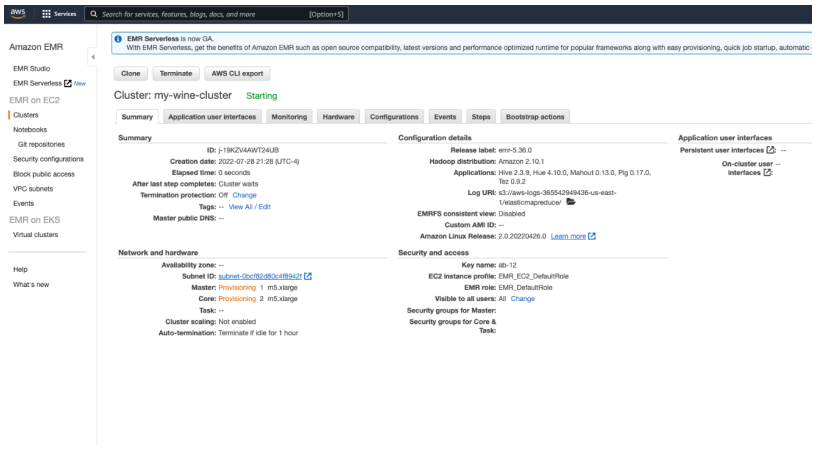
Wine Quality program deployment process

Name: Keerthi Kondisetty
UCID: rkk3

Github with dockerfile - https://github.com/keerthikondisetty/cs643-wineTaster

Deployment and preparation of the instances

1. Launch an EMR cluster to run spark on the AWS Console.

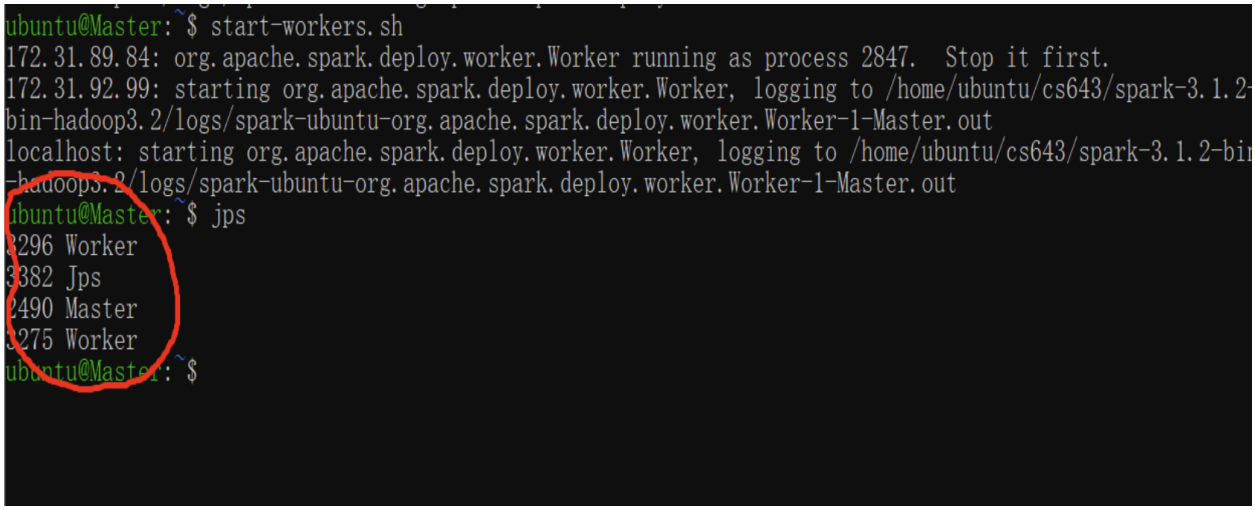


2. Using the command below, Copy the whole directory onto the master node.

   scp -i ~/.ssh/id_rsa -r ${pwd}/
   hadoop@<public_dns_master_node>:/home/master/wineTaster

3. Start the worker machines using the below command and use the next command to check if nodes are online.

   1. Start-workers.sh
   2. jps



4. Start the training by running the file using “python training.py <Dataset>”.

   Before running the above command, make sure you are in the same directory as you copied the files in step 2.

Running the actual prediction using the training model created above.

1. We will pull the docker image from the docker hub container registry

   docker pull rk33/cs643-wineTaster

2. Run the docker image using the below command

   docker container run –rm -it rk33/cs643-wineTaster

3. The result should be something similar to the below image.

```
 02:54:40 INFO MapOutputTrackerMasterEndpoint: Asked to send map output locations for shuffle 3 to 172.31.4

 02:54:40 INFO TaskSetManager: Finished task 0.0 in stage 126.0 (TID 125) in 18 ms on ip-172-31-49-119.ec2.
ecutor 1) (1/1)
 02:54:40 INFO YarnScheduler: Removed TaskSet 126.0, whose tasks have all completed, from pool
 02:54:40 INFO DAGScheduler: ResultStage 126 (collectAsMap at MulticlassMetrics.scala:53) finished in 0.024
 02:54:40 INFO DAGScheduler: Job 122 finished: collectAsMap at MulticlassMetrics.scala:53, took 0.114820 s
e: 0.5904050519731796
 02:54:40 INFO SparkContext: Invoking stop() from shutdown hook
 02:54:40 INFO SparkUI: Stopped Spark web UI at http://ip-172-31-60-95.ec2.internal:4040
 02:54:40 INFO YarnClientSchedulerBackend: Interrupting monitor thread
 02:54:40 INFO YarnClientSchedulerBackend: Shutting down all executors
 02:54:40 INFO YarnSchedulerBackend$YarnDriverEndpoint: Asking each executor to shut down
 02:54:40 INFO SchedulerExtensionServices: Stopping SchedulerExtensionServices
Option=None,
s=List(),
=false)
 02:54:40 INFO YarnClientSchedulerBackend: Stopped
 02:54:40 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoint stopped!
 02:54:40 INFO MemoryStore: MemoryStore cleared
 02:54:40 INFO BlockManager: BlockManager stopped
 02:54:40 INFO BlockManagerMaster: BlockManagerMaster stopped
 02:54:40 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: OutputCommitCoordinator stopped!
 02:54:40 INFO SparkContext: Successfully stopped SparkContext
 02:54:40 INFO ShutdownHookManager: Shutdown hook called
 02:54:40 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-38a9db92-1163-49b3-88bb-e37110fbe8c3
 02:54:40 INFO ShutdownHookManager: Deleting directory /mnt/tmp/spark-38a9db92-1163-49b3-88bb-e37110fbe8c3/
```