimated size 194.5 KB, free 412.7 MB)
18/07/19 04:31:03 INFO memory.MemoryStore: Block broadcast_7_piece0 stored as bytes in memory (estimated size 70.1 KB, free 412.6 MB)
18/07/19 04:31:03 INFO storage.BlockManagerInfo: Added broadcast_7_piece0 in memory on 192.168.5.119:45174 (size: 70.1 KB, free: 413.7 MB)
18/07/19 04:31:03 INFO spark.SparkContext: Created broadcast 7 from broadcast at DAGScheduler.scala:1006
18/07/19 04:31:03 INFO scheduler.DAGScheduler: Submitting 200 missing tasks from ResultStage 4 (MapPartitionsRDD[29] at sql at DataEnrichment
```

```
2: 127.0.0.1(acadgild) 3: 127.0.0.1(acadgild) 4: 127.0.0.1(acadgild) 5: 127.0.0.1(acadgild)
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/07/19 04:31:18 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/07/19 04:31:18 INFO datasources.SQLHadoopMapReduceCommitProtocol: Using output committer class org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
18/07/19 04:31:18 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/07/19 04:31:18 INFO datasources.SQLHadoopMapReduceCommitProtocol: Using output committer class org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
18/07/19 04:31:18 INFO mapred.SparkHadoopMapRedUtil: No need to commit output of task because needsTaskCommit=false: attempt_20180719043118_0 004_m_000094_0
18/07/19 04:31:18 INFO executor.Executor: Finished task 94.0 in stage 4.0 (TID 294). 4224 bytes result sent to driver
18/07/19 04:31:18 INFO mapred.SparkHadoopMapRedUtil: No need to commit output of task because needsTaskCommit=false: attempt_20180719043118_0 004_m_000096_0
18/07/19 04:31:18 INFO executor.Executor: Finished task 96.0 in stage 4.0 (TID 295). 4267 bytes result sent to driver
18/07/19 04:31:18 INFO scheduler.TaskSetManager: Starting task 97.0 in stage 4.0 (TID 296, localhost, executor driver, partition 97, PROCESS_LOCAL, 5054 bytes)
18/07/19 04:31:18 INFO scheduler.TaskSetManager: Finished task 94.0 in stage 4.0 (TID 294) in 248 ms on localhost (executor driver) (91/200)
18/07/19 04:31:18 INFO executor.Executor: Running task 97.0 in stage 4.0 (TID 296)
18/07/19 04:31:18 INFO scheduler.TaskSetManager: Starting task 98.0 in stage 4.0 (TID 297, localhost, executor driver, partition 98, PROCESS_LOCAL, 5054 bytes)
18/07/19 04:31:18 INFO scheduler.TaskSetManager: Finished task 96.0 in stage 4.0 (TID 295) in 222 ms on localhost (executor driver) (92/200)
18/07/19 04:31:18 INFO executor.Executor: Running task 98.0 in stage 4.0 (TID 297)
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 200 blocks
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 200 blocks
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/07/19 04:31:18 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/07/19 04:31:18 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/07/19 04:31:18 INFO datasources.SQLHadoopMapReduceCommitProtocol: Using output committer class org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
18/07/19 04:31:18 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/07/19 04:31:18 INFO datasources.SQLHadoopMapReduceCommitProtocol: Using output committer class org.apache.hadoop.mapreduce.lib.output.File
```

```
18/07/19 04:31:37 INFO scheduler.TaskSetManager: Finished task 199.0 in stage 4.0 (TID 403) in 439 ms on localhost (executor driver) (200/200)
18/07/19 04:31:37 INFO scheduler.DAGScheduler: ResultStage 4 (sql at DataEnrichment.scala:73) finished in 33.977 s
18/07/19 04:31:37 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 4.0, whose tasks have all completed, from pool
18/07/19 04:31:37 INFO scheduler.DAGScheduler: Job 0 finished: sql at DataEnrichment.scala:73, took 63.286214 s
18/07/19 04:31:38 INFO datasources.FileFormatWriter: Job null committed.
18/07/19 04:31:38 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:38 INFO parser.CatalystSqlParser: Parsing command: string
18/07/19 04:31:38 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:38 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:38 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:39 ERROR hdfs.KeyProviderCache: Could not find uri with key [dfs.encryption.key.provider.uri] to create a keyProvider !!
18/07/19 04:31:39 INFO metadata.Hive: Replacing src:hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=pass/part-00057-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=pass/part-00057-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, Status:true
18/07/19 04:31:39 INFO metadata.Hive: Replacing src:hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=pass/part-00087-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=pass/part-00087-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, Status:true
18/07/19 04:31:39 INFO metadata.Hive: Replacing src:hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=pass/part-00095-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=pass/part-00095-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, Status:true
18/07/19 04:31:39 INFO metadata.Hive: Replacing src:hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=pass/part-00107-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=pass/part-00107-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, Status:true
18/07/19 04:31:39 INFO metadata.Hive: Replacing src:hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=pass/part-00160-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=pass/part-00160-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, Status:true
```

```
tus:true
18/07/19 04:31:45 INFO metadata.Hive: Replacing src:hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=fail/part-00160-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=fail/part-00160-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, Status:true
18/07/19 04:31:45 INFO metadata.Hive: Replacing src:hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=fail/part-00161-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=fail/part-00161-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, Status:true
18/07/19 04:31:45 INFO metadata.Hive: Replacing src:hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=fail/part-00165-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=fail/part-00165-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, Status:true
18/07/19 04:31:45 INFO metadata.Hive: Replacing src:hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=fail/part-00199-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, dest: hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=fail/part-00199-170ff912-ed6c-4a04-b154-b3b0bfdb70c4.c000, Status:true
18/07/19 04:31:46 INFO metadata.Hive: New loading path = hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/-ext-10000/batchid=1/status=fail with partSpec {batchid=1, status=fail}
18/07/19 04:31:46 INFO execution.SparkSqlParser: Parsing command: project` .enriched_data`
18/07/19 04:31:46 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:46 INFO parser.CatalystSqlParser: Parsing command: string
18/07/19 04:31:46 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:46 INFO spark.SparkContext: Invoking stop() from shutdown hook
18/07/19 04:31:46 INFO server.AbstractConnector: Stopped Spark@5365c6f{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
18/07/19 04:31:46 INFO ui.SparkUI: Stopped Spark web UI at http://192.168.5.19:4040
18/07/19 04:31:46 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/07/19 04:31:47 INFO memory.MemoryStore: MemoryStore cleared
```

```

2. 127.0.0.1(acadgild) 3. 127.0.0.1(acadgild) 4. 127.0.0.1(acadgild) 5. 127.0.0.1(acadgild)
//localhost:8020/user/hive/warehouse/project.db/enriched_data/batchid=1/status=fail/part-00199-170ff912-ed6c-4a04-b154-b3b0bfbdb70c4.c000, Status: true
18/07/19 04:31:46 INFO metadata.Hive: New loading path = hdfs://localhost:8020/user/hive/warehouse/project.db/enriched_data/.hive-staging_hive_2018-07-19_04-30-08_317_6744492802765411400-1/ext-10000/batchid=1/status=fail with partSpec {batchid=1, status=fail}
18/07/19 04:31:46 INFO execution.SparkSqlParser: Parsing command: 'project'.'enriched_data'
18/07/19 04:31:46 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:46 INFO parser.CatalystSqlParser: Parsing command: string
18/07/19 04:31:46 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:46 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:46 INFO parser.CatalystSqlParser: Parsing command: int
18/07/19 04:31:46 INFO spark.SparkContext: Invoking stop() from shutdown hook
18/07/19 04:31:46 INFO server.AbstractConnector: Stopped Spark web UI at http://192.168.5.119:4040
18/07/19 04:31:46 INFO ui.SparkUI: Stopped Spark web UI at http://192.168.5.119:4040
18/07/19 04:31:46 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/07/19 04:31:46 INFO memory.MemoryStore: MemoryStore cleared
18/07/19 04:31:47 INFO storage.BlockManager: BlockManager stopped
18/07/19 04:31:47 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
18/07/19 04:31:47 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/07/19 04:31:47 INFO spark.SparkContext: Successfully stopped SparkContext
18/07/19 04:31:47 INFO util.ShutdownHookManager: Shutdown hook called
18/07/19 04:31:47 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-e469bf0-6852-4ae4-9a5a-795ed6c042bf
18/07/19 04:31:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/07/19 04:32:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Data Enrichment Complete
Lets run some use cases now...
USE CASES COMPLETE !!
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost music]$ 

```

hive> select * from enriched_data;	OK													
U110	S207	A303	1465490556	1468094889	1462863262	AU	ST401	3	1	1	1	1	fail	
U107	S204	A304	1465130523	1475130523	1465130523	AP	ST402	1	1	1	1	1	fail	
S204	A304	1465130523	1465230523	1475130523	J	ST413	0	0	1	1	1	1	fail	
U111	S204	A304	1468094889	1494297562	1494297562	J	ST413	0	1	0	1	1	fail	
U118	S204	A304	1465130523	1465130523	1485130523	NULL	ST415	0	0	1	1	1	fail	
U112	S209	A305	1465130523	1475130523	1475130523	E	ST414	0	1	1	1	1	fail	
U119	S209	A305	1475130523	1465230523	1465230523	J	ST403	2	1	0	1	1	fail	
U106	S209	A305	1462863262	1494297562	1465490556	A	ST410	0	1	1	1	1	fail	
U110	S208	A304	1465130523	1485130523	1485130523	E	ST408	3	1	1	1	1	fail	
S210	A302	NULL	1465230523	1485130523	1465130523	E	ST404	1	0	0	1	1	fail	
U112	S210	NULL	1465490556	1462863262	1494297562	E	ST404	3	1	1	1	1	fail	
U114	S210	NULL	1462863262	1465490556	1494297562	A	ST405	3	1	1	1	1	fail	
U118	S210	NULL	1475130523	1465230523	1465130523	NULL	ST415	1	0	1	1	1	fail	
U111	S210	NULL	1475130523	1475130523	1465230523	AP	ST412	0	0	0	1	1	fail	
S205	A301	1462863262	1494297562	1462863262	J	ST413	3	0	0	0	1	1	fail	
U117	S203	A303	1475130523	1465230523	1475130523	AU	ST406	0	1	1	1	1	fail	
S202	A302	1462863262	1465490556	1462863262	A	ST400	0	0	0	0	1	1	pass	
U107	S202	A302	1465490556	1465490556	1465490556	E	ST414	1	0	0	1	1	pass	
U103	S202	A302	1465130523	1475130523	1475130523	J	ST413	2	0	0	1	1	pass	
S204	A304	1495130523	1465130523	1485130523	AU	ST401	2	1	0	1	1	pass		
U106	S204	A304	1468094889	1468094889	1462863262	E	ST408	3	0	0	1	1	pass	
U103	S204	A304	1475130523	1465230523	1465130523	AU	ST406	0	0	0	1	1	pass	
U107	S209	A305	1462863262	1494297562	1494297562	E	ST404	1	0	1	1	1	pass	
S209	A305	1495130523	1475130523	1465230523	J	ST403	3	0	0	1	1	pass		
U118	S209	A305	1465490556	1494297562	1462863262	A	ST411	2	1	0	1	1	pass	
U108	S209	A305	1465490556	1494297562	1465490556	J	ST413	2	0	1	1	1	pass	
U101	S206	A302	1494297562	1494297562	1494297562	E	ST404	1	0	0	1	1	pass	
U113	S206	A302	1468094889	1468094889	1462863262	A	ST411	2	0	1	1	1	pass	
S206	A302	1465230523	1475130523	1465130523	AU	ST401	2	0	1	1	1	pass		
U105	S206	A302	1465230523	1475130523	1485130523	AU	ST401	2	0	1	1	1	pass	
U105	S208	A304	1465130523	1485130523	1475130523	E	ST414	0	1	0	1	1	pass	
S208	A304	1465490556	1494297562	1468094889	A	ST411	0	0	0	1	1	pass		
U106	S208	A304	1495130523	1475130523	1475130523	AP	ST407	2	0	1	1	1	pass	
U108	S205	A301	1465490556	1462863262	1462863262	AP	ST402	0	0	0	1	1	pass	
U104	S205	A301	1475130523	1465130523	1465230523	A	ST400	3	0	0	1	1	pass	
U117	S205	A301	1465130523	1465230523	1485130523	E	ST414	2	0	0	1	1	pass	
S205	A301	1462863262	1494297562	1462863262	E	ST409	1	0	0	1	1	pass		

U118	S204	A304	1465130523	1465130523	1485130523	NULL	ST415	0	0	1	1	fail
U112	S209	A305	1465130523	1475130523	1465230523	E	ST414	0	1	1	1	fail
U119	S209	A305	1475130523	1465230523	1465230523	J	ST403	2	1	0	1	fail
U106	S209	A305	1462863262	1494297562	1465490556	A	ST410	0	1	1	1	fail
U110	S208	A304	1465130523	1465130523	1485130523	E	ST408	3	1	1	1	fail
U102	S210	NULL	1465230523	1485130523	1465130523	E	ST404	1	0	0	1	fail
U112	S210	NULL	1465490556	1462863262	1494297562	E	ST404	3	1	1	1	fail
U114	S210	NULL	1462863262	1465490556	1494297562	A	ST405	3	1	1	1	fail
U118	S210	NULL	1475130523	1465230523	1465130523	NULL	ST415	1	0	1	1	fail
U111	S210	NULL	1475130523	1475130523	1465230523	AP	ST412	0	0	0	1	fail
NULL	S205	A301	1462863262	1494297562	1462863262	J	ST413	3	0	0	1	fail
U117	S203	A303	1475130523	1465230523	1475130523	AU	ST406	0	1	1	1	fail
U111	S202	A302	1462863262	1465490556	1462863262	A	ST400	0	0	0	1	pass
U107	S202	A302	1465490556	1465490556	1465490556	E	ST414	1	0	0	1	pass
U103	S202	A302	1465130523	1475130523	1475130523	J	ST413	2	0	0	1	pass
U117	S204	A304	1495130523	1465130523	1465130523	AU	ST401	2	1	0	1	pass
U106	S204	A304	1468094889	1468094889	1468094889	E	ST408	3	0	0	1	pass
U103	S204	A304	1475130523	1465230523	1465130523	AU	ST406	0	0	0	1	pass
U107	S209	A305	1462863262	1494297562	1494297562	E	ST404	1	0	1	1	pass
U109	S209	A305	1495130523	1475130523	1465230523	J	ST403	3	0	0	1	pass
U118	S209	A305	1465490556	1494207562	1462863262	A	ST411	2	1	0	1	pass
U108	S209	A305	1465490556	1494297562	1465490556	J	ST413	2	0	1	1	pass
U101	S206	A302	1494297562	1494297562	1494297562	E	ST404	1	0	0	1	pass
U113	S206	A302	1468094889	1468094889	1468094889	E	ST402	2	0	1	1	pass
U105	S206	A302	1465230523	1475130523	1485130523	AU	ST401	2	0	1	1	pass
U105	S208	A304	1465130523	1485130523	1475130523	E	ST414	0	1	0	1	pass
U118	S208	A304	1465490556	1494297562	1468094889	A	ST411	0	0	0	1	pass
U106	S208	A304	1495130523	1475130523	1475130523	AP	ST407	2	0	1	1	pass
U108	S205	A301	1465490556	1462863262	1462863262	AP	ST402	0	0	0	1	pass
U104	S205	A301	1475130523	1465130523	1465130523	A	ST400	3	0	0	1	pass
U117	S205	A301	1465130523	1465230523	1485130523	E	ST414	2	0	0	1	pass
U119	S205	A301	1462863262	1494297562	1462863262	E	ST409	1	0	0	1	pass
U114	S205	A301	1494297562	1494297562	1494297562	A	ST410	3	0	1	1	pass
U109	S205	A301	1462863262	1468094889	1468094889	E	ST408	1	0	1	1	pass
U102	S205	A301	1468094889	1462863262	1462863262	AP	ST407	2	0	0	1	pass
U103	S206	A300	1495130523	1475130523	1485130523	E	ST409	1	0	1	1	pass

Time taken: 60.704 seconds, Fetched: 40 row(s)

hive> 

At the end script will automatically divide the records based on status pass & fail and dump the result into processed_dir folder with valid and invalid folders.

Data Analysis using Spark

In this stage, we will do analysis on enriched data using Spark SQL and run the program using Spark Submit command.

Before running the spark-submit command we have to zip -d command to remove the bad manifests in created spark project jar file to avoid the invalid Signature exception.



data_analysis.sh



DataAnalysis.scala

```
[acadgild@localhost MusicDataAnalysis]$ sbt -v package
[process_args] java.version = '1.8'
# Executing command line:
java
-Xms1024m
-Xmx1024m
-XX:ReservedCodeCacheSize=128m
-XX:MaxMetaspaceSize=256m
-jar
/usr/share/sbt/bin/sbt-launch.jar
package

[info] Loading project definition from /home/acadgild/examples/music/MusicDataAnalysis/project
[info] Loading settings from build.sbt ...
[info] Set current project to MusicDataAnalysis (in build file:/home/acadgild/examples/music/MusicDataAnalysis/)
[info] Compiling 1 Scala source to /home/acadgild/examples/music/MusicDataAnalysis/target/scala-2.11/classes ...
[warn] there was one deprecation warning; re-run with -deprecation for details
[warn] one warning found
[info] Done compiling.
[warn] Multiple main classes detected. Run 'show discoveredMainClasses' to see the list
[info] Packaging /home/acadgild/examples/music/MusicDataAnalysis/target/scala-2.11/musicdataanalysis_2.11-1.0.jar ...
[info] Done packaging.
[success] Total time: 62 s, completed Aug 7, 2018 11:06:15 AM
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost MusicDataAnalysis]$
[acadgild@localhost MusicDataAnalysis]$ cd
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost music]$ ./music_project_master.sh
Preparing to execute python scripts to generate data...
Data Generated Successfully !
Starting the daemons....
3456 NameNode
5585 RunJar
4390 ResourceManager
6343 Jps
5464 RunJar
3913 HRegionServer
4489 NodeManager
```

```
4127 SecondaryNameNode
3759 HQuorumPeer
All hadoop daemons started !
Upload the look up tables now in Hbase...
Done with data population in look up tables !
Lets do some data formatting now...
Hive table with Hbase Mapping Complete !
Let us do data enrichment as per the requirement...
data formatting complete !
Creating hive tables on top of hbase tables for data enrichment and filtering...
Data Enrichment Complete
Lets run some use cases now...
18/08/07 11:09:14 INFO spark.SparkContext: Running Spark version 2.2.1
18/08/07 11:09:22 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
18/08/07 11:09:24 WARN util.Utils: Your hostname, localhost.localdomain resolves to a loopback address: 127.0.0.1; using 192.168.169.128 instead (on interface eth1)
18/08/07 11:09:24 WARN util.Utils: Set SPARK_LOCAL_IP if you need to bind to another address
18/08/07 11:09:25 INFO spark.SparkContext: Submitted application: Data Analysis
18/08/07 11:09:25 INFO spark.SecurityManager: Changing acls to: acadgild
18/08/07 11:09:25 INFO spark.SecurityManager: Changing modify acls to: acadgild
18/08/07 11:09:25 INFO spark.SecurityManager: Changing view acls groups to:
18/08/07 11:09:25 INFO spark.SecurityManager: Changing modify acls groups to:
18/08/07 11:09:25 INFO spark.SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(acadgild); groups with view permissions: Set(); users with modify permissions: Set(acadgild); groups with modify permissions: Set()
18/08/07 11:09:30 INFO util.Utils: Successfully started service 'sparkDriver' on port 43791.
18/08/07 11:09:31 INFO spark.SparkEnv: Registering MapOutputTracker
18/08/07 11:09:31 INFO spark.SparkEnv: Registering BlockManagerMaster
18/08/07 11:09:31 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
18/08/07 11:09:31 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
18/08/07 11:09:31 INFO storage.DiskBlockManager: Created local directory at /tmp/blockmgrr-01db9d77-b10d-46a9-b042-803cef38c3b3
18/08/07 11:09:32 INFO memory.MemoryStore: MemoryStore started with capacity 413.9 MB
18/08/07 11:09:32 INFO spark.SparkEnv: Registering OutputCommitCoordinator
18/08/07 11:09:34 INFO util.log: Logging initialized @40449ms
18/08/07 11:09:36 INFO server.Server: jetty-9.3.2-SNAPSHOT
18/08/07 11:09:36 INFO server.Server: Started @42773ms
18/08/07 11:09:37 INFO server.AbstractConnector: Started ServerConnector@4274b5a4{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
```

```
18/08/07 11:09:36 INFO server.Server: Started @42773ms
18/08/07 11:09:37 INFO server.AbstractConnector: Started ServerConnector@4274b5a4{[HTTP/1.1, [http://1.1]}{0.0.0.0:4040}
18/08/07 11:09:37 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.
18/08/07 11:09:37 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@l5bcf458{/jobs,null,AVAILABLE,@Spark}
18/08/07 11:09:37 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@73393584{/jobs/json,null,AVAILABLE,@Spark}
18/08/07 11:09:37 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@l827a81/{jobs/job,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4362d7df/{jobs/job/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@l2c5b8a7/{stages,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@705fe12e{/stages/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3e587920{/stages/stage,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1e11b55{/stages/stage/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@70e0accd{/stages/pool,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6ab72419{/stages/pool/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4df4fa676{/storage,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5be82d43{/storage/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@345e51a17{/storage/rdd,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@443d2e42{/storage/rdd/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@l734f689{/environment,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5ed190be{/environment/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5bbcc9f72{/executors,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@41382722{/executors/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@425357dd{/executors/threadDump,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@210386e0{/executors/threadDump/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6f5f87a2c{/static,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@652eac4b{/api,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@2a5f51a63{/api/json,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@29539e36{/jobs/job/kill,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@f5c79a6f{/stages/stage/kill,null,AVAILABLE,@Spark}
18/08/07 11:09:38 INFO ui.SparkUI: Bound SparkUI to 0.0.0.0, and started at http://192.168.169.128:4040
18/08/07 11:09:38 INFO spark.SparkContext: Added JAR file:/home/acgadild/install/hive/apache-hive-2.3.2-bin/lib/hive-hbase-handler-2.3.2.jar at spark://192.168.169.128:43791/jars/hive-hbase-handler-2.3.2.jar with timestamp 1533620378813
18/08/07 11:09:38 INFO spark.SparkContext: Added JAR file:/home/acgadild/install/hive/apache-hive-2.3.2-bin/lib/hbase-client-1.1.1.jar at spark://192.168.169.128:43791/jars/hbase-client-1.1.1.jar with timestamp 1533620378829
18/08/07 11:09:38 INFO spark.SparkContext: Added JAR file:/home/acgadild/install/hive/apache-hive-2.3.2-bin/lib/hbase-common-1.1.1.jar at spark://192.168.169.128:43791/jars/hbase-common-1.1.1.jar with timestamp 1533620378829
18/08/07 11:09:38 INFO spark.SparkContext: Added JAR file:/home/acgadild/install/hive/apache-hive-2.3.2-bin/lib/hbase-hadoop-compat-1.1.1.jar at spark://192.168.169.128:43791/jars/hbase-hadoop-compat-1.1.1.jar with timestamp 1533620378833
```

```
2. 192.168.169.128 3. 192.168.169.128 4. 192.168.169.128 5. 192.168.169.128
18/08/07 11:11:30 INFO executor.Executor: Fetching spark://192.168.169.128:43791/jars/hbase-hadoop-compat-1.1.1.jar with timestamp 1533620378829
18/08/07 11:11:30 INFO util.Utils: Fetching spark://192.168.169.128:43791/jars/hbase-hadoop-compat-1.1.1.jar to /tmp/spark-ce4b0b1d-2c7a-40ab-808a-39dalac75d87/userFiles-a5bbbf88-8e67-42ae-8086-5d31d3652525/fetchFileTemp3694656950875252315.tmp
18/08/07 11:11:30 INFO executor.Executor: Adding file:/tmp/spark-ce4b0b1d-2c7a-40ab-808a-39dalac75d87/userFiles-a5bbbf88-8e67-42ae-8086-5d31d3652525/hbase-hadoop-compat-1.1.1.jar to class loader
18/08/07 11:11:30 INFO executor.Executor: Fetching spark://192.168.169.128:43791/jars/hbase-common-1.1.1.jar with timestamp 1533620378829
18/08/07 11:11:30 INFO util.Utils: Fetching spark://192.168.169.128:43791/jars/hbase-common-1.1.1.jar to /tmp/spark-ce4b0b1d-2c7a-40ab-808a-39dalac75d87/userFiles-a5bbbf88-8e67-42ae-8086-5d31d3652525/fetchFileTemp8693952579170910117.tmp
18/08/07 11:11:30 INFO executor.Executor: Adding file:/tmp/spark-ce4b0b1d-2c7a-40ab-808a-39dalac75d87/userFiles-a5bbbf88-8e67-42ae-8086-5d31d3652525/hbase-common-1.1.1.jar to class loader
18/08/07 11:11:30 INFO executor.Executor: Fetching spark://192.168.169.128:43791/jars/hbase-server-1.1.1.jar with timestamp 1533620378834
18/08/07 11:11:30 INFO util.Utils: Fetching spark://192.168.169.128:43791/jars/hbase-server-1.1.1.jar to /tmp/spark-ce4b0b1d-2c7a-40ab-808a-39dalac75d87/userFiles-a5bbbf88-8e67-42ae-8086-5d31d3652525/fetchFileTemp2982530517119663813.tmp
18/08/07 11:11:31 INFO executor.Executor: Adding file:/tmp/spark-ce4b0b1d-2c7a-40ab-808a-39dalac75d87/userFiles-a5bbbf88-8e67-42ae-8086-5d31d3652525/hbase-server-1.1.1.jar to class loader
18/08/07 11:11:32 INFO codegen.CodeGenerator: Code generated in 158.615412 ms
18/08/07 11:11:32 INFO codegen.CodeGenerator: Code generated in 73.143932 ms
18/08/07 11:11:32 INFO codegen.CodeGenerator: Code generated in 96.565068 ms
18/08/07 11:11:32 INFO executor.Executor: Finished task 0.0 in stage 0.0 (TID 0). 1954 bytes result sent to driver
18/08/07 11:11:33 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0) in 4906 ms on localhost (executor driver) (1/1)
18/08/07 11:11:33 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks have all completed, from pool
18/08/07 11:11:33 INFO scheduler.DAGScheduler: ShuffleMapStage 0 (sql at DataAnalysis.scala:152) finished in 5.195 s
18/08/07 11:11:33 INFO scheduler.DAGScheduler: looking for newly runnable stages
18/08/07 11:11:33 INFO scheduler.DAGScheduler: running: Set()
18/08/07 11:11:33 INFO scheduler.DAGScheduler: waiting: Set(ShuffleMapStage 1, ShuffleMapStage 2, ResultStage 3)
18/08/07 11:11:33 INFO scheduler.DAGScheduler: failed: Set()
18/08/07 11:11:33 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 1 (MapPartitionsRDD[6] at sql at DataAnalysis.scala:152), which
has no missing parents
18/08/07 11:11:33 INFO memory.MemoryStore: Block broadcast_1 stored as values in memory (estimated size 33.5 KB, free 413.9 MB)
18/08/07 11:11:33 INFO memory.MemoryStore: Block broadcast_1_piece0 stored as bytes in memory (estimated size 14.0 KB, free 413.9 MB)
18/08/07 11:11:33 INFO storage.BlockManagerInfo: Added broadcast_1_piece0 in memory on 192.168.169.128:40491 (size: 14.0 KB, free: 413.9 M
B)
18/08/07 11:11:33 INFO spark.SparkContext: Created broadcast 1 from broadcast at DAGScheduler.scala:1006
18/08/07 11:11:33 INFO scheduler.DAGScheduler: Submitting 200 missing tasks from ShuffleMapStage 1 (MapPartitionsRDD[6] at sql at DataAnal
ysis.scala:152) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14))
18/08/07 11:11:33 INFO scheduler.TaskSchedulerImpl: Adding task set 1.0 with 200 tasks
18/08/07 11:11:33 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 1.0 (TID 1, localhost, executor driver, partition 0, PROCESS_L
```

```
2. 192.168.169.128 3. 192.168.169.128 4. 192.168.169.128 5. 192.168.169.128
18/08/07 11:11:37 INFO executor.Executor: Running task 24.0 in stage 1.0 (TID 25)
18/08/07 11:11:37 INFO scheduler.TaskSetManager: Starting task 25.0 in stage 1.0 (TID 26, localhost, executor driver, partition 25, PROCES
S_LOCAL, 4715 bytes)
18/08/07 11:11:37 INFO scheduler.TaskSetManager: Finished task 22.0 in stage 1.0 (TID 23) in 164 ms on localhost (executor driver) (24/200
)
18/08/07 11:11:37 INFO executor.Executor: Running task 25.0 in stage 1.0 (TID 26)
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/08/07 11:11:37 INFO executor.Executor: Finished task 24.0 in stage 1.0 (TID 25). 2993 bytes result sent to driver
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/08/07 11:11:37 INFO executor.Executor: Finished task 25.0 in stage 1.0 (TID 26). 2993 bytes result sent to driver
18/08/07 11:11:37 INFO scheduler.TaskSetManager: Starting task 26.0 in stage 1.0 (TID 27, localhost, executor driver, partition 26, PROCES
S_LOCAL, 4715 bytes)
18/08/07 11:11:37 INFO executor.Executor: Running task 26.0 in stage 1.0 (TID 27)
18/08/07 11:11:37 INFO scheduler.TaskSetManager: Finished task 24.0 in stage 1.0 (TID 25) in 117 ms on localhost (executor driver) (25/200
)
18/08/07 11:11:37 INFO scheduler.TaskSetManager: Starting task 27.0 in stage 1.0 (TID 28, localhost, executor driver, partition 27, PROCES
S_LOCAL, 4715 bytes)
18/08/07 11:11:37 INFO scheduler.TaskSetManager: Finished task 25.0 in stage 1.0 (TID 26) in 133 ms on localhost (executor driver) (26/200
)
18/08/07 11:11:37 INFO executor.Executor: Running task 27.0 in stage 1.0 (TID 28)
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/08/07 11:11:37 INFO executor.Executor: Finished task 26.0 in stage 1.0 (TID 27). 2993 bytes result sent to driver
18/08/07 11:11:37 INFO scheduler.TaskSetManager: Starting task 28.0 in stage 1.0 (TID 29, localhost, executor driver, partition 28, PROCES
S_LOCAL, 4715 bytes)
18/08/07 11:11:37 INFO executor.Executor: Running task 28.0 in stage 1.0 (TID 29)
18/08/07 11:11:37 INFO scheduler.TaskSetManager: Finished task 26.0 in stage 1.0 (TID 27) in 136 ms on localhost (executor driver) (27/200
)
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 1 blocks
18/08/07 11:11:37 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/08/07 11:11:37 INFO executor.Executor: Finished task 27.0 in stage 1.0 (TID 28). 2993 bytes result sent to driver
18/08/07 11:11:37 INFO scheduler.TaskSetManager: Starting task 29.0 in stage 1.0 (TID 30, localhost, executor driver, partition 29, PROCES
S_LOCAL, 4715 bytes)
18/08/07 11:11:37 INFO executor.Executor: Running task 29.0 in stage 1.0 (TID 30)
```



```
2. 192.168.169.128 3. 192.168.169.128 4. 192.168.169.128 5. 192.168.169.128
18/08/07 11:12:31 INFO zookeeper.ClientCnxn: Socket connection established to localhost/127.0.0.1:2181, initiating session
18/08/07 11:12:31 INFO zookeeper.ClientCnxn: Session establishment complete on server localhost/127.0.0.1:2181, sessionid = 0x16512d2577900
006, negotiated timeout = 90000
18/08/07 11:12:33 INFO util.RegionSizeCalculator: Calculating region sizes for table "subscribed-users".
18/08/07 11:12:39 INFO client.ConnectionManager$HConnectionImplementation: Closing master protocol: MasterService
18/08/07 11:12:39 INFO client.ConnectionManager$HConnectionImplementation: Closing zookeeper sessionid=0x16512d257790006
18/08/07 11:12:39 INFO zookeeper.ClientCnxn: EventThread shut down
18/08/07 11:12:39 INFO zookeeper.ZooKeeper: Session: 0x16512d257790006 closed
18/08/07 11:12:41 INFO spark.SparkContext: Starting job: sql at DataAnalysis.scala:154
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Registering RDD 20 (sql at DataAnalysis.scala:154)
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Registering RDD 27 (sql at DataAnalysis.scala:154)
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Registering RDD 32 (sql at DataAnalysis.scala:154)
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Got job 1 (sql at DataAnalysis.scala:154) with 200 output partitions
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Final stage: ResultStage 7 (sql at DataAnalysis.scala:154)
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Parents of final stage: List(ShuffleMapStage 6)
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Missing parents: List(ShuffleMapStage 6)
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 4 (MapPartitionsRDD[20] at sql at DataAnalysis.scala:154), which
has no missing parents
18/08/07 11:12:41 INFO memory.MemoryStore: Block broadcast_6 stored as values in memory (estimated size 10.9 KB, free 413.2 MB)
18/08/07 11:12:41 INFO memory.MemoryStore: Block broadcast_6_piece0 stored as bytes in memory (estimated size 5.6 KB, free 413.2 MB)
18/08/07 11:12:41 INFO storage.BlockManagerInfo: Added broadcast_6_piece0 in memory on 192.168.169.128:40491 (size: 5.6 KB, free: 413.9 MB
)
18/08/07 11:12:41 INFO spark.SparkContext: Created broadcast 6 from broadcast at DAGScheduler.scala:1006
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Submitting 7 missing tasks from ShuffleMapStage 4 (MapPartitionsRDD[20] at sql at DataAnaly
sis.scala:154) (first 15 tasks are for partitions Vector(0, 1, 2, 3, 4, 5, 6))
18/08/07 11:12:41 INFO scheduler.TaskSchedulerImpl: Adding task set 4.0 with 7 tasks
18/08/07 11:12:41 INFO scheduler.DAGScheduler: Submitting ShuffleMapStage 5 (MapPartitionsRDD[27] at sql at DataAnalysis.scala:154), which
has no missing parents
18/08/07 11:12:42 INFO memory.MemoryStore: Block broadcast_7 stored as values in memory (estimated size 16.7 KB, free 413.2 MB)
18/08/07 11:12:42 INFO memory.MemoryStore: Block broadcast_7_piece0 stored as bytes in memory (estimated size 9.1 KB, free 413.2 MB)
18/08/07 11:12:42 INFO storage.BlockManagerInfo: Added broadcast_7_piece0 in memory on 192.168.169.128:40491 (size: 9.1 KB, free: 413.8 MB
)
18/08/07 11:12:42 INFO spark.SparkContext: Created broadcast 7 from broadcast at DAGScheduler.scala:1006
18/08/07 11:12:42 INFO scheduler.DAGScheduler: Submitting 1 missing tasks from ShuffleMapStage 5 (MapPartitionsRDD[27] at sql at DataAnaly
sis.scala:154) (first 15 tasks are for partitions Vector(0))
18/08/07 11:12:42 INFO scheduler.TaskSchedulerImpl: Adding task set 5.0 with 1 tasks
18/08/07 11:12:42 INFO scheduler.TaskSetManager: Starting task 0.0 in stage 4.0 (TID 402, localhost, executor driver, partition 0, PROCESS
_LOCAL, 5078 bytes)
```

```
2. 192.168.169.128 3. 192.168.169.128 4. 192.168.169.128 5. 192.168.169.128
18/08/07 11:13:35 INFO executor.Executor: Finished task 169.0 in stage 7.0 (TID 777). 3687 bytes result sent to driver
18/08/07 11:13:35 INFO scheduler.TaskSetManager: Starting task 172.0 in stage 7.0 (TID 780, localhost, executor driver, partition 172, PRO
CESS_LOCAL, 4726 bytes)
18/08/07 11:13:35 INFO scheduler.TaskSetManager: Finished task 169.0 in stage 7.0 (TID 777) in 530 ms on localhost (executor driver) (169/
200)
18/08/07 11:13:35 INFO executor.Executor: Running task 172.0 in stage 7.0 (TID 780)
18/08/07 11:13:35 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 200 blocks
18/08/07 11:13:35 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 0 ms
18/08/07 11:13:35 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/08/07 11:13:35 INFO datasources.SQLHadoopMapReduceCommitProtocol: Using output committer class org.apache.hadoop.mapreduce.lib.output.F
ileOutputCommitter
18/08/07 11:13:35 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 200 blocks
18/08/07 11:13:35 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
18/08/07 11:13:35 INFO output.FileOutputCommitter: File Output Committer Algorithm version is 1
18/08/07 11:13:35 INFO datasources.SQLHadoopMapReduceCommitProtocol: Using output committer class org.apache.hadoop.mapreduce.lib.output.F
ileOutputCommitter
18/08/07 11:13:35 INFO output.FileOutputCommitter: Saved output of task 'attempt_20180807111335_0007_m_000171_0' to hdfs://localhost:8020/
user/hive/warehouse/project.db/usersBehaviour/.hive-staging_hive_2018-08-07_11-12-20_775_8329065990233436026-1-ext-10000/_temporary/0/t
ask_20180807111335_0007_m_000171
18/08/07 11:13:35 INFO mapred.SparkHadoopMapRedUtil: attempt_20180807111335_0007_m_000171_0: Committed
18/08/07 11:13:35 INFO executor.Executor: Finished task 171.0 in stage 7.0 (TID 779). 3687 bytes result sent to driver
18/08/07 11:13:35 INFO scheduler.TaskSetManager: Starting task 173.0 in stage 7.0 (TID 781, localhost, executor driver, partition 173, PRO
CESS_LOCAL, 4726 bytes)
18/08/07 11:13:35 INFO executor.Executor: Running task 173.0 in stage 7.0 (TID 781)
18/08/07 11:13:35 INFO scheduler.TaskSetManager: Finished task 171.0 in stage 7.0 (TID 779) in 304 ms on localhost (executor driver) (170/
200)
18/08/07 11:13:35 INFO output.FileOutputCommitter: Saved output of task 'attempt_20180807111335_0007_m_000172_0' to hdfs://localhost:8020/
user/hive/warehouse/project.db/usersBehaviour/.hive-staging_hive_2018-08-07_11-12-20_775_8329065990233436026-1-ext-10000/_temporary/0/t
ask_20180807111335_0007_m_000172
18/08/07 11:13:35 INFO mapred.SparkHadoopMapRedUtil: attempt_20180807111335_0007_m_000172_0: Committed
18/08/07 11:13:35 INFO executor.Executor: Finished task 172.0 in stage 7.0 (TID 780). 3687 bytes result sent to driver
18/08/07 11:13:35 INFO scheduler.TaskSetManager: Starting task 174.0 in stage 7.0 (TID 782, localhost, executor driver, partition 174, PRO
CESS_LOCAL, 4726 bytes)
18/08/07 11:13:35 INFO executor.Executor: Running task 174.0 in stage 7.0 (TID 782)
18/08/07 11:13:35 INFO scheduler.TaskSetManager: Finished task 172.0 in stage 7.0 (TID 780) in 367 ms on localhost (executor driver) (171/
200)
18/08/07 11:13:36 INFO storage.ShuffleBlockFetcherIterator: Getting 0 non-empty blocks out of 200 blocks
18/08/07 11:13:36 INFO storage.ShuffleBlockFetcherIterator: Started 0 remote fetches in 1 ms
```

```
18/08/07 11:14:42 INFO execution.SparkSqlParser: Parsing command: CREATE TABLE IF NOT EXISTS top_10_royalty_songs
(
  song_id STRING,
  duration INT
)
PARTITIONED BY (batchid INT)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
18/08/07 11:14:42 INFO execution.SparkSqlParser: Parsing command: INSERT OVERWRITE TABLE top_10_royalty_songs
PARTITION(batchid=1)
SELECT song_id,
       SUM(ABS(CAST(end_ts AS DECIMAL(20,0))-CAST(start_ts AS DECIMAL(20,0)))) AS duration
  FROM enriched_data
 WHERE status='pass'
   AND batchid=1
   AND ('like'='1' OR song_end_type=0)
 GROUP BY song_id
 ORDER BY duration DESC
LIMIT 10
18/08/07 11:14:43 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:14:43 INFO parser.CatalystSqlParser: Parsing command: string
18/08/07 11:14:43 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:14:43 INFO parser.CatalystSqlParser: Parsing command: string
18/08/07 11:14:43 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:14:43 INFO common.FileUtils: Creating directory if it doesn't exist: hdfs://localhost:8020/user/hive/warehouse/project.db/top_
10_royalty_songs/hive-staging_hive_2018-08-07_11-14-43_781_1080811041543145558-1
```

```
2. 192.168.169.128 3. 192.168.169.128 4. 192.168.169.128 x 5. 192.168.169.128
18/08/07 11:14:55 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:14:55 INFO execution.SparkSqlParser: Parsing command: CREATE TABLE IF NOT EXISTS top_10_unsubscribed_users
(
  user_id STRING,
  duration INT
)
PARTITIONED BY (batchid INT)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
18/08/07 11:14:55 INFO execution.SparkSqlParser: Parsing command: INSERT OVERWRITE TABLE top_10_unsubscribed_users
PARTITION(batchid=1)
SELECT
  ed.user_id,
  SUM(ABS(CAST(ed.end_ts AS DECIMAL(20,0))-CAST(ed.start_ts AS DECIMAL(20,0)))) AS duration
FROM enriched_data ed
LEFT OUTER JOIN subscribed_users su
ON ed.user_id=su.user_id
WHERE ed.status='pass'
AND ed.batchid=1
AND (su.user_id IS NULL OR (CAST(ed.timestamp AS DECIMAL(20,0)) > CAST(su.subscn_end_dt AS DECIMAL(20,0))))
GROUP BY ed.user_id
ORDER BY duration DESC
LIMIT 10
18/08/07 11:14:56 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:14:56 INFO parser.CatalystSqlParser: Parsing command: string
18/08/07 11:14:56 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:14:56 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:14:56 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:14:56 INFO parser.CatalystSqlParser: Parsing command: string
```

```

2. 192.168.169.128 3. 192.168.169.128 4. 192.168.169.128 5. 192.168.169.128
18/08/07 11:15:22 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 19.0 (TID 1635) in 399 ms on localhost (executor driver) (1/1) ^
18/08/07 11:15:22 INFO scheduler.DAGScheduler: ResultStage 19 (sql at DataAnalysis.scala:160) finished in 0.387 s
18/08/07 11:15:22 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 19.0, whose tasks have all completed, from pool
18/08/07 11:15:22 INFO datasources.FileFormatWriter: Job null committed.
18/08/07 11:15:22 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:15:22 INFO parser.CatalystSqlParser: Parsing command: string
18/08/07 11:15:24 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:15:24 INFO parser.CatalystSqlParser: Parsing command: string
18/08/07 11:15:24 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:15:24 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:15:25 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:15:25 INFO parser.CatalystSqlParser: Parsing command: string
18/08/07 11:15:25 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:15:26 INFO common.FileUtils: Creating directory if it doesn't exist: hdfs://localhost:8020/user/hive/warehouse/project.db/top_
10_unsubscribed_users/batchid=1
18/08/07 11:15:26 INFO metadata.Hive: Renaming src: hdfs://localhost:8020/user/hive/warehouse/project.db/top_10_unsubscribed_users/.hive-s
taaging_hive_2018-08-07_11-14-57_075_9065818365352712808-1-ext-10000/part-00000-457258dd-fca7-4394-a868-7167f4853d86-c000, dest: hdfs://lo
calhost:8020/user/hive/warehouse/project.db/top_10_unsubscribed_users/batchid=1/part-00000-457258dd-fca7-4394-a868-7167f4853d86-c000, Stat
us:true
18/08/07 11:15:28 INFO execution.SparkSqlParser: Parsing command: 'project'.top_10_unsubscribed_users'
18/08/07 11:15:29 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:15:29 INFO parser.CatalystSqlParser: Parsing command: string
18/08/07 11:15:29 INFO parser.CatalystSqlParser: Parsing command: int
18/08/07 11:15:29 INFO spark.SparkContext: Invoking stop() from shutdown hook
18/08/07 11:15:29 INFO server.AbstractConnector: Stopped Spark@427ab5a4[HTTP/1.1]{0.0.0.0:4040}
18/08/07 11:15:29 INFO ui.SparkUI: Stopped Spark web UI at http://192.168.169.128:4040
18/08/07 11:15:29 INFO spark.MapOutputTrackerEndpoint: MapOutputTrackerMasterEndpoint stopped!
18/08/07 11:15:30 INFO memory.MemoryStore: MemoryStore cleared
18/08/07 11:15:30 INFO storage.BlockManager: BlockManager stopped
18/08/07 11:15:30 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
18/08/07 11:15:30 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
18/08/07 11:15:30 INFO spark.SparkContext: Successfully stopped SparkContext
18/08/07 11:15:30 INFO util.shutdownHookManager: Shutdown hook called
18/08/07 11:15:30 INFO util.shutdownHookManager: Deleting directory /tmp/spark-ce4b0b1d-2c7a-40ab-808a-39dalac75d87
USE CASES COMPLETE !!
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost music]$ 
```

Query 2: Determine total duration of songs played by each type of user, where type of user can be 'subscribed' or 'unsubscribed'. An unsubscribed user is the one whose record is either not present in Subscribed_users lookup table or has subscription_end_date earlier than the timestamp of the song played by him.

```

hive> select user_type,duration from usersBehaviour;
OK
UNSUBSCRIBED      177678279
SUBSCRIBED        33290548
Time taken: 1.57 seconds, Fetched: 2 row(s)

```

Query 3: Determine top 10 connected artists. Connected artists are those whose songs are most listened by the unique users who follow them.

```

hive> select artist_id from connected_artists;
OK
A302
A301
Time taken: 1.505 seconds, Fetched: 2 row(s)

```

Query 4: Determine top 10 songs who have generated the maximum revenue. Royalty applies to a song only if it was liked or was completed successfully or both.

```

hive> select song_id from top_10_royalty_songs;
OK
S208
S202
S204
S205
Time taken: 2.684 seconds, Fetched: 4 row(s)

```

Query 5: Determine top 10 unsubscribed users who listened to the songs for the longest duration.

```
hive> select user_id,duration from top_10_unsubscribed_users;
OK
U118      57636973
U117      39900000
U119      31434300
U108      28807006
U103      10000000
U109      9900000
U107      0
U114      0
U106      0
U101      0
Time taken: 1.546 seconds, Fetched: 10 row(s)
```

Job Scheduling

Creating crontab to schedule the music_project_master.sh script to run for every 3 hour interval.

```
[acadgild@localhost ~]$ sudo crontab -e
[sudo] password for acadgild:
no crontab for root - using an empty one
crontab: installing new crontab
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

```
2.192.168.169.128 * */3 * * *date>>/home/acadgild/examples/music/music_project_master.sh >> /home/acadgild/examples/music/jobscheduling.log
```

Highlights of the Project

- No join of query is used while analysis. Data is already enriched with new fields and using broadcast maps on Lookup tables so as to avoid any join.
- We used full automated bash scripts from start to end.