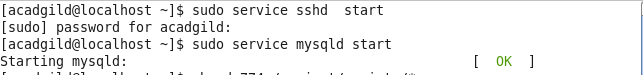
Project

MUSIC DATA ANALYSIS

**Music Data Analysis**

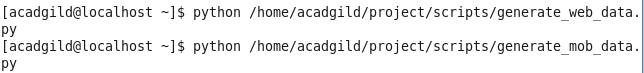
**Starting Services:**

****

**Permission set:**

****

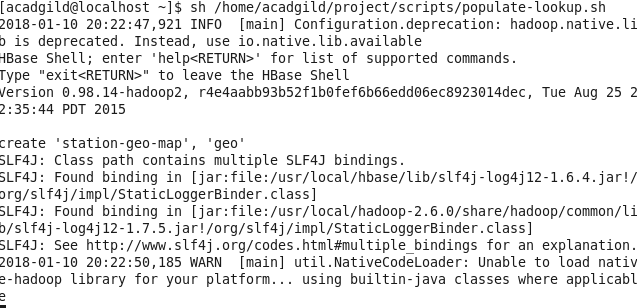
**Generating Web and Mobile Data:**

****

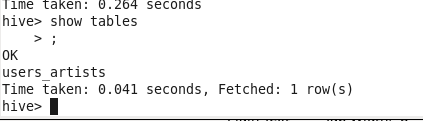
**Starting daemons**

****

**Creating hbase tables and populating in hive**

****

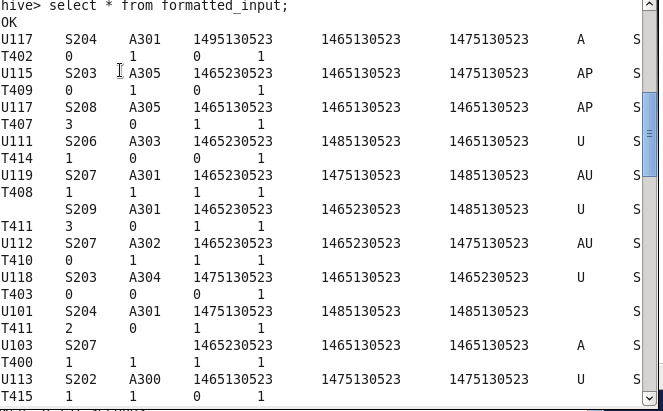
**Created Table user artist in hive**

****

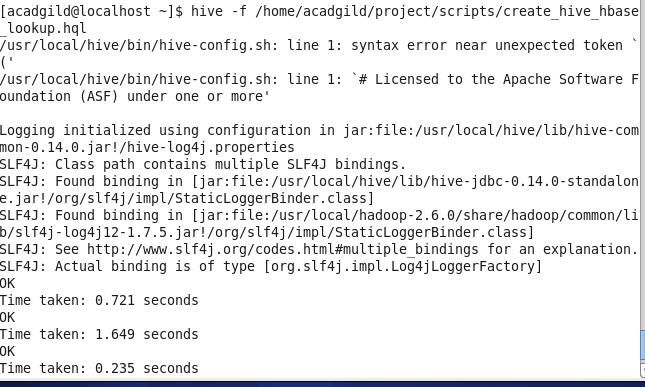
**Data Formatting**

****

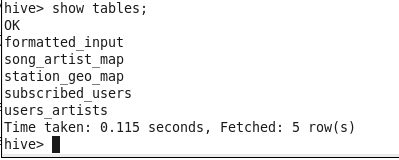
**Formatted input table in hive**

****

**Creating tables in hive**

****

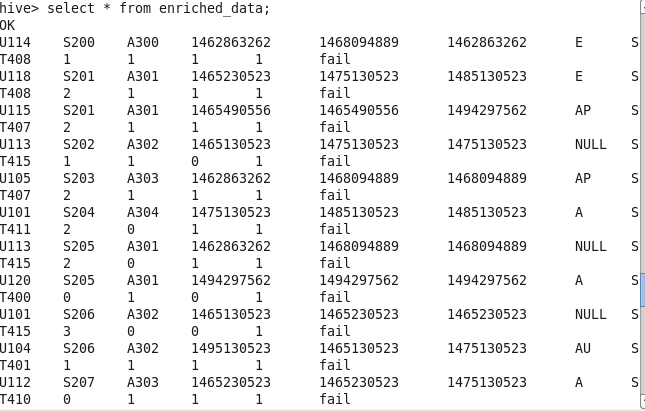
**Created hive tables**

****

**Data enrichment:**

****

**Enriched Data:**

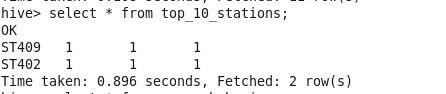
****

**After Data Analysis created Tables:**

****

**1. Determine top 10 station\_id(s) where maximum number of songs were played, which were**

**liked by unique users.**

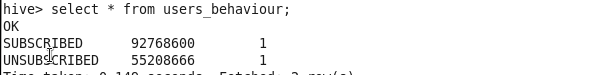
****

**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**

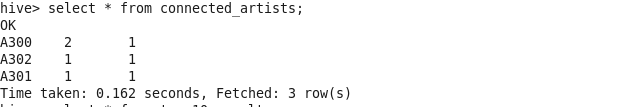
**present in Subscribed\_users lookup table or has subscription\_end\_date earlier than the**

**timestamp of the song played by him.**

****

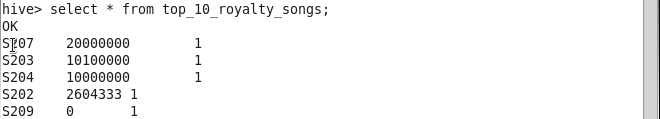
**3. Determine top 10 connected artists. Connected artists are those whose songs are most**

**listened by the unique users who follow them.**

****

**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.**

****

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

****