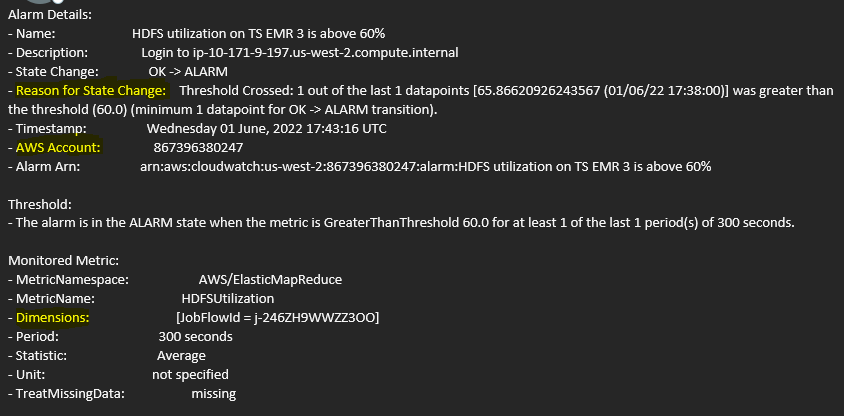
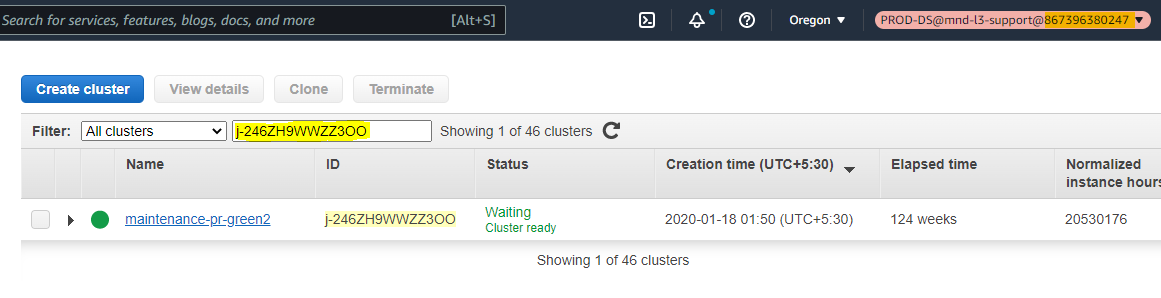
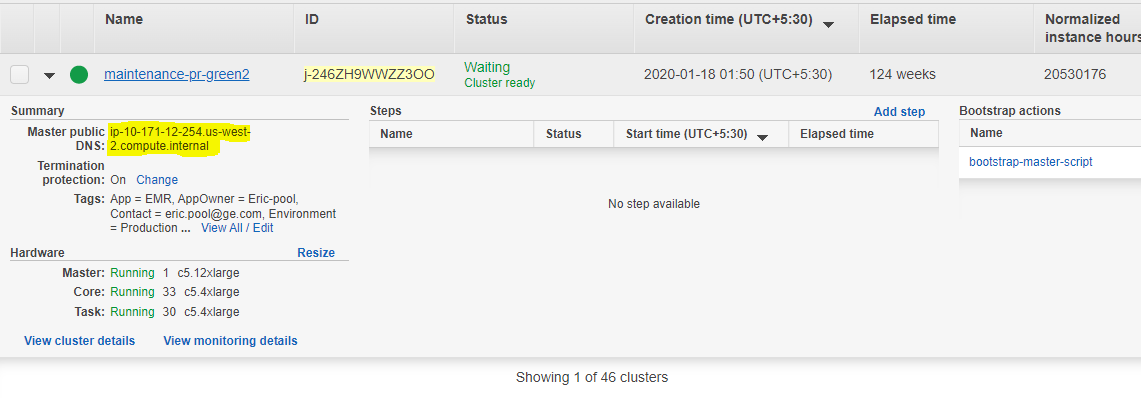
EXT: ALARM: "HDFS utilization on TS EMR 3 is above 60%" in US West (Oregon)

1. We must check the threshold value of the alert and check which EMR is affecting in which account.



1. We must find the EMR from the Console and select the IP connect to that EMR.



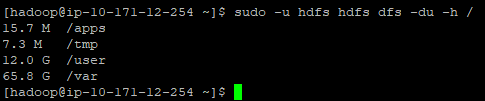


3. Connect EMR using the below command.

ssh hadoop@ip-10-171-12-254

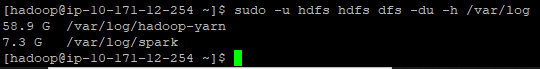
4. Check the Disk Used in the EMR using below command

sudo -u hdfs hdfs dfs -du -h /



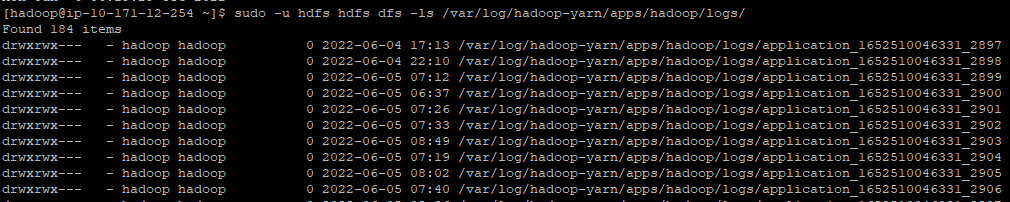
5. Logs are writing in /var mount point, so check the disk used for /var/log using below command.

sudo -u hdfs hdfs dfs -du -h /var/log

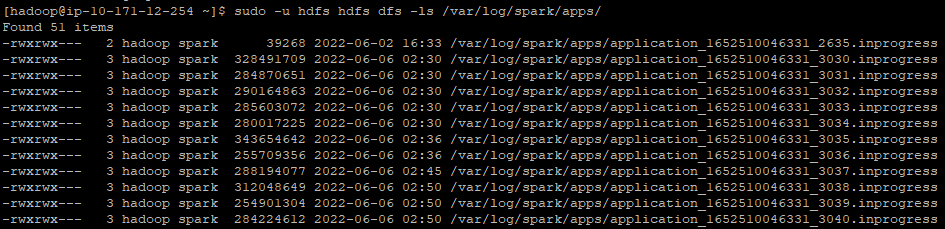


6. Check the log files under Hadoop-yarn application and spark application using below commands.

sudo -u hdfs hdfs dfs -ls /var/log/hadoop-yarn/apps/hadoop/logs/



sudo -u hdfs hdfs dfs -ls /var/log/spark/apps/



7. Take the approvals from Authorized Approvers before deleting the logs.

8. By using below commands we can clean the logs.

A. For Hadoop-Yarn

sudo -u hdfs hdfs dfs -rm -R /var/log/hadoop-yarn/apps/hadoop/logs/ -skipTrash

B. For Spark App

sudo -u hdfs hdfs dfs -rm -R /var/log/spark/apps/ -skipTrash

NOTE: Make sure all the compaction jobs (Raw & Derived) should complete before clearing the logs.