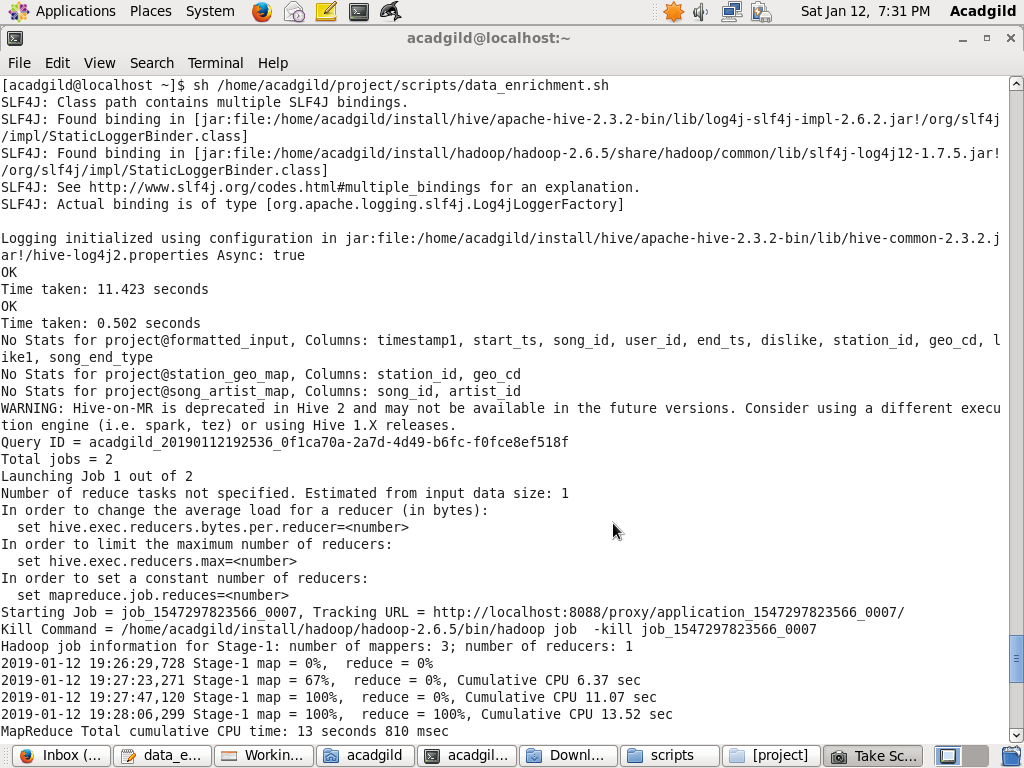
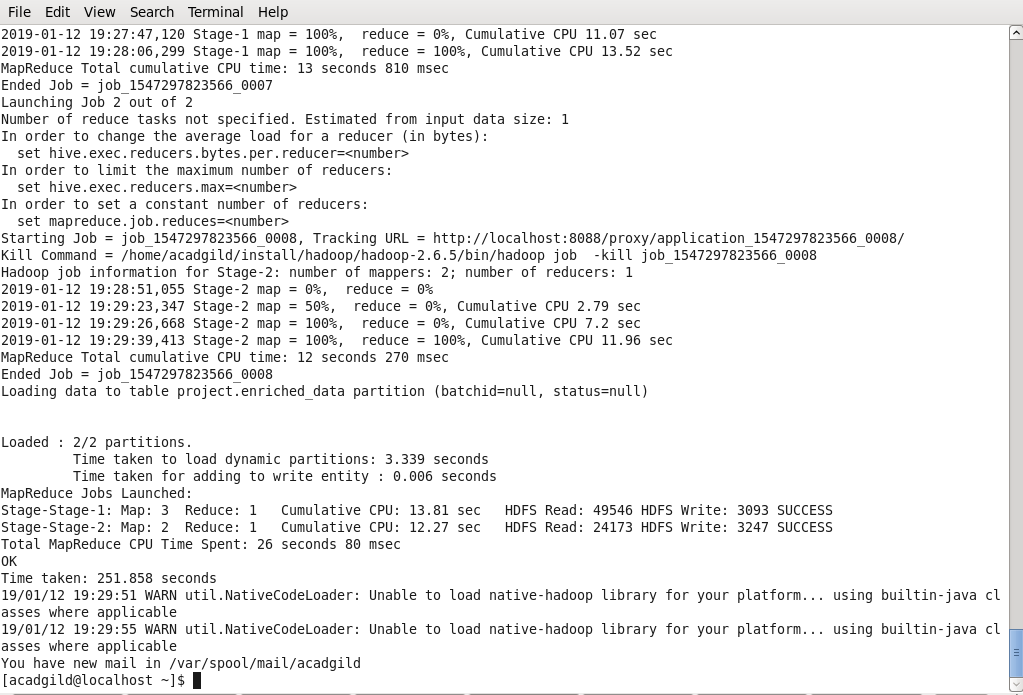
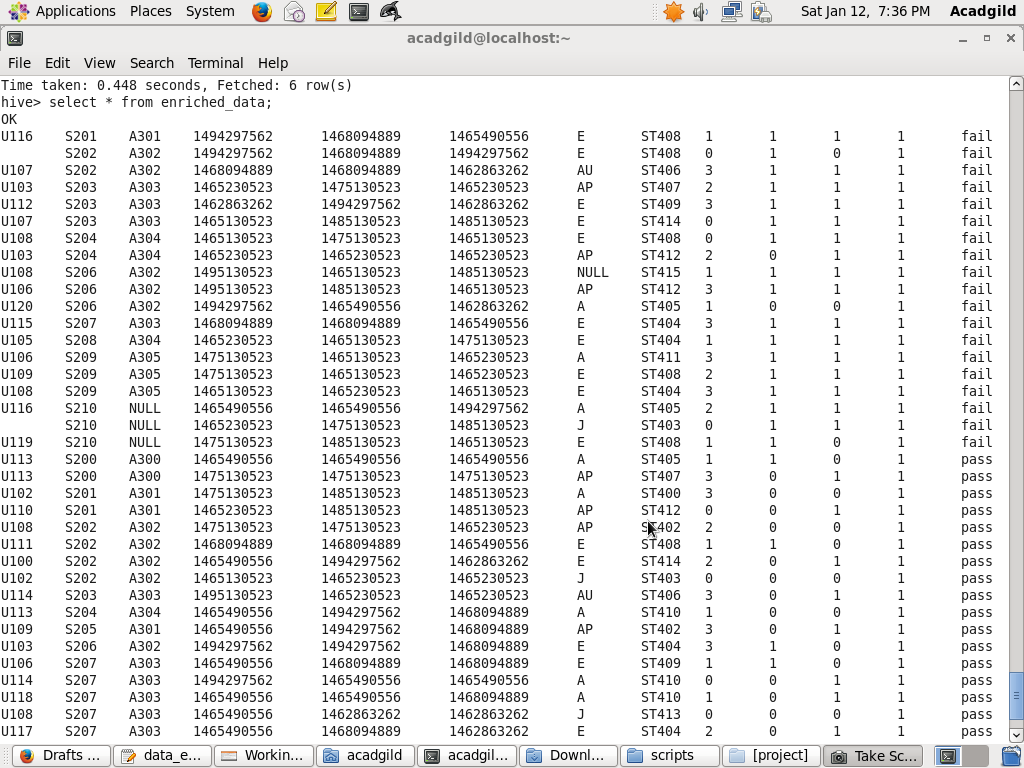
**Data Enrichment:**

Note : The scripts are placed in separate folder named as Scripts – Kinldy refer to it.

****

****

****

[acadgild@localhost ~]$ sh /home/acadgild/project/scripts/data\_enrichment.sh  
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/StaticLoggerBinder.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/slf4j/impl/StaticLoggerBinder.class]  
SLF4J: See [http://www.slf4j.org/codes.html#multiple\_bindings](http://www.slf4j.org/codes.html" \l "multiple_bindings" \t "_blank) 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-log4j2.properties Async: true  
OK  
Time taken: 11.423 seconds  
OK  
Time taken: 0.502 seconds  
No Stats for project@formatted\_input, Columns: timestamp1, start\_ts, song\_id, user\_id, end\_ts, dislike, station\_id, geo\_cd, like1, song\_end\_type  
No Stats for project@station\_geo\_map, Columns: station\_id, geo\_cd  
No Stats for project@song\_artist\_map, Columns: song\_id, artist\_id  
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.  
Query ID = acadgild\_20190112192536\_0f1ca70a-2a7d-4d49-b6fc-f0fce8ef518f  
Total jobs = 2  
Launching Job 1 out of 2  
Number of reduce tasks not specified. Estimated from input data size: 1  
In order to change the average load for a reducer (in bytes):  
  set hive.exec.reducers.bytes.per.reducer=<number>  
In order to limit the maximum number of reducers:  
  set hive.exec.reducers.max=<number>  
In order to set a constant number of reducers:  
  set mapreduce.job.reduces=<number>  
Starting Job = job\_1547297823566\_0007, Tracking URL = [http://localhost:8088/proxy/application\_1547297823566\_0007/](http://localhost:8088/proxy/application_1547297823566_0007/" \t "_blank)  
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job\_1547297823566\_0007  
Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 1  
2019-01-12 19:26:29,728 Stage-1 map = 0%,  reduce = 0%  
2019-01-12 19:27:23,271 Stage-1 map = 67%,  reduce = 0%, Cumulative CPU 6.37 sec  
2019-01-12 19:27:47,120 Stage-1 map = 100%,  reduce = 0%, Cumulative CPU 11.07 sec  
2019-01-12 19:28:06,299 Stage-1 map = 100%,  reduce = 100%, Cumulative CPU 13.52 sec  
MapReduce Total cumulative CPU time: 13 seconds 810 msec  
Ended Job = job\_1547297823566\_0007  
Launching Job 2 out of 2  
Number of reduce tasks not specified. Estimated from input data size: 1  
In order to change the average load for a reducer (in bytes):  
  set hive.exec.reducers.bytes.per.reducer=<number>  
In order to limit the maximum number of reducers:  
  set hive.exec.reducers.max=<number>  
In order to set a constant number of reducers:  
  set mapreduce.job.reduces=<number>  
Starting Job = job\_1547297823566\_0008, Tracking URL = [http://localhost:8088/proxy/application\_1547297823566\_0008/](http://localhost:8088/proxy/application_1547297823566_0008/" \t "_blank)  
Kill Command = /home/acadgild/install/hadoop/hadoop-2.6.5/bin/hadoop job  -kill job\_1547297823566\_0008  
Hadoop job information for Stage-2: number of mappers: 2; number of reducers: 1  
2019-01-12 19:28:51,055 Stage-2 map = 0%,  reduce = 0%  
2019-01-12 19:29:23,347 Stage-2 map = 50%,  reduce = 0%, Cumulative CPU 2.79 sec  
2019-01-12 19:29:26,668 Stage-2 map = 100%,  reduce = 0%, Cumulative CPU 7.2 sec  
2019-01-12 19:29:39,413 Stage-2 map = 100%,  reduce = 100%, Cumulative CPU 11.96 sec  
MapReduce Total cumulative CPU time: 12 seconds 270 msec  
Ended Job = job\_1547297823566\_0008  
Loading data to table project.enriched\_data partition (batchid=null, status=null)  
  
  
Loaded : 2/2 partitions.  
     Time taken to load dynamic partitions: 3.339 seconds  
     Time taken for adding to write entity : 0.006 seconds  
MapReduce Jobs Launched:   
Stage-Stage-1: Map: 3  Reduce: 1   Cumulative CPU: 13.81 sec   HDFS Read: 49546 HDFS Write: 3093 SUCCESS  
Stage-Stage-2: Map: 2  Reduce: 1   Cumulative CPU: 12.27 sec   HDFS Read: 24173 HDFS Write: 3247 SUCCESS  
Total MapReduce CPU Time Spent: 26 seconds 80 msec  
OK  
Time taken: 251.858 seconds  
19/01/12 19:29:51 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
19/01/12 19:29:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$