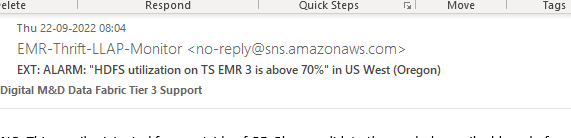
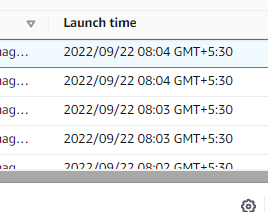
Hdfs alert initial findings

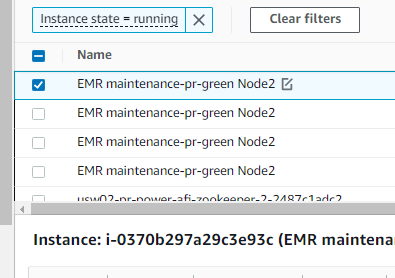
We get alert 3 times

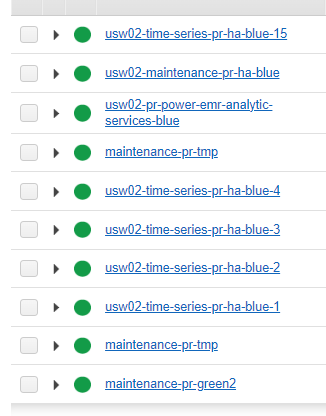
3:34, 8:04 , 20:39

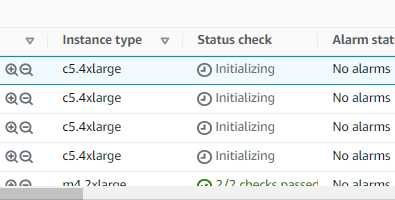
Need to get info what EMR gets job at what time

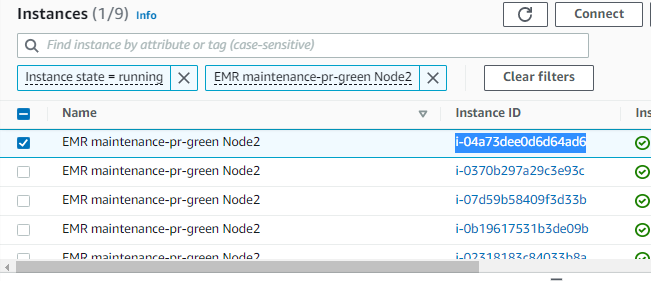


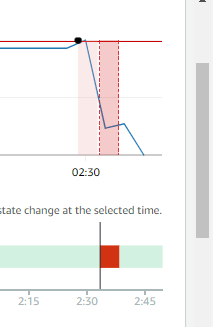


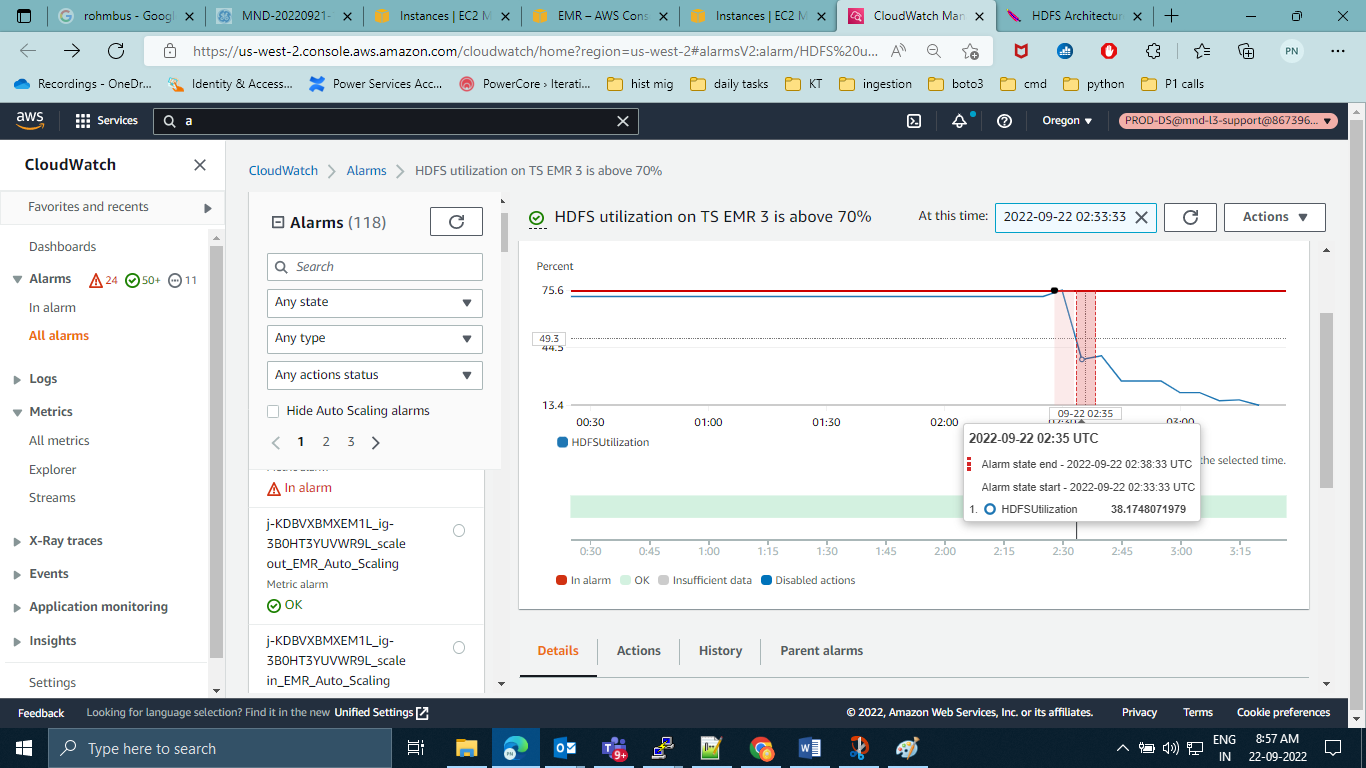












Hadoop classpath

cd /etc/hadoop/conf

minor runs in usw02-pri-power-afi-airflow-worker-analytics-77a488a888 instance

major runs in compaction emr :-

[hadoop@ip-10-171-12-254 ~]$ crontab -l

\*/30 \* \* \* \* /home/hadoop/cleanup.sh

0 16 \* \* \* nohup sh /home/hadoop/Khine/auroracompwritesqs2s3kpi.sh &

0 23 \* \* \* nohup sh /home/hadoop/Khine/auroracompwritesqs2s3kpi.sh &

0 2 \* \* \* nohup sh /home/hadoop/Khine/auroracompwritesqs2s3major.sh &

cat /home/hadoop/Khine/auroracompwritesqs2s3kpi.sh

#!/usr/bin/env bash

LOGFILE=/mnt/logs/aurora\_kpisqs\_`date +%Y%m%d\_%H%M%S`.log

START\_DATE=`date +%Y%m%d\_%H%M%S`

echo "Job start time is $START\_DATE" >> $LOGFILE

spark-submit --master yarn --deploy-mode client \

--conf spark.dynamicAllocation.enabled=false \

--driver-memory 7G \

--num-executors 4 \

--executor-memory 5G \

--executor-cores 5 \

--files /home/hadoop/Khine/log4j.properties \

--conf "spark.executor.extraJavaOptions=-Dlog4j.configuration=file:log4j-prod.properties" \

--conf "spark.driver.extraJavaOptions=-Dlog4j.configuration=file:log4j-prod.properties" \

--conf spark.serializer=org.apache.spark.serializer.KryoSerializer \

--conf spark.hadoop.fs.s3.consistent=false \

--conf spark.shuffle.blockTransferService=nio \

--conf spark.port.maxRetries=200 /home/hadoop/Khine/writesqs2s3kpi.py >> $LOGFILE &

[hadoop@ip-10-171-12-254 logs]$ cat /home/hadoop/Khine/auroracompwritesqs2s3major.sh

#!/usr/bin/env bash

LOGFILE=/mnt/logs/aurora\_rawsqs\_`date +%Y%m%d\_%H%M%S`.log

START\_DATE=`date +%Y%m%d\_%H%M%S`

echo "Job start time is $START\_DATE" >> $LOGFILE

spark-submit --master yarn --deploy-mode client \

--conf spark.dynamicAllocation.enabled=false \

--driver-memory 7G \

--num-executors 4 \

--executor-memory 5G \

--executor-cores 5 \

--files /home/hadoop/Khine/log4j.properties \

--conf "spark.executor.extraJavaOptions=-Dlog4j.configuration=file:log4j-prod.properties" \

--conf "spark.driver.extraJavaOptions=-Dlog4j.configuration=file:log4j-prod.properties" \

--conf spark.serializer=org.apache.spark.serializer.KryoSerializer \

--conf spark.hadoop.fs.s3.consistent=false \

--conf spark.shuffle.blockTransferService=nio \

--conf spark.port.maxRetries=200 /home/hadoop/Khine/writesqs2s3major.py >> $LOGFILE &

Printenv