


HW3: Audio Scrobbler

Submitted by: ng2565



RUNNING Applications

Cluster

- About
- Nodes
- Node Labels
- Applications
 - NEW
 - NEW SAVING
 - SUBMITTED
 - ACCEPTED
 - RUNNING
 - FINISHED
 - FAILED
 - KILLED

Scheduler

Tools

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Memory Used	Memory Total	Memory Reserved	VCores Used	VCores Total	VCores Reserved	Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes
1	0	1	0	3	21.63 GB	22.50 GB	0 B	3	16	0	2	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation
Capacity Scheduler	[MEMORY]	<memory:32, vCores:1>	<memory:11520, vCores:8>

Show 20 entries

ID	User	Name	Application Type	Queue	StartTime	FinishTime	State	FinalStatus	Progress	Tracking UI
application_1480862419846_0001	hadoop	AudioScrobbler	SPARK	default	Sun Dec 4 16:22:45 -0500 2016	N/A	RUNNING	UNDEFINED		ApplicationMaster 0

Showing 1 to 1 of 1 entries

Services

Resource Groups

Amazon EMR

- Cluster list
- Security configurations
- VPC subnets
- Help

EC2 instance EMR_EC2_DefaultRole

profile: EMR_EC2_DefaultRole

EMR role: EMR_DefaultRole

Visible to all: All [Change](#)

users:

Security sg-00defa79

groups for (ElasticMapReduce-Master: master)

Security sg-03defa7a

groups for (ElasticMapReduce-slave)

Core & Task:

Monitoring

Hardware

Steps

Add step

Clone step

Cancel step

Steps

Filter: All steps [Filter steps ...](#) 4 steps (all loaded)

ID	Name	Status	Start time (UTC-5)	Elapsed time
s-2E7BFGKX3NGQ	Custom JAR	Running	2016-12-04 16:22 (UTC-5)	5 minutes

JAR location: s3://elasticmapreduce/libs/script-runner/script-runner.jar

Main class: None

Arguments: /usr/bin/spark-submit /home/hadoop/python_code/AudioScrobbler.py

Action on failure: Continue

1

Step added and completed screenshots:

Custom JAR

Status: Completed

ID: s-2E7BFYGKXSN3Q

Log files: [controller](#) | [syslog*](#) | [stderr](#) | [stdout](#)

JAR location: s3://elasticmapreduce/libs/script-runner/script-runner.jar

Main class: None

Arguments: /usr/bin/spark-submit /home/hadoop/python_code/AudioScrobbler.py

Action on failure: Continue

Start time: 2016-12-04 16:22 (UTC-5)

Elapsed time: 8 minutes

Jobs

Jobs for: s-2E7BFYGKXSN3Q

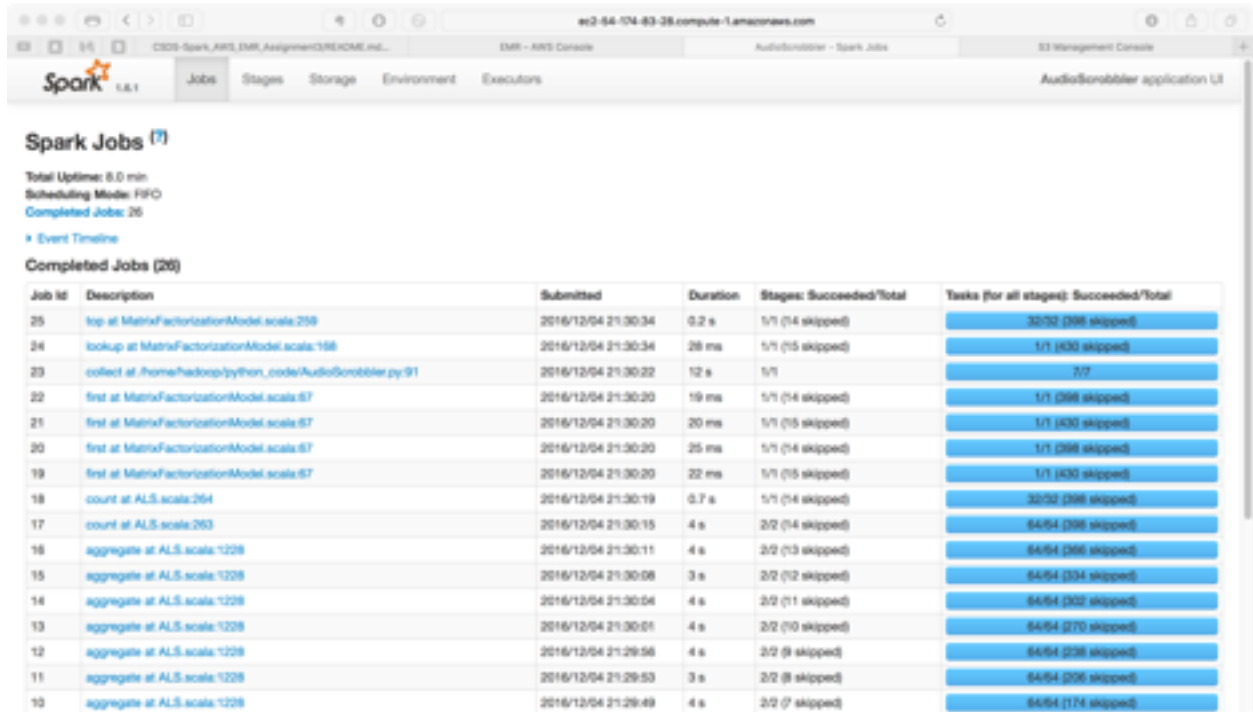
Filter:

Job	State	Start time (UTC-5)	Actions
-----	-------	--------------------	---------

No job available

Cancel

Sunday, December 4, 2016



Spark Jobs ⁽⁷⁾

Total Uptime: 8.0 min
Scheduling Model: 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.scale:258	2016/12/04 21:30:34	0.2 s	1/1 (14 skipped)	32/32 (298 skipped)
24	lookup at MatrixFactorizationModel.scale:198	2016/12/04 21:30:34	28 ms	1/1 (15 skipped)	1/1 (400 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.scale:67	2016/12/04 21:30:20	19 ms	1/1 (14 skipped)	1/1 (298 skipped)
21	first at MatrixFactorizationModel.scale:67	2016/12/04 21:30:20	20 ms	1/1 (15 skipped)	1/1 (400 skipped)
20	first at MatrixFactorizationModel.scale:67	2016/12/04 21:30:20	25 ms	1/1 (14 skipped)	1/1 (298 skipped)
19	first at MatrixFactorizationModel.scale:67	2016/12/04 21:30:20	22 ms	1/1 (15 skipped)	1/1 (400 skipped)
18	count at ALS.scale:264	2016/12/04 21:30:19	0.7 s	1/1 (14 skipped)	32/32 (298 skipped)
17	count at ALS.scale:263	2016/12/04 21:30:15	4 s	2/2 (14 skipped)	64/64 (298 skipped)
16	aggregate at ALS.scale:1228	2016/12/04 21:30:11	4 s	2/2 (13 skipped)	64/64 (298 skipped)
15	aggregate at ALS.scale:1228	2016/12/04 21:30:08	3 s	2/2 (12 skipped)	64/64 (334 skipped)
14	aggregate at ALS.scale:1228	2016/12/04 21:30:04	4 s	2/2 (11 skipped)	64/64 (352 skipped)
13	aggregate at ALS.scale:1228	2016/12/04 21:30:01	4 s	2/2 (10 skipped)	64/64 (270 skipped)
12	aggregate at ALS.scale:1228	2016/12/04 21:29:56	4 s	2/2 (9 skipped)	64/64 (238 skipped)
11	aggregate at ALS.scale:1228	2016/12/04 21:29:53	3 s	2/2 (8 skipped)	64/64 (206 skipped)
10	aggregate at ALS.scale:1228	2016/12/04 21:29:49	4 s	2/2 (7 skipped)	64/64 (174 skipped)

Sunday, December 4, 2016

A screenshot of the AWS Logs console interface. The browser address bar shows 'aws-logs-648564116187-us-east-1.s3.amazonaws.com'. The console displays a list of artists: [u'David Gray', u'The Sew Doctors', u'Blackalicious', u'Jurassic 5', u'Kazibit']. Below this, a list of 10 recommended artists is shown: 50 Cent, Eminem, 2Pac, Jay-Z, Snoop Dogg, Kanye West, [unknown], Dr. Dre, Outkast, and D12.

```
[u'David Gray', u'The Sew Doctors', u'Blackalicious', u'Jurassic 5', u'Kazibit']
50 Cent
Eminem
2Pac
Jay-Z
Snoop Dogg
Kanye West
[unknown]
Dr. Dre
Outkast
D12
```

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
```