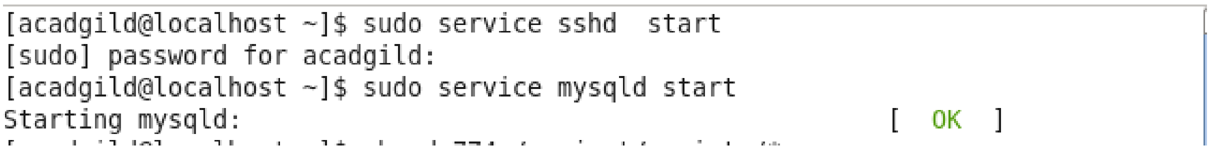
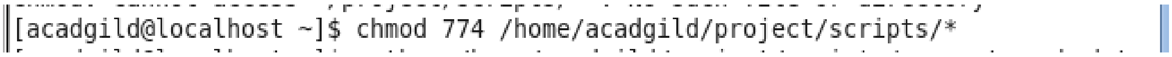
Music Data Analysis

Starting Services:



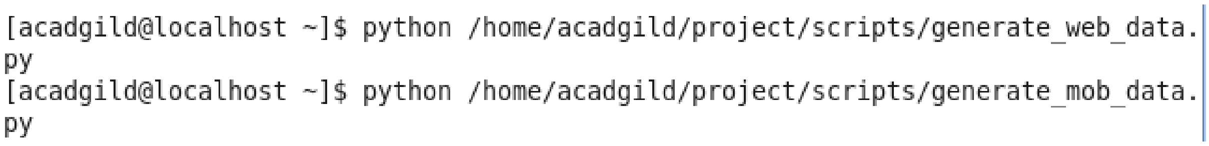
```
[acadgild@localhost ~]$ sudo service sshd  start
[sudo] password for acadgild:
[acadgild@localhost ~]$ sudo service mysqld start
Starting mysqld:                                           [  OK  ]
```

Permission set:



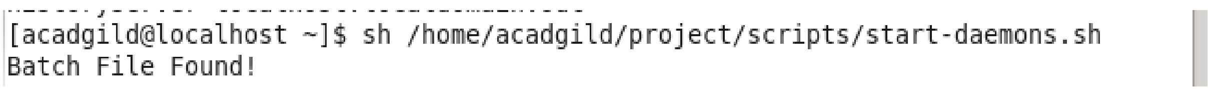
```
[acadgild@localhost ~]$ chmod 774 /home/acadgild/project/scripts/*
```

Generating Web and Mobile Data:



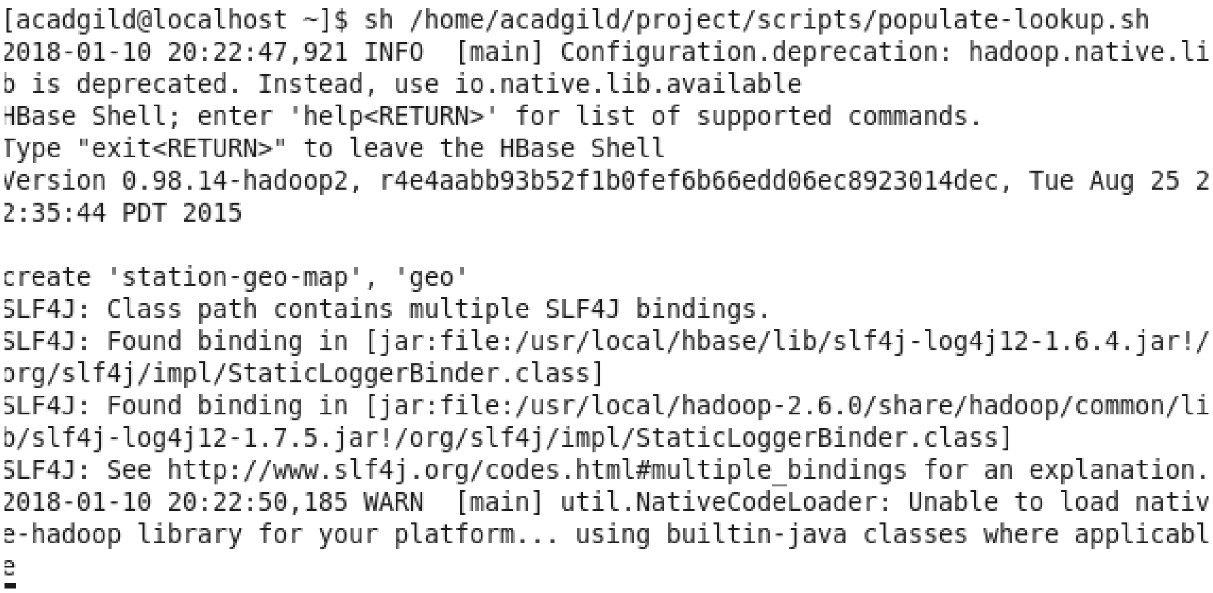
```
[acadgild@localhost ~]$ python /home/acadgild/project/scripts/generate_web_data.
py
[acadgild@localhost ~]$ python /home/acadgild/project/scripts/generate_mob_data.
py
```

Starting daemons



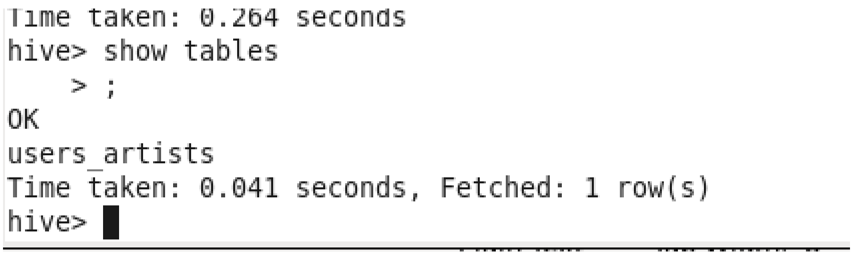
```
[acadgild@localhost ~]$ sh /home/acadgild/project/scripts/start-daemons.sh
Batch File Found!
```

Creating hbase tables and populating in hive



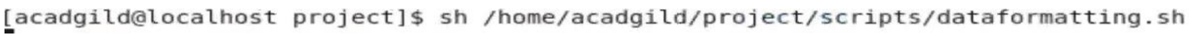
```
[acadgild@localhost ~]$ sh /home/acadgild/project/scripts/populate-lookup.sh
2018-01-10 20:22:47,921 INFO  [main] Configuration.deprecation: hadoop.native.li
b is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 0.98.14-hadoop2, r4e4aabb93b52f1b0fef6b66edd06ec8923014dec, Tue Aug 25 2
2:35:44 PDT 2015

create 'station-geo-map', 'geo'
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hbase/lib/slf4j-log4j12-1.6.4.jar!/
org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/li
b/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
2018-01-10 20:22:50,185 WARN  [main] util.NativeCodeLoader: Unable to load nativ
e-hadoop library for your platform... using builtin-java classes where applicabl
e
```

Created Table user artist in hive



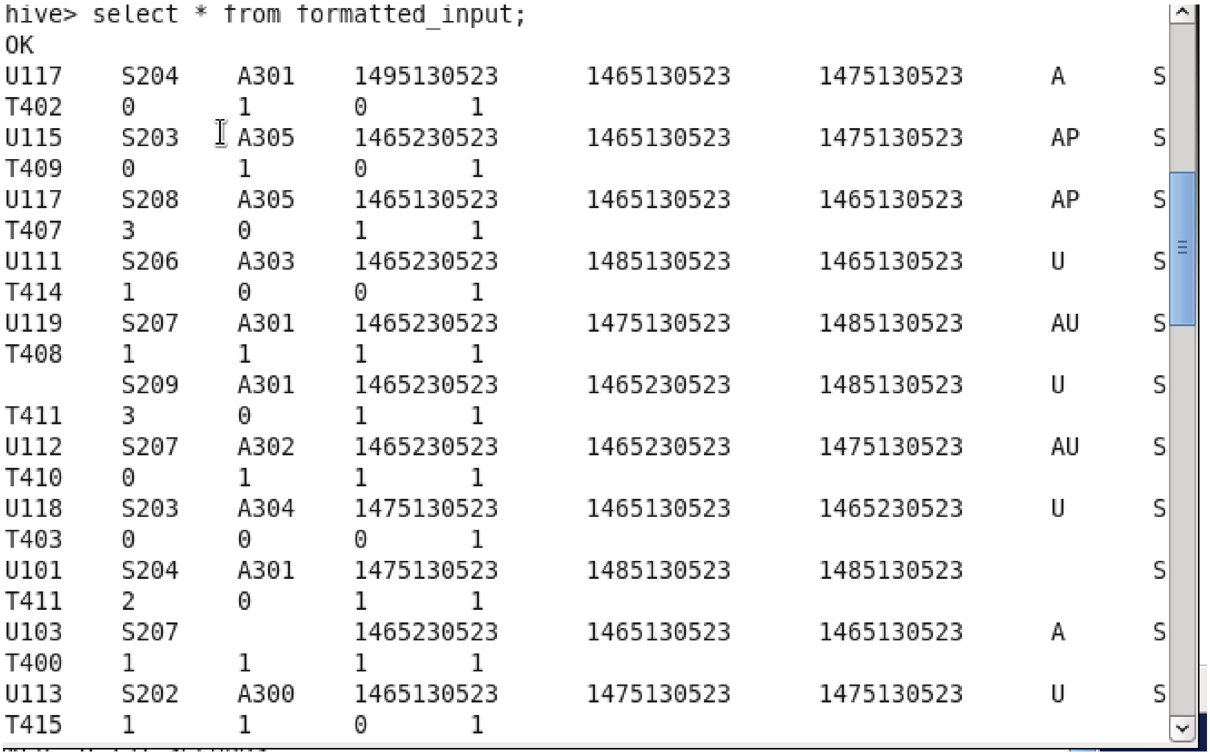
```
Time taken: 0.264 seconds
hive> show tables
    > ;
OK
users_artists
Time taken: 0.041 seconds, Fetched: 1 row(s)
hive>
```

Data Formatting



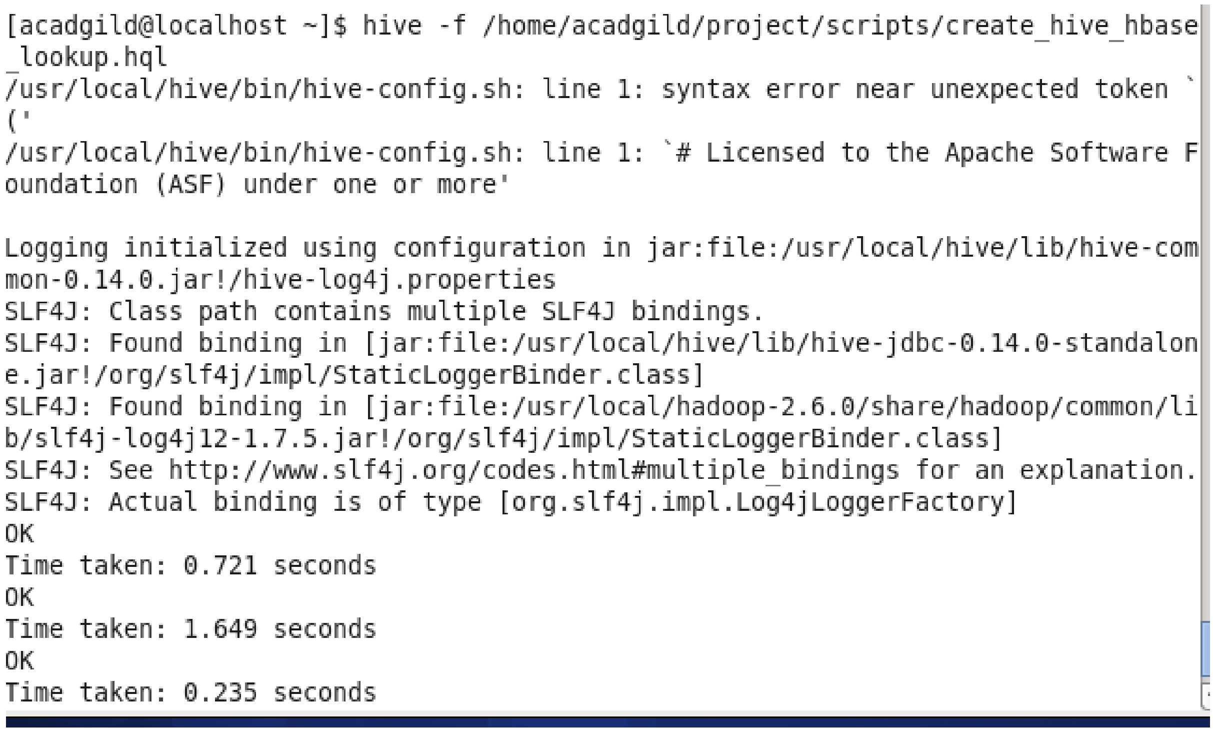
```
[acadgild@localhost project]$ sh /home/acadgild/project/scripts/dataformatting.sh
```

Formatted input table in hive



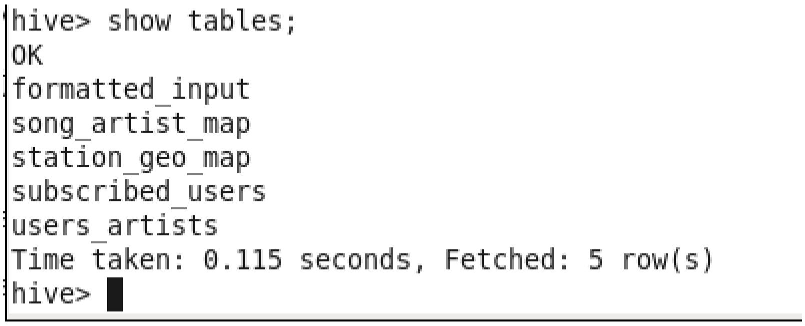
```
hive> select * from formatted_input;
OK
U117    S204    A301    1495130523      1465130523      1475130523      A      S
T402    0       1       0       1
U115    S203    A305    1465230523      1465130523      1475130523      AP     S
T409    0       1       0       1
U117    S208    A305    1465130523      1465130523      1465130523      AP     S
T407    3       0       1       1
U111    S206    A303    1465230523      1485130523      1465130523      U      S
T414    1       0       0       1
U119    S207    A301    1465230523      1475130523      1485130523      AU     S
T408    1       1       1       1
        S209    A301    1465230523      1465230523      1485130523      U      S
T411    3       0       1       1
U112    S207    A302    1465230523      1465230523      1475130523      AU     S
T410    0       1       1       1
U118    S203    A304    1475130523      1465130523      1465230523      U      S
T403    0       0       0       1
U101    S204    A301    1475130523      1485130523      1485130523             S
T411    2       0       1       1
U103    S207            1465230523      1465130523      1465130523      A      S
T400    1       1       1       1
U113    S202    A300    1465130523      1475130523      1475130523      U      S
T415    1       1       0       1
```

Creating tables in hive



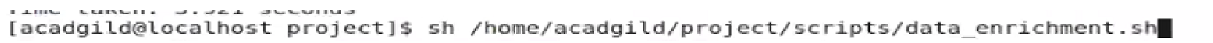
```
[acadgild@localhost ~]$ hive -f /home/acadgild/project/scripts/create_hive_hbase
_lookup.hql
/usr/local/hive/bin/hive-config.sh: line 1: syntax error near unexpected token `
('
/usr/local/hive/bin/hive-config.sh: line 1: `# Licensed to the Apache Software F
oundation (ASF) under one or more'

Logging initialized using configuration in jar:file:/usr/local/hive/lib/hive-com
mon-0.14.0.jar!/hive-log4j.properties
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/local/hive/lib/hive-jdbc-0.14.0-standalon
e.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/local/hadoop-2.6.0/share/hadoop/common/li
b/slf4j-log4j12-1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
OK
Time taken: 0.721 seconds
OK
Time taken: 1.649 seconds
OK
Time taken: 0.235 seconds
```

Created hive tables



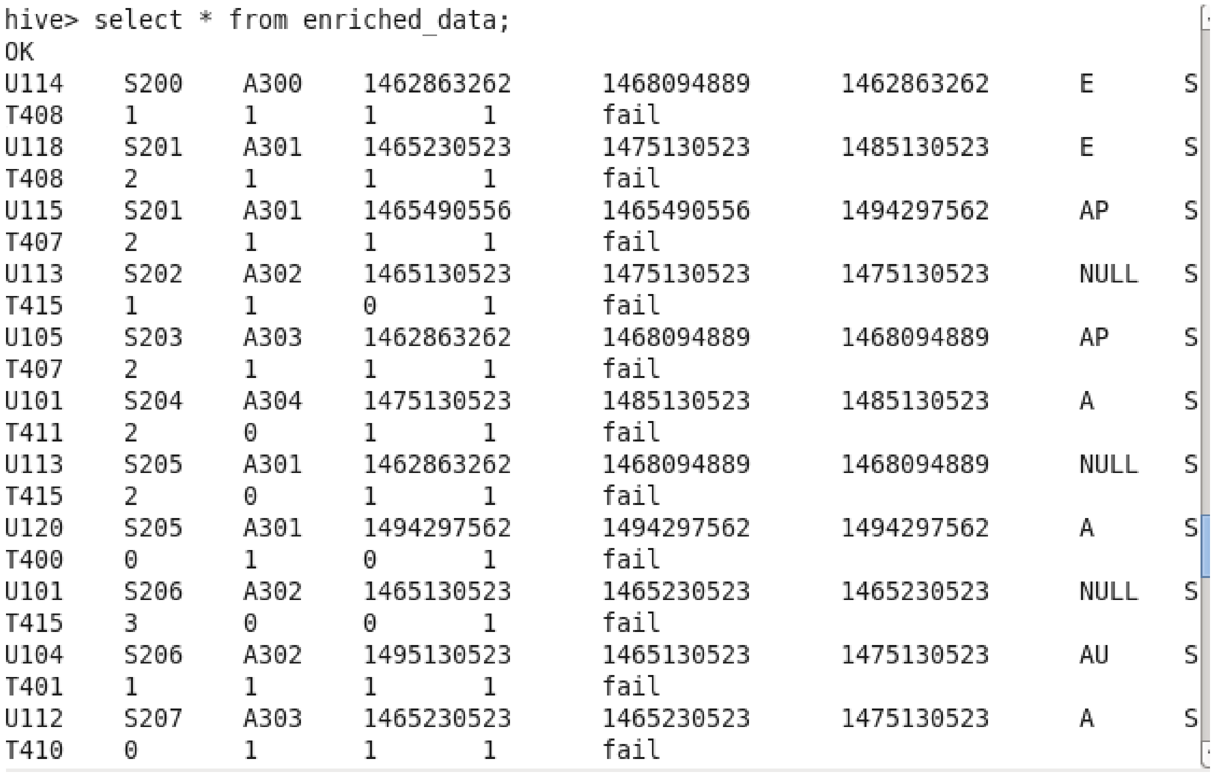
```
hive> show tables;
OK
formatted_input
song_artist_map
station_geo_map
subscribed_users
users_artists
Time taken: 0.115 seconds, Fetched: 5 row(s)
hive>
```

Data enrichment:



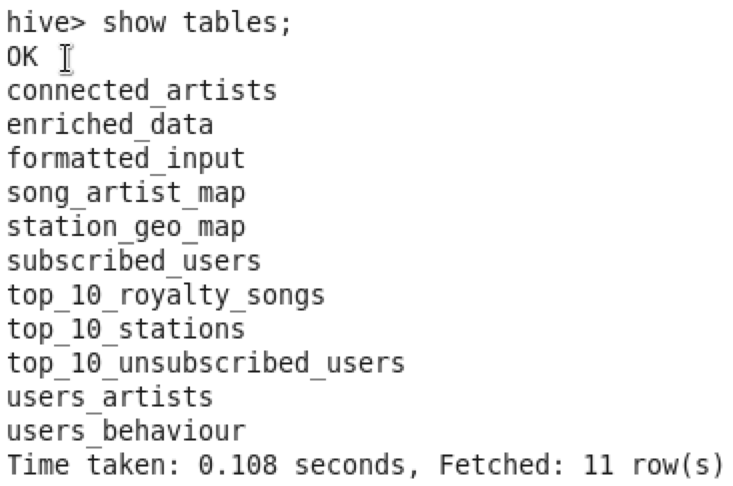
```
Time taken: 3.321 seconds
[acadgild@localhost project]$ sh /home/acadgild/project/scripts/data_enrichment.sh
```

Enriched Data:



```
hive> select * from enriched_data;
OK
U114    S200    A300    1462863262      1468094889      1462863262      E       S
T408    1       1       1       1       fail
U118    S201    A301    1465230523      1475130523      1485130523      E       S
T408    2       1       1       1       fail
U115    S201    A301    1465490556      1465490556      1494297562      AP      S
T407    2       1       1       1       fail
U113    S202    A302    1465130523      1475130523      1475130523      NULL    S
T415    1       1       0       1       fail
U105    S203    A303    1462863262      1468094889      1468094889      AP      S
T407    2       1       1       1       fail
U101    S204    A304    1475130523      1485130523      1485130523      A       S
T411    2       0       1       1       fail
U113    S205    A301    1462863262      1468094889      1468094889      NULL    S
T415    2       0       1       1       fail
U120    S205    A301    1494297562      1494297562      1494297562      A       S
T400    0       1       0       1       fail
U101    S206    A302    1465130523      1465230523      1465230523      NULL    S
T415    3       0       0       1       fail
U104    S206    A302    1495130523      1465130523      1475130523      AU      S
T401    1       1       1       1       fail
U112    S207    A303    1465230523      1465230523      1475130523      A       S
T410    0       1       1       1       fail
```

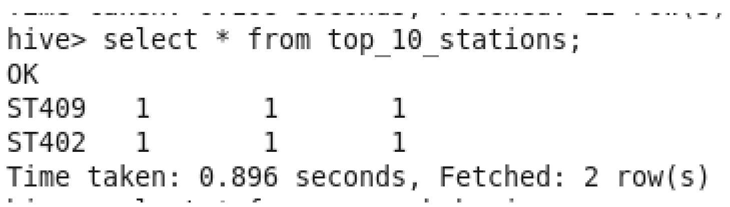
After Data Analysis created Tables:



```
hive> show tables;
OK
connected_artists
enriched_data
formatted_input
song_artist_map
station_geo_map
subscribed_users
top_10_royalty_songs
top_10_stations
top_10_unsubscribed_users
users_artists
users_behaviour
Time taken: 0.108 seconds, Fetched: 11 row(s)
```

1. Determine top 10 station_id(s) where maximum number of songs were played, which were

liked by unique users.



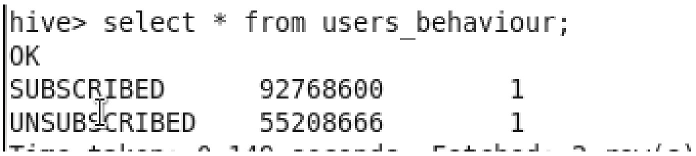
```
hive> select * from top_10_stations;
OK
ST409   1       1       1
ST402   1       1       1
Time taken: 0.896 seconds, Fetched: 2 row(s)
```

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

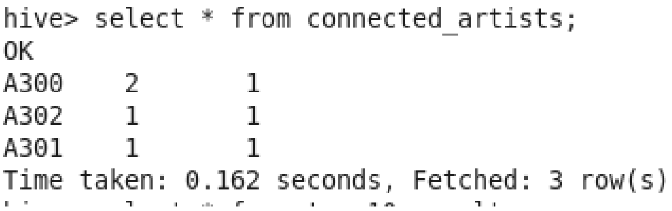
timestamp of the song played by him.



```
hive> select * from users_behaviour;
OK
SUBSCRIBED      92768600        1
UNSUBSCRIBED    55208666        1
Time taken: 0.140 seconds, Fetched: 2 row(s)
```

3. Determine top 10 connected artists. Connected artists are those whose songs are most

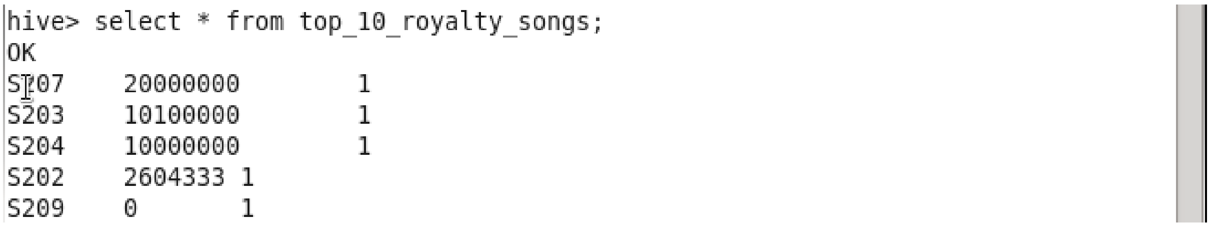
listened by the unique users who follow them.



```
hive> select * from connected_artists;
OK
A300    2       1
A302    1       1
A301    1       1
Time taken: 0.162 seconds, Fetched: 3 row(s)
```

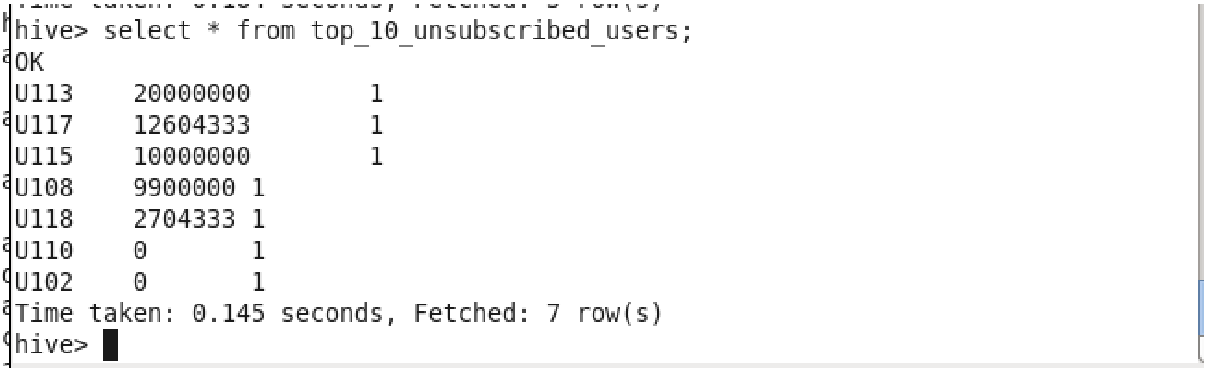
4. Determine top 10 songs who have generated the maximum revenue. Royalty applies to a

song only if it was liked or was completed successfully or both.



```
hive> select * from top_10_royalty_songs;
OK
S207    20000000        1
S203    10100000        1
S204    10000000        1
S202    2604333 1
S209    0       1
```

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



```
hive> select * from top_10_unsubscribed_users;
OK
U113    20000000        1
U117    12604333        1
U115    10000000        1
U108    9900000 1
U118    2704333 1
U110    0       1
U102    0       1
Time taken: 0.145 seconds, Fetched: 7 row(s)
hive> 
```