Background	1
EMR cluster configuration steps	2
Cluster setup - 1 - software and steps	3
Cluster setup - 2 - hardware	4
Cluster setup - 3 - general cluster settings	5
Cluster setup - 4 - security	6
The bootstrap action	6
Cluster initialization error	7
Following the logs	7
The bootstrapping log (stdout)	7
The provision-node apps-phase error messages	13
Things we tried	17
References	18

## Background

We are in the process of configuring an EMR cluster with a Step intended to run a Tensorflow-based machine learning job.

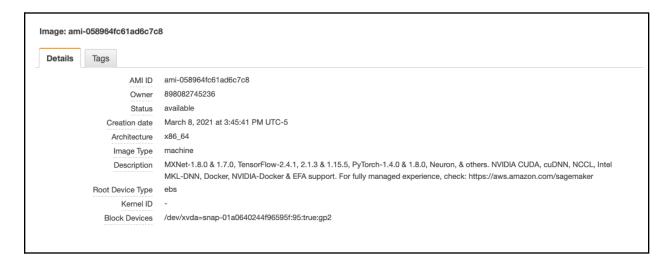
We want to leverage the support for GPU-based deep learning in Tensorflow and in AWS in order to distribute and scale the model training for our ML job.

Ideally, we would like to have a single step comprise three phases: a) data preparation with Spark; b) training a Deep Learning model (Tensorflow+Spark); and c) post-processing the results and persisting them into AWS S3, also with Spark.

For this reason, we're looking into integrating a ready-made Amazon DL AMI by loading it into the EMR cluster.

## EMR cluster configuration steps

In terms of the selected DLAMI, we're focusing on the very latest version 42.0 (ami ID = ami-058964fc61ad6c7c8):



Note that this DLAMI has TensorFlow 2.4.1 and does not specify the NVIDIA CUDA version.

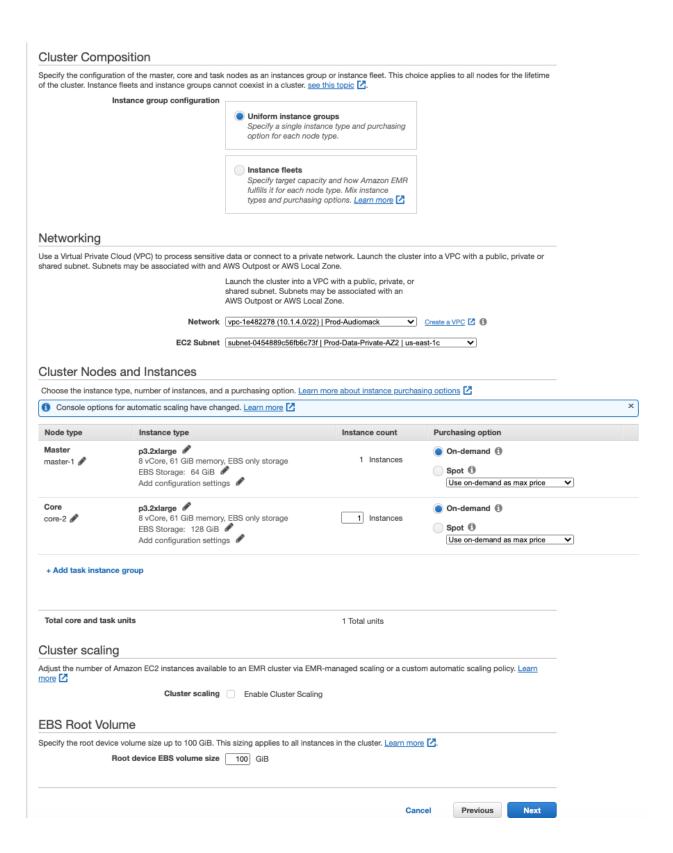
Below are the EMR cluster configuration steps, executed with EMR 6.2.0 selected and Spark 3.0.1:

# Cluster setup - 1 - software and steps

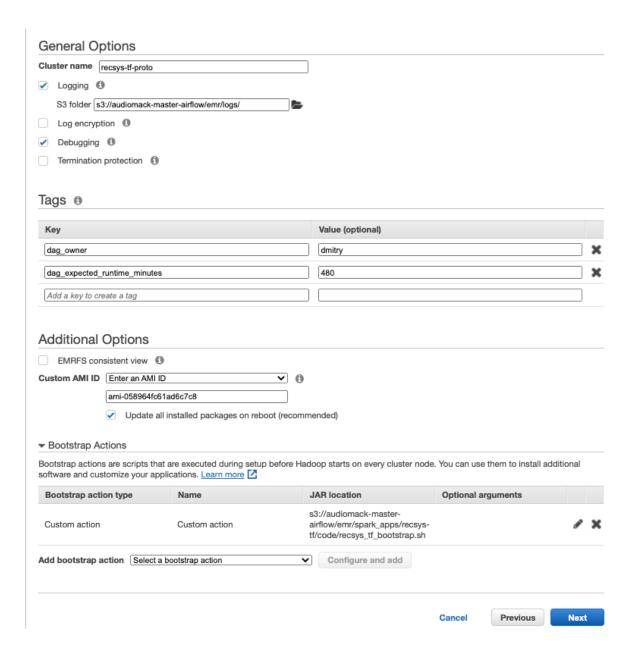
Create Cluster - Advanced Options Go to quick options

Step 1: Software and Steps	Software Configura	ation				
Step 2: Hardware	Release emr-6.2.0					
Step 3: General Cluster Settings	✓ Hadoop 3.2.1	Zeppelin	0.9.0	Livy 0.7.0		
Step 4: Security	JupyterHub 1.1.0	Tez 0.9.2		Flink 1.11.2		
	Ganglia 3.7.2	HBase 2	.2.6-amzn-0	Pig 0.17.0		
	✓ Hive 3.1.2	Presto 0.	238.3	PrestoSQL 343		
	ZooKeeper 3.4.14	JupyterE	nterpriseGateway 2.1.0	MXNet 1.7.0		
	Sqoop 1.4.7	Hue 4.8.	0	Phoenix 5.0.0		
	Oozie 5.2.0	✓ Spark 3.	0.1	HCatalog 3.1.2		
	TensorFlow 2.3.1					
	Multiple master nodes (optional)					
	Use multiple master nod	Use multiple master nodes to improve cluster availability. Learn more				
	AWS Glue Data Catalog s	AWS Glue Data Catalog settings (optional)				
	Use for Hive table metac	Use for Hive table metadata 6				
	Use for Spark table metadata					
	Edit software settings					
	Enter configuration	■ Enter configuration   Load JSON from S3				
	[{"configurations":[{"classification":"export","properties": {"PYSPARK_PYTHON":'/usr/bin/python3'})],"classification":"spark-env","properties":{}}, {"classification":"yarn-site","properties":{"yarn.nodemanager.disk-health-checker.max- Steps (optional)					
	Concurrency: Run multiple steps at the same time to improve cluster utilization  After last step completes: Clusters waiting state  Cluster auto-terminates  Step type Select a step					
	Name	Action on failure	JAR location	Arguments		
	recsys_tfrs_proto	Cancel and wait	command-runner.jar	users=s3://audiomack-master- datalake/dw01/artist_hist/artist_ hist_v001/partition_0963D20210 218 in-path- music=s3://audiomack-master- datalake/dw01/music_hist/music_ pist_v001/partition_0963D202 10218 in-path-songs-rank- daily=s3://audiomack-master- datalake/ingest/am_redis_stats/ song_rank_daily/20210218 in- path-songs-rank- weekly=s3://audiomack-master- datalake/ingest/am_redis_stats/ song_rank_weekly/20210218 in- path-songs-rank- monthly=s3://audiomack- master- datalake/ingest/am_redis_stats/ song_rank_monthly/20210218 in- path-songs-rank- monthly=s3://audiomack- master- datalake/ingest/am_redis_stats/ song_rank_monthly/20210218 in-path-songs-rank- in-path-songs-rank-		
				in-path-songs-rank- 		

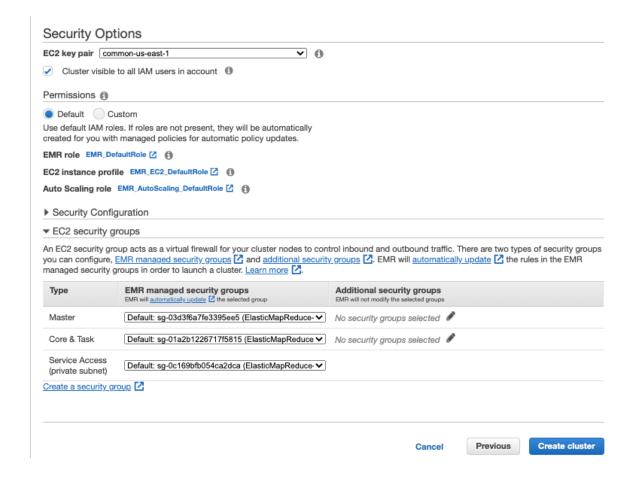
#### Cluster setup - 2 - hardware



### Cluster setup - 3 - general cluster settings



### Cluster setup - 4 - security



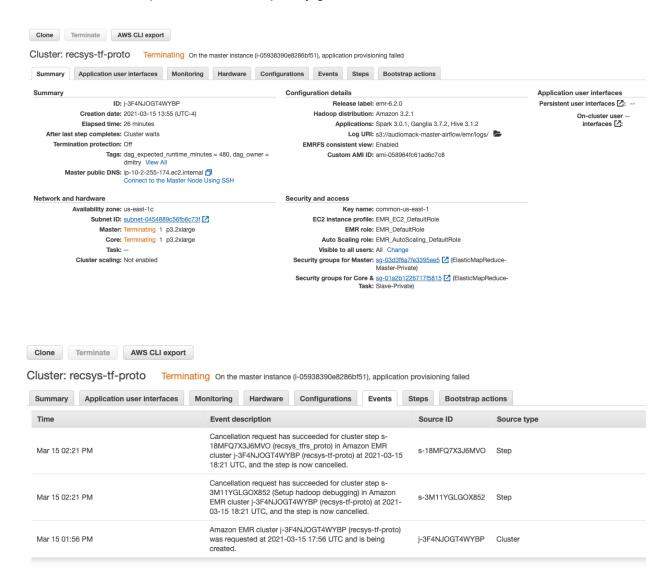
### The bootstrap action

Currently, the bootstrap action we're using simply provides a couple of dependency libraries for the Spark program:

```
#!/bin/bash
pip3 install --user --upgrade pip
pip3 install --user tensorflow_recommenders==v0.4.0
pip3 install tensorflow-io-nightly==0.17.0.dev20210208174016
pip3 install --user boto3
```

#### Cluster initialization error

The cluster bootstraps but then subsequently generates an error and self-terminates:



### Following the logs

The bootstrapping log (stdout)

Bootstrapping seems OK (no errors in stderr).

Bootstrapping Recsys-TF... Requirement already satisfied: pip in /usr/local/lib/python3.7/site-packages (21.0.1) Collecting tensorflow recommenders==v0.4.0 Downloading tensorflow recommenders-0.4.0-py3-none-any.whl (60 kB) Collecting tensorflow==2.4 Downloading tensorflow-2.4.0-cp37-cp37m-manylinux2010 x86 64.whl (394.7 MB) Collecting absl-pv>=0.1.6 Downloading absl py-0.12.0-py3-none-any.whl (129 kB) Requirement already satisfied: wrapt~=1.12.1 in /usr/local/lib64/python3.7/site-packages (from tensorflow==2.4->tensorflow recommenders==v0.4.0) (1.12.1) Collecting h5py~=2.10.0 Downloading h5py-2.10.0-cp37-cp37m-manylinux1\_x86\_64.whl (2.9 MB) Collecting termcolor~=1.1.0 Downloading termcolor-1.1.0.tar.gz (3.9 kB) Requirement already satisfied: six~=1.15.0 in /usr/local/lib/python3.7/site-packages (from tensorflow==2.4->tensorflow recommenders==v0.4.0) (1.15.0) Collecting arpcio~=1.32.0 Downloading grpcio-1.32.0-cp37-cp37m-manylinux2014 x86 64.whl (3.8 MB) Collecting keras-preprocessing~=1.1.2 Downloading Keras Preprocessing-1.1.2-py2.py3-none-any.whl (42 kB) Requirement already satisfied: typing-extensions~=3.7.4 in /usr/local/lib/python3.7/site-packages (from tensorflow==2.4->tensorflow recommenders==v0.4.0) (3.7.4.3) Collecting opt-einsum~=3.3.0 Downloading opt einsum-3.3.0-py3-none-any.whl (65 kB) Requirement already satisfied: numpy~=1.19.2 in /usr/local/lib64/python3.7/site-packages (from tensorflow==2.4->tensorflow recommenders==v0.4.0) (1.19.5) Collecting gast==0.3.3 Downloading gast-0.3.3-py2.py3-none-any.whl (9.7 kB) Collecting tensorboard~=2.4 Downloading tensorboard-2.4.1-py3-none-any.whl (10.6 MB) Collecting tensorflow-estimator<2.5.0,>=2.4.0rc0 Downloading tensorflow estimator-2.4.0-py2.py3-none-any.whl (462 kB) Collecting google-pasta~=0.2 Downloading google pasta-0.2.0-py3-none-any.whl (57 kB) Requirement already satisfied: wheel~=0.35 in /usr/local/lib/python3.7/site-packages (from tensorflow==2.4->tensorflow recommenders==v0.4.0) (0.36.2) Requirement already satisfied: protobuf>=3.9.2 in /usr/local/lib64/python3.7/site-packages (from tensorflow==2.4->tensorflow recommenders==v0.4.0) (3.15.5) Collecting flatbuffers~=1.12.0 Downloading flatbuffers-1.12-py2.py3-none-any.whl (15 kB) Collecting astunparse~=1.6.3 Downloading astunparse-1.6.3-py2.py3-none-any.whl (12 kB) Collecting google-auth-oauthlib<0.5,>=0.4.1 Downloading google auth oauthlib-0.4.3-py2.py3-none-any.whl (18 kB) Collecting google-auth<2,>=1.6.3 Downloading google auth-1.27.1-py2.py3-none-any.whl (136 kB)

Collecting werkzeug>=0.11.15

Downloading Werkzeug-1.0.1-py2.py3-none-any.whl (298 kB)

Collecting tensorboard-plugin-wit>=1.6.0

Downloading tensorboard\_plugin\_wit-1.8.0-py3-none-any.whl (781 kB)

Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.7/site-packages (from tensorboard~=2.4->tensorflow==2.4->tensorflow\_recommenders==v0.4.0) (2.25.1) Collecting markdown>=2.6.8

Downloading Markdown-3.3.4-py3-none-any.whl (97 kB)

Requirement already satisfied: setuptools>=41.0.0 in /usr/local/lib/python3.7/site-packages (from tensorboard~=2.4->tensorflow==2.4->tensorflow\_recommenders==v0.4.0) (54.1.0) Collecting pyasn1-modules>=0.2.1

Downloading pyasn1\_modules-0.2.8-py2.py3-none-any.whl (155 kB)

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.7/site-packages (from google-auth<2,>=1.6.3->tensorboard~=2.4->tensorflow==2.4->tensorflow\_recommenders==v 0.4.0) (4.7.2)

Collecting cachetools<5.0,>=2.0.0

Downloading cachetools-4.2.1-py3-none-any.whl (12 kB)

Collecting requests-oauthlib>=0.7.0

Downloading requests\_oauthlib-1.3.0-py2.py3-none-any.whl (23 kB)

Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/site-packages (from

markdown>=2.6.8->tensorboard~=2.4->tensorflow==2.4->tensorflow\_recommenders==v0.4.0 ) (3.7.0)

Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in /usr/local/lib/python3.7/site-packages (from

pyasn1-modules>=0.2.1->google-auth<2,>=1.6.3->tensorboard~=2.4->tensorflow==2.4->tensorflow\_recommenders==v0.4.0) (0.4.8)

Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/site-packages (from requests<3,>=2.21.0->tensorboard~=2.4->tensorflow==2.4->tensorflow\_recommenders==v0.4.0) (2.10)

Requirement already satisfied: chardet<5,>=3.0.2 in /usr/local/lib/python3.7/site-packages (from

requests<3,>=2.21.0->tensorboard~=2.4->tensorflow==2.4->tensorflow\_recommenders==v0. 4.0) (4.0.0)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/site-packages (from

requests<3,>=2.21.0->tensorboard~=2.4->tensorflow==2.4->tensorflow\_recommenders==v0. 4.0) (2020.12.5)

Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.7/site-packages (from

requests<3,>=2.21.0->tensorboard~=2.4->tensorflow==2.4->tensorflow\_recommenders==v0. 4.0) (1.26.3)

Collecting oauthlib>=3.0.0

Downloading oauthlib-3.1.0-py2.py3-none-any.whl (147 kB)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/site-packages (from importlib-metadata->markdown>=2.6.8->tensorboard~=2.4->tensorflow==2.4->tensorflow\_rec ommenders==v0.4.0) (3.4.1)

Building wheels for collected packages: termcolor Building wheel for termcolor (setup.py): started

Building wheel for termcolor (setup.py): finished with status 'done'

Created wheel for termcolor: filename=termcolor-1.1.0-py3-none-any.whl size=4829 sha256=e10a11156083813785be8da3059bfa57a6e32f672a3a81bdbe81ad1fa71f0e0f Stored in directory:

/home/hadoop/.cache/pip/wheels/3f/e3/ec/8a8336ff196023622fbcb36de0c5a5c218cbb24111d 1d4c7f2

Successfully built termcolor

Installing collected packages: pyasn1-modules, oauthlib, cachetools, requests-oauthlib, google-auth, werkzeug, tensorboard-plugin-wit, markdown, grpcio, google-auth-oauthlib, absl-py, termcolor, tensorflow-estimator, tensorboard, opt-einsum, keras-preprocessing, h5py, google-pasta, gast, flatbuffers, astunparse, tensorflow, tensorflow-recommenders Successfully installed absl-py-0.12.0 astunparse-1.6.3 cachetools-4.2.1 flatbuffers-1.12 gast-0.3.3 google-auth-1.27.1 google-auth-oauthlib-0.4.3 google-pasta-0.2.0 grpcio-1.32.0 h5py-2.10.0 keras-preprocessing-1.1.2 markdown-3.3.4 oauthlib-3.1.0 opt-einsum-3.3.0 pyasn1-modules-0.2.8 requests-oauthlib-1.3.0 tensorboard-2.4.1 tensorboard-plugin-wit-1.8.0 tensorflow-2.4.0 tensorflow-estimator-2.4.0 tensorflow-recommenders-0.4.0 termcolor-1.1.0 werkzeug-1.0.1

Defaulting to user installation because normal site-packages is not writeable

Collecting tensorflow-io-nightly==0.17.0.dev20210208174016

Downloading

tensorflow\_io\_nightly-0.17.0.dev20210208174016-cp37-cp37m-manylinux2010\_x86\_64.whl (25.4 MB)

Requirement already satisfied: tensorflow<2.5.0,>=2.4.0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow-io-nightly==0.17.0.dev20210208174016) (2.4.0)

Requirement already satisfied: protobuf>=3.9.2 in /usr/local/lib64/python3.7/site-packages (from tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (3.15.5)

Requirement already satisfied: absl-py~=0.10 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (0.12.0)

Requirement already satisfied: opt-einsum~=3.3.0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (3.3.0)

Requirement already satisfied: six~=1.15.0 in /usr/local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.15.0)

Requirement already satisfied: numpy~=1.19.2 in /usr/local/lib64/python3.7/site-packages

(from tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.19.5)

Requirement already satisfied: termcolor~=1.1.0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.1.0)

Requirement already satisfied: wrapt~=1.12.1 in /usr/local/lib64/python3.7/site-packages

(from tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.12.1)

Requirement already satisfied: gast==0.3.3 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0.>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (0.3.3)

Requirement already satisfied: keras-preprocessing~=1.1.2 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.1.2)

Requirement already satisfied: h5py~=2.10.0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (2.10.0)

Requirement already satisfied: typing-extensions~=3.7.4 in

/usr/local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (3.7.4.3)

Requirement already satisfied: wheel~=0.35 in /usr/local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (0.36.2)

Requirement already satisfied: grpcio~=1.32.0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.32.0)

Requirement already satisfied: tensorflow-estimator<2.5.0,>=2.4.0rc0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (2.4.0)

Requirement already satisfied: google-pasta~=0.2 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (0.2.0)

Requirement already satisfied: astunparse~=1.6.3 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.6.3)

Requirement already satisfied: tensorboard~=2.4 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (2.4.1)

Requirement already satisfied: flatbuffers~=1.12.0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.12)

Requirement already satisfied: google-auth<2,>=1.6.3 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174 016) (1.27.1)

Requirement already satisfied: google-auth-oauthlib<0.5,>=0.4.1 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174 016) (0.4.3)

Requirement already satisfied: werkzeug>=0.11.15 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174 016) (1.0.1)

Requirement already satisfied: requests<3,>=2.21.0 in /usr/local/lib/python3.7/site-packages (from

tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174 016) (2.25.1)

Requirement already satisfied: markdown>=2.6.8 in

/home/hadoop/.local/lib/python3.7/site-packages (from

tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174 016) (3.3.4)

Requirement already satisfied: tensorboard-plugin-wit>=1.6.0 in

/home/hadoop/.local/lib/pvthon3.7/site-packages (from

tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174 016) (1.8.0)

Requirement already satisfied: setuptools>=41.0.0 in /usr/local/lib/python3.7/site-packages (from

tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174

016) (54.1.0)

Requirement already satisfied: cachetools<5.0,>=2.0.0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

google-auth<2,>=1.6.3->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (4.2.1)

Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.7/site-packages (from google-auth<2,>=1.6.3->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (4.7.2)

Requirement already satisfied: pyasn1-modules>=0.2.1 in

/home/hadoop/.local/lib/python3.7/site-packages (from

google-auth<2,>=1.6.3->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (0.2.8)

Requirement already satisfied: requests-oauthlib>=0.7.0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

google-auth-oauthlib<0.5,>=0.4.1->tensorboard $\sim$ =2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.3.0)

Requirement already satisfied: importlib-metadata in /usr/local/lib/python3.7/site-packages (from

markdown>=2.6.8->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.1 7.0.dev20210208174016) (3.7.0)

Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in /usr/local/lib/python3.7/site-packages (from

pyasn1-modules>=0.2.1->google-auth<2,>=1.6.3->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (0.4.8)

Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.7/site-packages (from

requests<3,>=2.21.0->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (1.26.3)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/site-packages (from

requests<3,>=2.21.0->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (2020.12.5)

Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/site-packages (from requests<3,>=2.21.0->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (2.10)

Requirement already satisfied: chardet<5,>=3.0.2 in /usr/local/lib/python3.7/site-packages (from

requests<3,>=2.21.0->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (4.0.0)

Requirement already satisfied: oauthlib>=3.0.0 in

/home/hadoop/.local/lib/python3.7/site-packages (from

requests-oauthlib>=0.7.0->google-auth-oauthlib<0.5,>=0.4.1->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (3.1.0)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/site-packages (from importlib-metadata->markdown>=2.6.8->tensorboard~=2.4->tensorflow<2.5.0,>=2.4.0->tensorflow-io-nightly==0.17.0.dev20210208174016) (3.4.1)

Installing collected packages: tensorflow-io-nightly

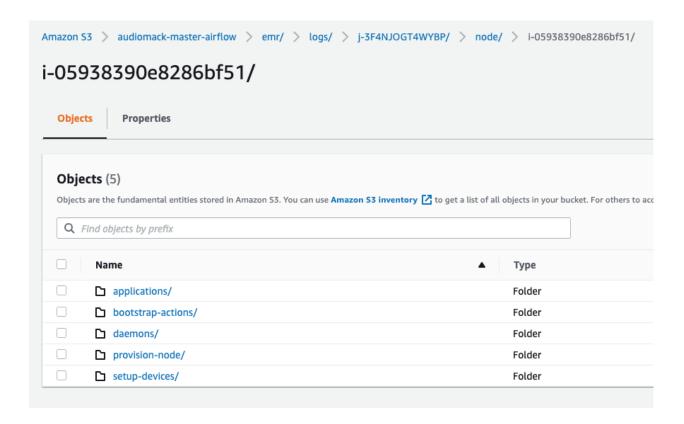
Successfully installed tensorflow-io-nightly-0.17.0.dev20210208174016

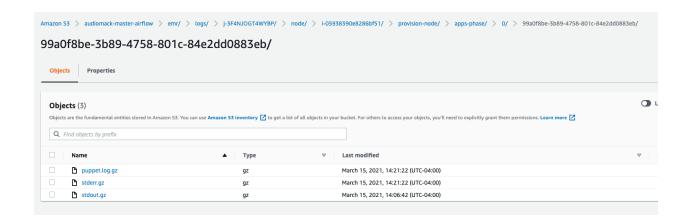
Requirement already satisfied: boto3 in /usr/local/lib/python3.7/site-packages (1.17.21)

Requirement already satisfied: jmespath<1.0.0,>=0.7.1 in

/usr/local/lib/python3.7/site-packages (from boto3) (0.10.0)
Requirement already satisfied: s3transfer<0.4.0,>=0.3.0 in
/usr/local/lib/python3.7/site-packages (from boto3) (0.3.4)
Requirement already satisfied: botocore<1.21.0,>=1.20.21 in
/usr/local/lib/python3.7/site-packages (from boto3) (1.20.21)
Requirement already satisfied: urllib3<1.27,>=1.25.4 in /usr/local/lib/python3.7/site-packages
(from botocore<1.21.0,>=1.20.21->boto3) (1.26.3)
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in
/usr/local/lib/python3.7/site-packages (from botocore<1.21.0,>=1.20.21->boto3) (2.8.1)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/site-packages (from python-dateutil<3.0.0,>=2.1->botocore<1.21.0,>=1.20.21->boto3) (1.15.0)
Bootstrapping of Recsys-TF: done.

#### The provision-node apps-phase error messages





In that node log directory, the stderr file contains the below error.

2021-03-15 18:20:58,205 ERROR main: Encountered a problem while provisioning com.amazonaws.emr.node.provisioner.puppet.api.PuppetException: Unable to complete transaction and some changes were applied. at com.amazonaws.emr.node.provisioner.puppet.api.ApplyCommand.handleExitcode(ApplyCom mand.java:74) at com.amazonaws.emr.node.provisioner.puppet.api.ApplyCommand.call(ApplyCommand.java:5 6) com.amazonaws.emr.node.provisioner.bigtop.BigtopPuppeteer.applyPuppet(BigtopPuppeteer. java:74) com.amazonaws.emr.node.provisioner.bigtop.BigtopDeployer.deploy(BigtopDeployer.java:22) com.amazonaws.emr.node.provisioner.NodeProvisioner.provision(NodeProvisioner.java:25) com.amazonaws.emr.node.provisioner.workflow.NodeProvisionerWorkflow.doWork(NodeProvi sionerWorkflow.java:230) com. a mazon a ws. emr. node. provisioner. work flow. Node Provisioner Work flow. work (Node Provisioner) and the provisioner work flow. Node Provisioner work flow. NodnerWorkflow.java:113) at com.amazonaws.emr.node.provisioner.Program.main(Program.java:31)

The puppet.log file contains the following error message(s):

```
2021-03-15 18:05:24 +0000

/Stage[main]/Nvidia::Common/Package[kernel-devel-4.14.219-161.340.amzn2.x86_64] (info):

Starting to evaluate the resource (470 of 818)

2021-03-15 18:05:24 +0000

/Stage[main]/Nvidia::Common/Package[kernel-devel-4.14.219-161.340.amzn2.x86_64] (info):

Evaluated in 0.07 seconds
```

```
2021-03-15 18:05:24 +0000 /Stage[main]/Nvidia::Common/Package[gcc] (info): Starting to
evaluate the resource (471 of 818)
2021-03-15 18:05:24 +0000 /Stage[main]/Nvidia::Common/Package[gcc] (info): Evaluated in
0.00 seconds
2021-03-15 18:05:24 +0000 /Stage[main]/Nvidia::Common/Package[nvidia-cuda] (info):
Starting to evaluate the resource (472 of 818)
2021-03-15 18:07:51 +0000 /Stage[main]/Nvidia::Common/Package[nvidia-cuda]/ensure
(notice): created
2021-03-15 18:07:51 +0000 /Stage[main]/Nvidia::Common/Package[nvidia-cuda] (info):
Evaluated in 147.45 seconds
2021-03-15 18:07:51 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]
(info): Starting to evaluate the resource (473 of 818)
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): ++ uname -r
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): +
KERNEL VERSION=4.14.219-161.340.amzn2.x86 64
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): +
KERNEL SOURCE PATH=/usr/src/kernels/4.14.219-161.340.amzn2.x86 64/
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): +
NVIDIA DIR=/mnt/nvidia
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): +
INSTALL DIR=/mnt/nvidia
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): +
CUDA_VERSION=10.1
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): +
CUDA BINARY='cuda 10.1* linux.run'
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): +
CUDNN TGZ='cudnn-10.1-linux-x64-v*.tgz'
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): +
NVIDIA BINARY='NVIDIA-Linux-x86 64-*.run'
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): + export
CC=qcc
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): + CC=gcc
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): +
NVIDIA OPTS='-s --kernel-source-path=/usr/src/kernels/4.14.219-161.340.amzn2.x86 64/
--no-opengl-files'
2021-03-15 18:08:01 +0000
/Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): + chmod +x
/mnt/nvidia/NVIDIA-Linux-x86 64-418.116.00.run
```

2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): + /mnt/nvidia/NVIDIA-Linux-x86 64-418.116.00.run -s --kernel-source-path=/usr/src/kernels/4.14.219-161.340.amzn2.x86 64/ --no-opengl-files 2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): Verifying archive integrity... OK 2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): Uncompressing NVIDIA Accelerated Graphics Driver for Linux-x86 64 418.116.00..... 2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): 2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): ERROR: An NVIDIA kernel module 'nvidia' appears to already be loaded in your kernel. This may be because it is in use (for example, by an X server, a CUDA program, or the NVIDIA Persistence Daemon), but this may also happen if your kernel was configured without support for module unloading. Please be sure to exit any programs that may be using the GPU(s) before attempting to upgrade your driver. If no GPU-based programs are running, you know that your kernel supports module unloading, and you still receive this message, then an error may have occured that has corrupted an NVIDIA kernel module's usage count, for which the simplest remedy is to reboot your computer. 2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): 2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): 2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (notice): ERROR: Installation has failed. Please see the file '/var/log/nvidia-installer.log' for details. You may find suggestions on fixing installation problems in the README available on the Linux driver download page at www.nvidia.com. 2021-03-15 18:08:01 +0000 Puppet (err): '/mnt/nvidia/install-nvidia' returned 1 instead of one 2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]/returns (err): change from 'notrun' to ['0'] failed: '/mnt/nvidia/install-nvidia' returned 1 instead of one of [0]

2021-03-15 18:08:01 +0000 /Stage[main]/Nvidia::Common/Exec[/mnt/nvidia/install-nvidia]

(info): Evaluated in 10.18 seconds

# Things we tried

We have tried a few things from the bootstrap action:

sudo yum remove nvidia-\* -y

also

sudo yum remove nvidia-\* -y

sudo yum install -y nvidia-container-runtime nvidia-container-toolkit nvidia-docker2 nvidia-fabricmanager-450

to no avail.

### References

Below are some of the reference documentation we've been using:

- 1. TensorFlow GPU support doc page: https://www.tensorflow.org/install/gpu
- 2. AWS TensorFlow tutorial page:
  - https://docs.aws.amazon.com/dlami/latest/devguide/tutorial-tensorflow.html
- 3. Blog article on choosing the right GPU for DL on AWS: https://towardsdatascience.com/choosing-the-right-gpu-for-deep-learning-on-aws-d69c1 57d8c86
- 4. **Our post in StackOverflow about the issue we're seeing**: https://stackoverflow.com/questions/66446887/how-to-install-cudatoolkit-on-aws-emr-to-enable-distributed-training-in-tensorfl