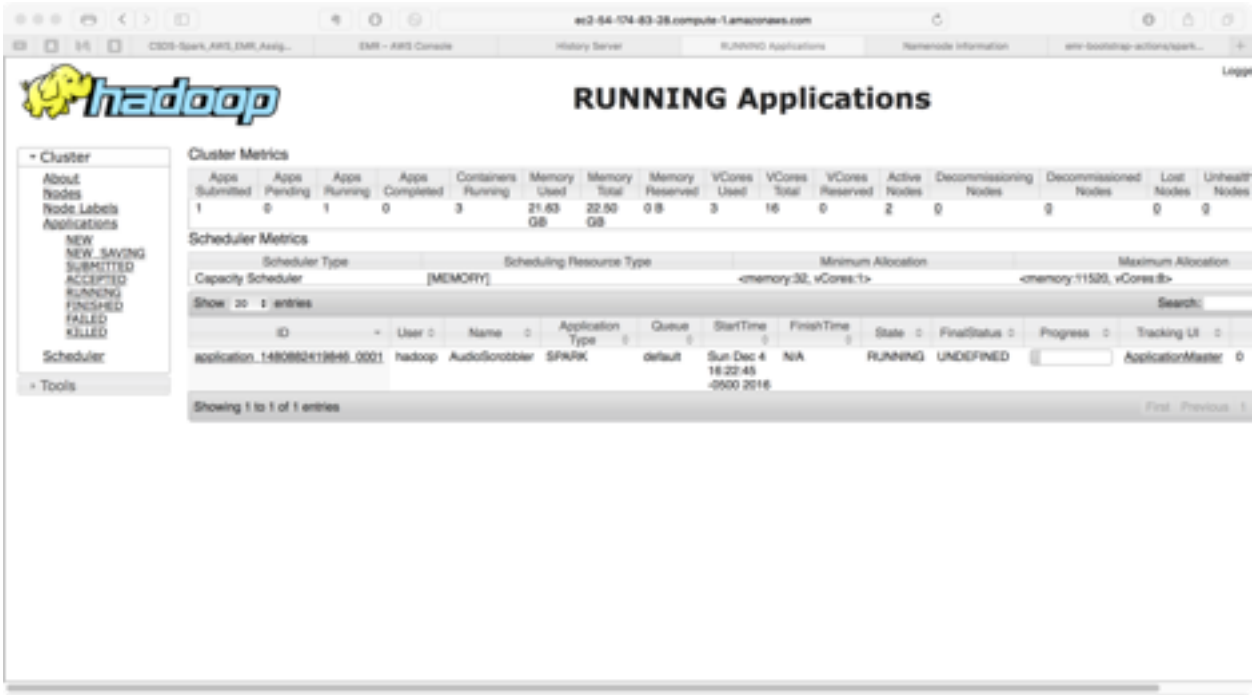
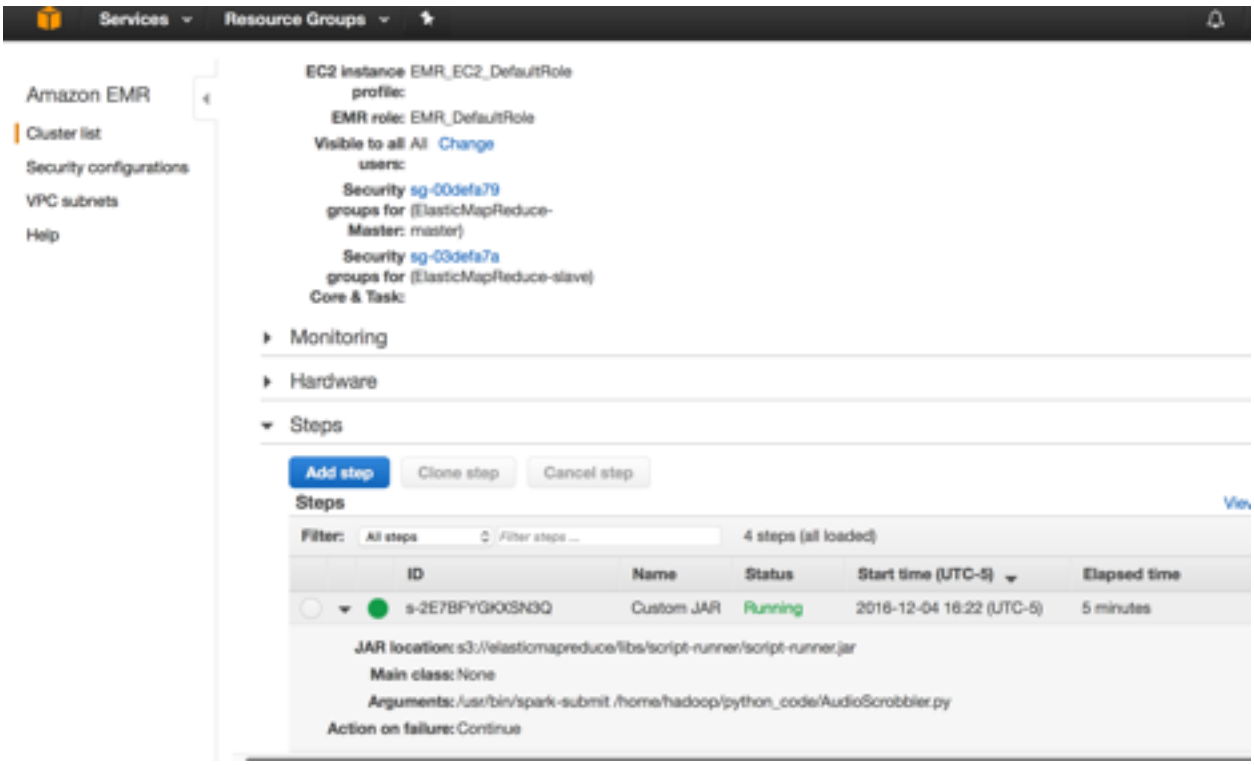
Sunday, December 4, 2016

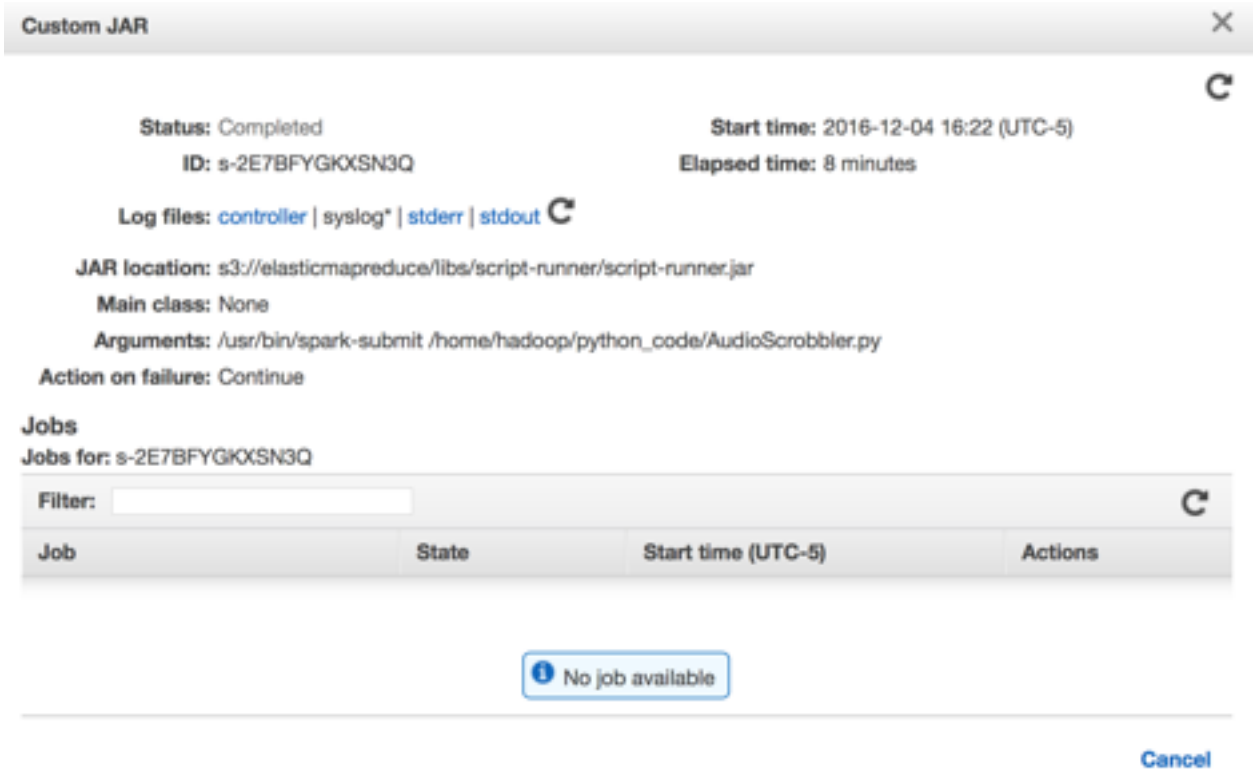
HW3: Audio Scrobbler

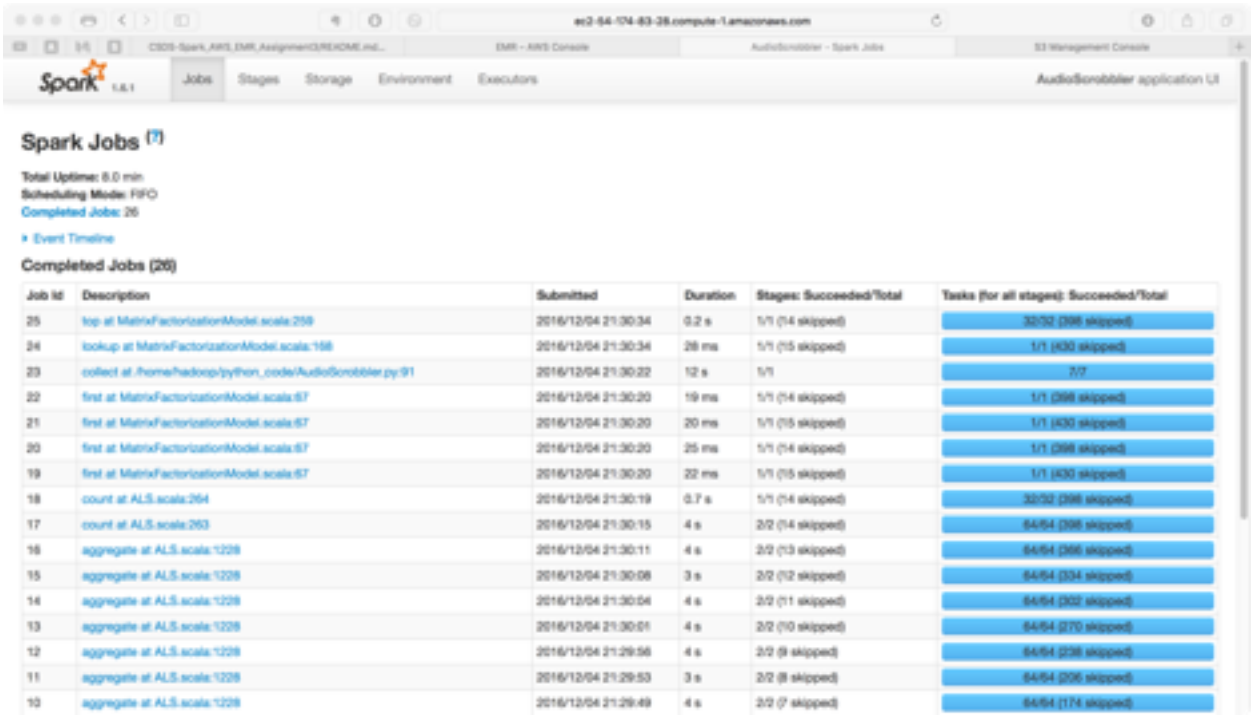
Submitted by: ng2565





Step added and completed screenshots:




Spark 1.6.1
Jobs
Stages
Storage
Environment
Executors
AudioScrobbler application UI
Spark Jobs [7]
Total Uptime: 8.0 min
Scheduling Mode: FIFO
Completed Jobs: 26
Event Timeline
Completed Jobs (26)
Job Id
Description
Submitted
Duration
Stages: Succeeded/Total
Tasks (for all stages): Succeeded/Total
25 top at MatrixFactorizationModel.scala:259 2016/12/04 21:30:34 0.2 s 1/1 (14 skipped) 32/32 (398 skipped)
24 lookup at MatrixFactorizationModel.scala:168 2016/12/04 21:30:34 28 ms 1/1 (15 skipped) 1/1 (430 skipped)
23 collect at /home/hadoop/python_code/AudioScrobbler.py:91 2016/12/04 21:30:22 12 s 1/1 7/7
22 first at MatrixFactorizationModel.scala:67 2016/12/04 21:30:20 19 ms 1/1 (14 skipped) 1/1 (398 skipped)
21 first at MatrixFactorizationModel.scala:67 2016/12/04 21:30:20 20 ms 1/1 (15 skipped) 1/1 (430 skipped)
20 first at MatrixFactorizationModel.scala:67 2016/12/04 21:30:20 25 ms 1/1 (14 skipped) 1/1 (398 skipped)
19 first at MatrixFactorizationModel.scala:67 2016/12/04 21:30:20 22 ms 1/1 (15 skipped) 1/1 (430 skipped)
18 count at ALS.scala:264 2016/12/04 21:30:19 0.7 s 1/1 (14 skipped) 32/32 (398 skipped)
17 count at ALS.scala:263 2016/12/04 21:30:15 4 s 2/2 (14 skipped) 64/64 (398 skipped)
16 aggregate at ALS.scala:1228 2016/12/04 21:30:11 4 s 2/2 (13 skipped) 64/64 (366 skipped)
15 aggregate at ALS.scala:1228 2016/12/04 21:30:08 3 s 2/2 (12 skipped) 64/64 (334 skipped)
14 aggregate at ALS.scala:1228 2016/12/04 21:30:04 4 s 2/2 (11 skipped) 64/64 (302 skipped)
13 aggregate at ALS.scala:1228 2016/12/04 21:30:01 4 s 2/2 (10 skipped) 64/64 (270 skipped)
12 aggregate at ALS.scala:1228 2016/12/04 21:29:56 4 s 2/2 (9 skipped) 64/64 (238 skipped)
11 aggregate at ALS.scala:1228 2016/12/04 21:29:53 3 s 2/2 (8 skipped) 64/64 (206 skipped)
10 aggregate at ALS.scala:1228 2016/12/04 21:29:49 4 s 2/2 (7 skipped) 64/64 (174 skipped)

Screenshot of output of the model: hyper parameters used are: rank=50, iterations=5, lambda_=1.0, alpha=40.0)

Printing artists listened to by User ID: 2093760 and the 10 recommendations for this user: Highlighted output shows the artists listened to by the user

Spark ALS model recommends the 10 artists printed below:

```
50 Cent
Eminem
2Pac
Jay-Z
Snoop Dogg
Kanye West
[unknown]
Dr. Dre
Outkast
D12
```