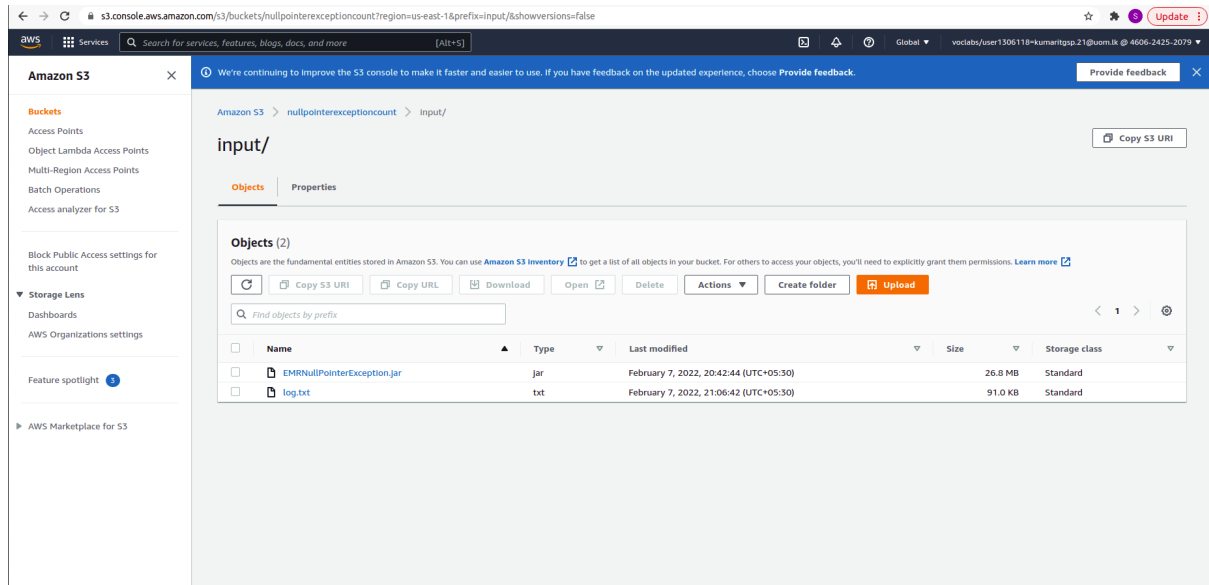


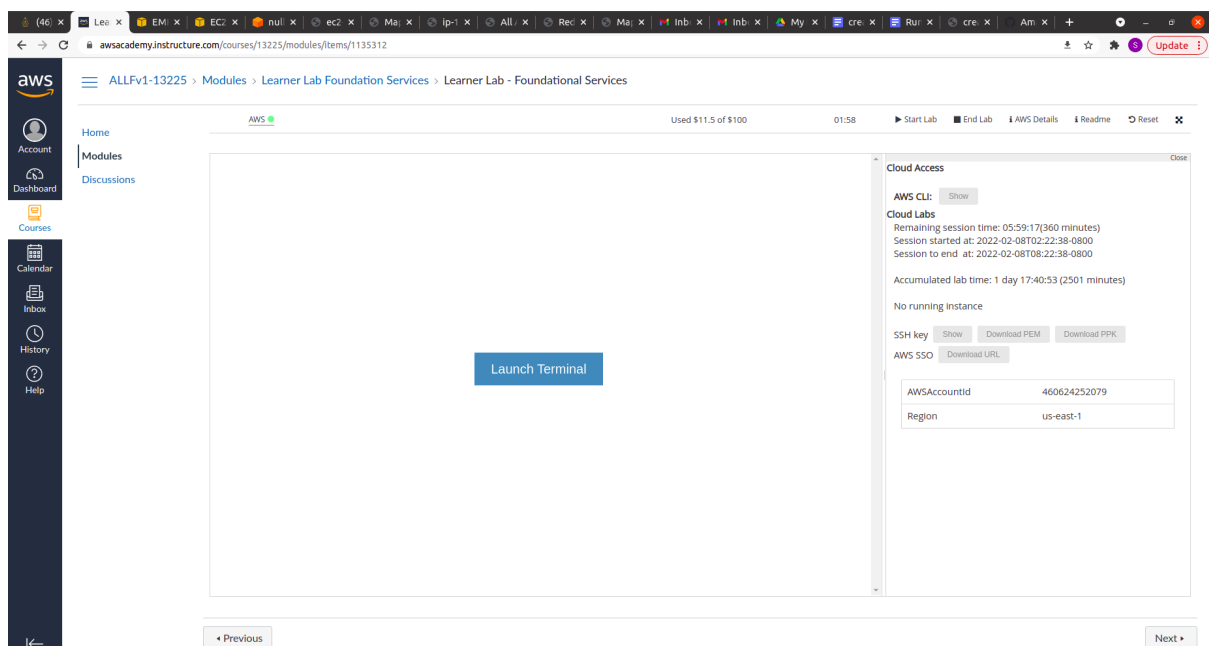
Save executable jar and log file AWS s3 bucket

Step 01 : goto AWS S3 bucket and create a folder . create a input folder and put jar and log file

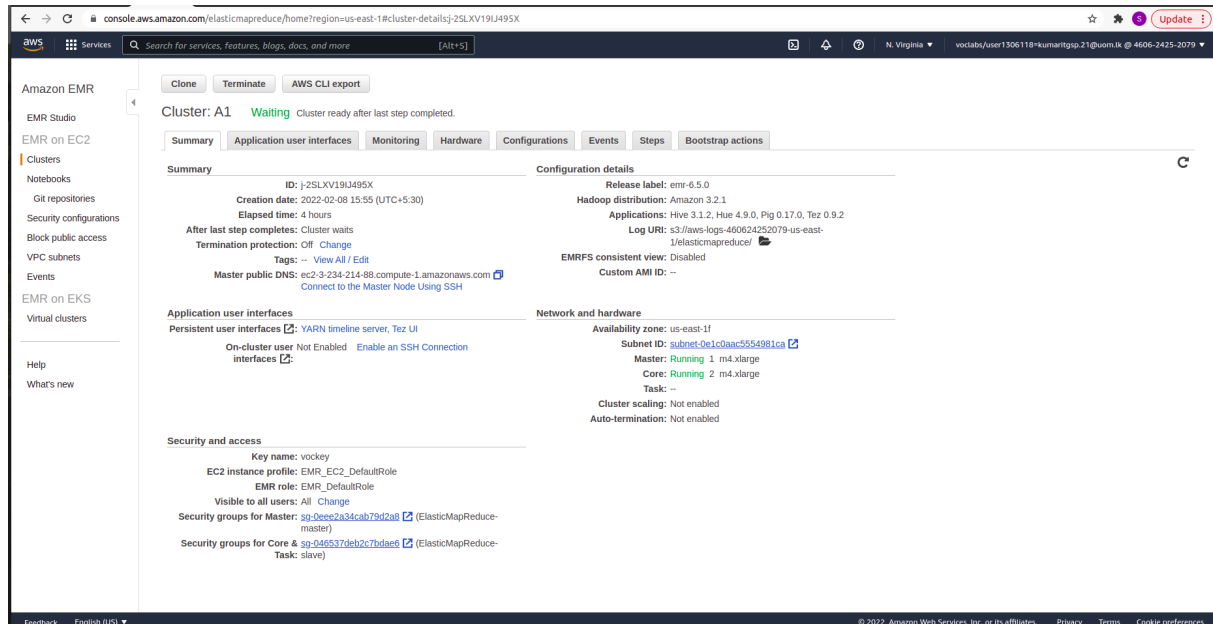


Run application on AWS EMR

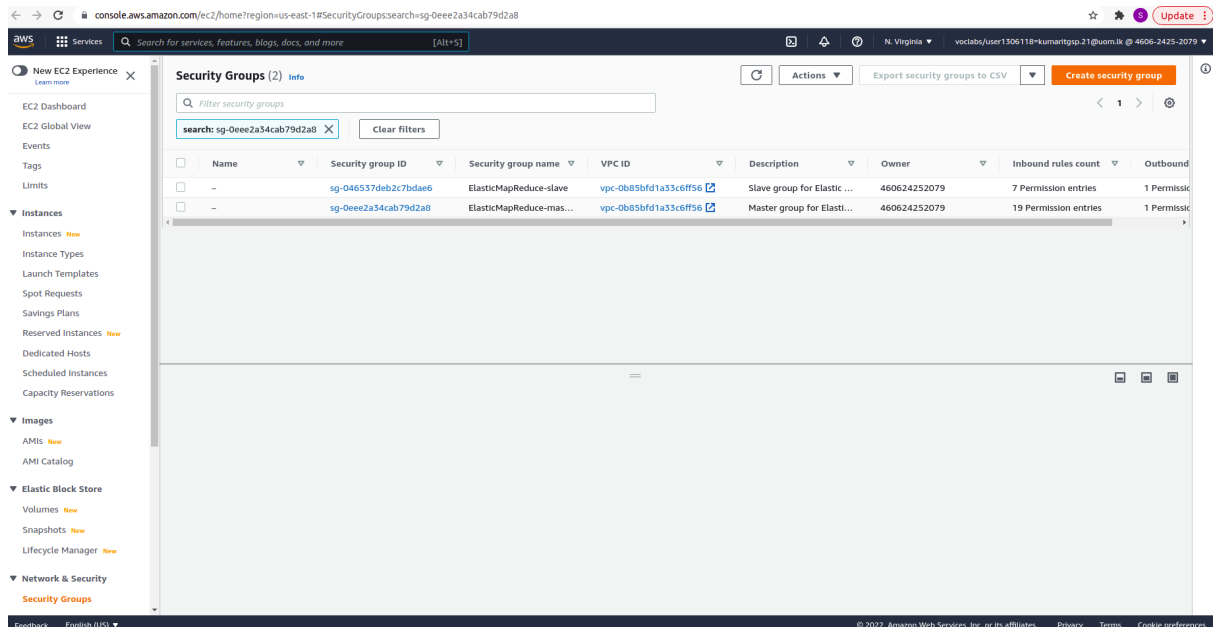
Step 01 : start learner lab



Step 02 : click AWS and create cluster



Step 03 : add security group for both master and slave which is in “Security and access” section



Step 04 : add security group both master and slave as follows



Step 05 : connect to the master node and enable SSH connection

```
hell@dell-Inspiron-5593:~/softwares/big_data$ ssh -i labsuser.pem hadoop@ec2-3-234-214-88.compute-1.amazonaws.com
Last login: Tue Feb  8 11:33:04 2022

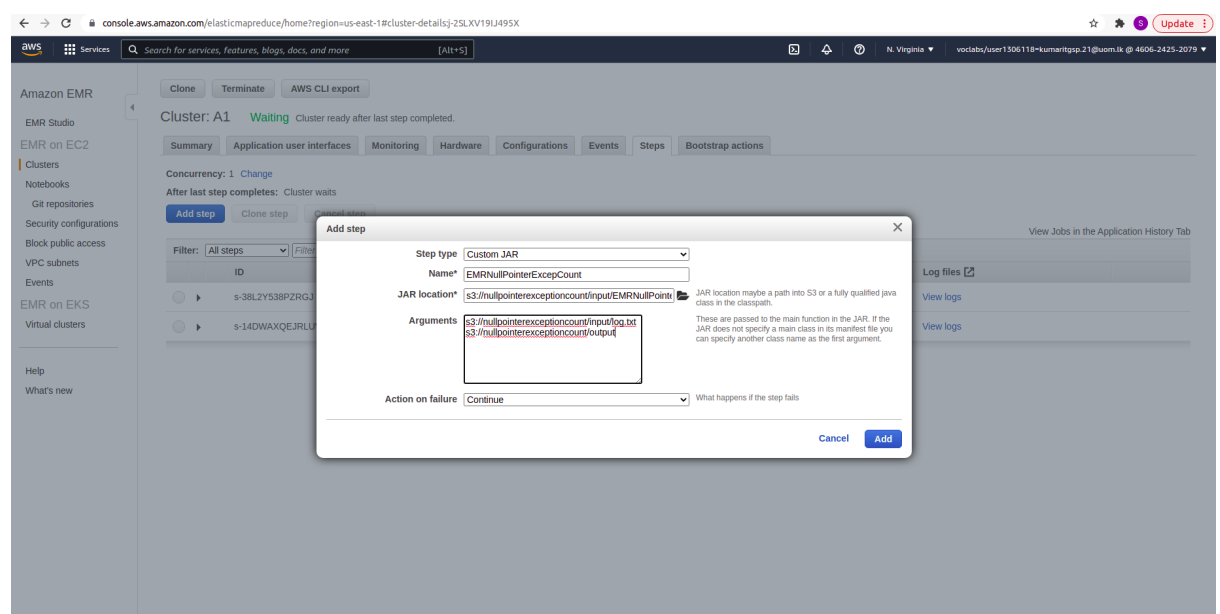
  _|_  _|_ )
 _|_ ( _|_ /   Amazon Linux 2 AMI
--|\\_||_||

https://aws.amazon.com/amazon-linux-2/
19 package(s) needed for security, out of 24 available
Run "sudo yum update" to apply all updates.

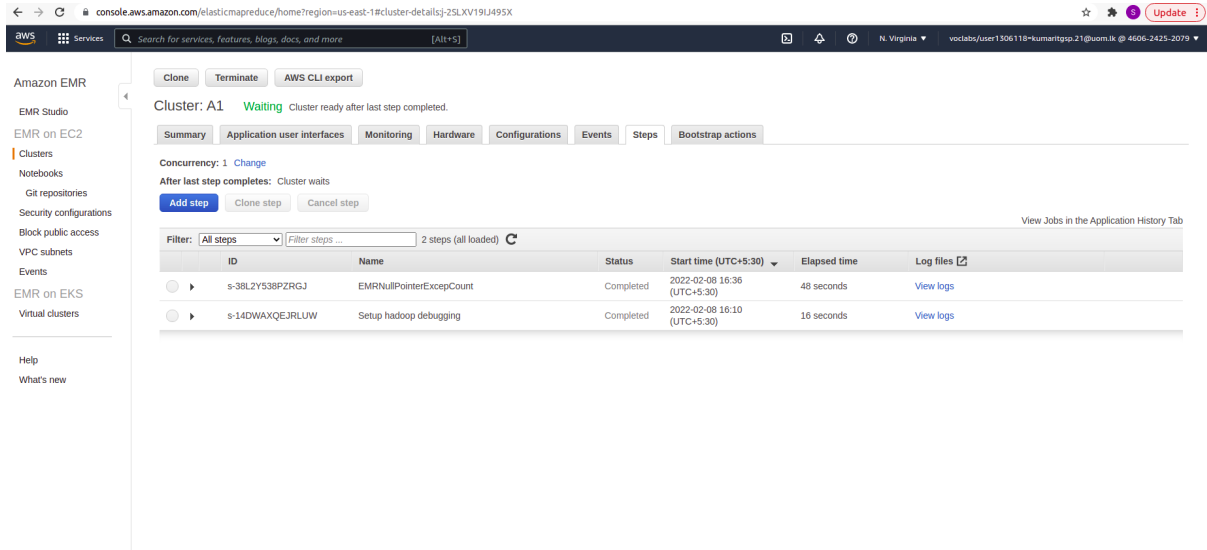
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRRRRRRRRRR
E::::::::::::::::::::E M::::::::M M::::::::M R::::::::R
EE::::::::EEEEEEEE::E M::::::::M M::::::::M R::::RRRRRR::::R
 E::::E EEEEE M::::::::M M::::::::M RR::::R R::::R
 E::::E M::::::::M M::M::::M R:::R R::::R
 E::::EEEEEEEEEE M::::M M::M M::::M R::RRRRRR::::R
 E::::EEEEEEEEEE M::::M M::M::::M M::::M R::::::::::RR
 E::::EEEEEEEEEE M::::M M::M M::::M R::RRRRRR::::R
 E::::E M::::M M::M M::::M R:::R R::::R
 E::::E EEEEE M::::M MMM M::::M R:::R R::::R
EE::::EEEEEEEE::E M::::M M::::M R:::R R::::R
E::::::::::::::::::::E M::::M M::::M RR::::R R::::R
EEEEEEEEEEEEEEEEEEEE MMMMMMMM MMMMMMMM RRRRRRRR RRRRRR
```

```
hell@dell-Inspiron-5593:~/softwares/big_data$ ssh -i labsuser.pem -ND 8157 hadoop@ec2-3-234-214-88.compute-1.amazonaws.com
```

Step 06 : go to cluster and add step as follows



Step 07: after sometime later you can see your map reduce task complete or not



Cluster: A1 **Waiting** Cluster ready after last step completed.

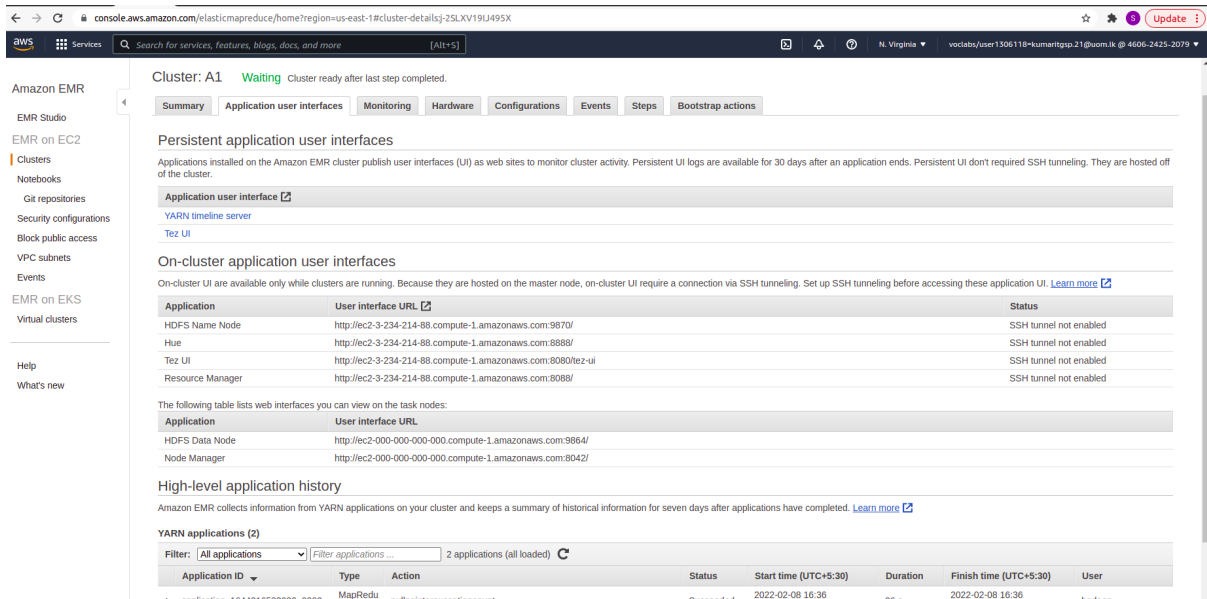
Summary Application user interfaces Monitoring Hardware Configurations Events Steps Bootstrap actions

Concurrency: 1 [Change](#)
After last step completes: Cluster waits
[Add step](#) [Clone step](#) [Cancel step](#)

Filter: All steps Filter steps ... 2 steps (all loaded)

ID	Name	Status	Start time (UTC+5:30)	Elapsed time	Log files
s-38LZY538P2RGJ	EMRNullPointerExcepCount	Completed	2022-02-08 16:36 (UTC+5:30)	48 seconds	View logs
s-14DWAXQEJRLUW	Setup hadoop debugging	Completed	2022-02-08 16:10 (UTC+5:30)	16 seconds	View logs

Step 09 : goto the cluster and click the “Application and user interfaces” tab.



Cluster: A1 **Waiting** Cluster ready after last step completed.

Summary Application user interfaces Monitoring Hardware Configurations Events Steps Bootstrap actions

Persistent application user interfaces

Applications installed on the Amazon EMR cluster publish user interfaces (UI) as web sites to monitor cluster activity. Persistent UI logs are available for 30 days after an application ends. Persistent UI don't required SSH tunneling. They are hosted off of the cluster.

[Application user interface](#)
[YARN timeline server](#)
[Tez UI](#)

On-cluster application user interfaces

On-cluster UI are available only while clusters are running. Because they are hosted on the master node, on-cluster UI require a connection via SSH tunneling. Set up SSH tunneling before accessing these application UI. [Learn more](#)

Application	User interface URL	Status
HDFS Name Node	http://ec2-3-234-214-88.compute-1.amazonaws.com:9870/	SSH tunnel not enabled
Hue	http://ec2-3-234-214-88.compute-1.amazonaws.com:8888/	SSH tunnel not enabled
Tez UI	http://ec2-3-234-214-88.compute-1.amazonaws.com:8080/tez-ui/	SSH tunnel not enabled
Resource Manager	http://ec2-3-234-214-88.compute-1.amazonaws.com:8080/	SSH tunnel not enabled

The following table lists web interfaces you can view on the task nodes:

Application	User interface URL
HDFS Data Node	http://ec2-000-000-000-000.compute-1.amazonaws.com:9864/
Node Manager	http://ec2-000-000-000-000.compute-1.amazonaws.com:8042/

High-level application history

Amazon EMR collects information from YARN applications on your cluster and keeps a summary of historical information for seven days after applications have completed. [Learn more](#)

YARN applications (2)

Filter: All applications Filter applications ... 2 applications (all loaded)

Application ID	Type	Action	Status	Start time (UTC+5:30)	Duration	Finish time (UTC+5:30)	User
application_1644316532039_0002	MapRedu	nullpointerexceptioncount	Succeeded	2022-02-08 16:36	36 s	2022-02-08 16:36	hadoop

Step 10 : copy “HDFS Name Node” and paste in chrome.

Note: you should install FroxyProxy chrome extension and setup the environment

Overview

Started:	Tue Feb 08 16:05:02 +0530 2022
Version:	3.2.1-amzn-5, (Unknown)
Compiled:	Wed Nov 10 14:16:00 +0530 2021 by release from Unknown
Cluster ID:	CID-276ee636-5101-4efd-b564-44732f69ba8d
Block Pool ID:	BP-1177051720-172.31.69.101-1644316449747

Summary

Security is off.	
SafeMode is off.	
1,241 files and directories, 1,172 blocks (1,172 replicated blocks, 0 erasure coded block groups) = 2,413 total filesystem object(s).	
Heap Memory used 62.63 MB of 453 MB Heap Memory. Max Heap Memory is 1.6 GB.	
Non Heap Memory used 78.9 MB of 80.81 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.	
Configured Capacity:	115.94 GB
Configured Remote Capacity:	0 B
DFS Used:	1.02 GB (0.88%)
Non DFS Used:	858.17 MB
DFS Remaining:	114.08 GB (99.4%)
Block Pool Used:	1.02 GB (0.88%)
Datanodes: 16active% (Min/Median/Max)DFSHea	0.84% / 0.92% / 0.94%

Step 11: you can see logs by clicking utilities -> logs -> log file


Directory: /logs/

Name ↑	Last Modified	Size
hadoop-hdfs-namenode-ip-172-31-69-101.log	Feb 8, 2022 2:45:59 PM	7,402 bytes
hadoop-hdfs-namenode-ip-172-31-69-101.log.2022-02-08-10.gz	Feb 8, 2022 11:04:25 AM	55,339 bytes
hadoop-hdfs-namenode-ip-172-31-69-101.log.2022-02-08-11	Feb 8, 2022 11:55:01 AM	44,325 bytes
hadoop-hdfs-namenode-ip-172-31-69-101.log.2022-02-08-12	Feb 8, 2022 12:55:01 PM	8,604 bytes
hadoop-hdfs-namenode-ip-172-31-69-101.log.2022-02-08-13	Feb 8, 2022 1:55:01 PM	8,604 bytes
hadoop-hdfs-namenode-ip-172-31-69-101.out	Feb 8, 2022 11:03:02 AM	6,670 bytes
nn.format.log	Feb 8, 2022 10:34:59 AM	30,786 bytes
SecurityAuth-hdfs.audit	Feb 8, 2022 10:34:58 AM	0 bytes

```
2022-02-08 14:00:01,782 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:00:01,786 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:00:01,793 WARN com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
2022-02-08 14:00:01,793 INFO com.amazon.ws.enr.hadoop.metrics2.sink.cloudwatch.CloudWatchSink (cloudwatch): Metrics pushed
2022-02-08 14:05:01,780 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:05:01,784 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:05:01,791 WARN com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
2022-02-08 14:05:01,791 INFO com.amazon.ws.enr.hadoop.metrics2.sink.cloudwatch.CloudWatchSink (cloudwatch): Metrics pushed
2022-02-08 14:05:01,792 INFO org.apache.hadoop.util.JvmPauseMonitor (org.apache.hadoop.util.JvmPauseMonitor$Monitor@2f67a4d3): Detected pause in JVM or host machine (eg GC): pause of approximately 3228ms
No GCs detected
2022-02-08 14:10:01,781 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:10:01,785 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:10:01,791 WARN com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
2022-02-08 14:10:01,791 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:10:01,784 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:15:01,791 WARN com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
2022-02-08 14:15:01,791 INFO com.amazon.ws.enr.hadoop.metrics2.sink.cloudwatch.CloudWatchSink (cloudwatch): Metrics pushed
2022-02-08 14:15:01,792 INFO org.apache.hadoop.util.JvmPauseMonitor (org.apache.hadoop.util.JvmPauseMonitor$Monitor@2f67a4d3): Detected pause in JVM or host machine (eg GC): pause of approximately 3599ms
No GCs detected
2022-02-08 14:20:01,784 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:20:01,797 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:20:01,797 WARN com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
2022-02-08 14:20:01,797 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:25:01,783 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:25:01,793 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
2022-02-08 14:25:01,793 INFO com.amazon.ws.enr.hadoop.metrics2.sink.cloudwatch.CloudWatchSink (cloudwatch): Metrics pushed
2022-02-08 14:30:01,781 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:30:01,785 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:30:01,792 WARN com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
2022-02-08 14:30:01,792 INFO com.amazon.ws.enr.hadoop.metrics2.sink.cloudwatch.CloudWatchSink (cloudwatch): Metrics pushed
2022-02-08 14:32:27,802 INFO org.apache.hadoop.util.JvmPauseMonitor (org.apache.hadoop.util.JvmPauseMonitor$Monitor@2f67a4d3): Detected pause in JVM or host machine (eg GC): pause of approximately 3334ms
No GCs detected
2022-02-08 14:35:01,781 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:35:01,785 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:35:01,791 WARN com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
2022-02-08 14:35:01,791 INFO com.amazon.ws.enr.hadoop.metrics2.sink.cloudwatch.CloudWatchSink (cloudwatch): Metrics pushed
2022-02-08 14:40:01,781 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:40:01,785 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:40:01,792 WARN com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
2022-02-08 14:40:01,792 INFO com.amazon.ws.enr.hadoop.metrics2.sink.cloudwatch.CloudWatchSink (cloudwatch): Metrics pushed
2022-02-08 14:41:28,552 INFO org.apache.hadoop.hdfs.server.namenode.FSEditLog (FSEditLogAsync): Number of transactions: 6260 Total time for transactions(ms): 191 Number of transactions batched in Syncs: 4714 Number of syncs: 1546
SyncTimes(ms): 2069 3476
2022-02-08 14:45:01,781 INFO com.amazon.ws.enr.hadoop.util.ProcessRunnerImpl (cloudwatch): Executing command, timeout 10 SECONDS command /usr/bin/systemctl is-active metricscollector
2022-02-08 14:45:01,785 INFO com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): MetricsCollector Running Status is: active
2022-02-08 14:45:01,792 WARN com.amazon.ws.enr.hadoop.rpc.MetricsCollectorRpcClient (cloudwatch): Successfully call transferMetrics on Metrics Collector.
```

Step 12: again goto the cluster and click the “Application and user interfaces ” tab. Then copy “Resource Manager” and paste in chrome.

← → ↻ Not secure | ec2-3-234-214-88.compute-1.amazonaws.com:8088/cluster ☆ ⓘ Update ⓘ



All Applications

Cluster

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

Tools

Cluster Metrics

Apps Submitted	0	Apps Pending	0	Apps Running	2	Apps Completed	0	Containers Running	0 B	Memory Used	24 GB	Memory Total	0 B	Memory Reserved	0	VCoers Used	16	VCo	
----------------	---	--------------	---	--------------	---	----------------	---	--------------------	-----	-------------	-------	--------------	-----	-----------------	---	-------------	----	-----	--

Cluster Nodes Metrics

Active Nodes	0	Decommissioning Nodes	0	Decommissioned Nodes	0	Lost Nodes	0	Unhealthy Nodes	0	Rebooted Nodes	0
--------------	---	-----------------------	---	----------------------	---	------------	---	-----------------	---	----------------	---

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cl
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:32, vCores:1>	<memory:12288, vCores:8>	0

Show 20 ▾ entries

ID	User	Name	Application Type	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCoers	Allocated Memory MB	Reserved CPU VCoers	Reserved Memory MB	% of Queue	% Clk
application_1644316532039_0002	hadoop	nullpointerexceptioncount	MAPREDUCE	default	0	Tue Feb 8 16:36:09 +0550 2022	Tue Feb 8 16:36:10 +0550 2022	Tue Feb 8 16:36:46 +0550 2022	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	0.0	0.0
application_1644316532039_0001	hive	HIVE-sacdefcd-5cc3-4ab0-92c3-069df9460d4	TEZ	default	0	Tue Feb 8 16:08:15 +0550 2022	Tue Feb 8 16:08:16 +0550 2022	Tue Feb 8 16:13:26 +0550 2022	FINISHED	SUCCEEDED	N/A	N/A	N/A	N/A	N/A	0.0	0.0


Showing 1 to 2 of 2 entries

You can see “nullpointerexceptioncount” step is finished or not. Also succeeded or failed

Step 13 : click “ID” and then you will see your task is succeeded and all history details as follows

← → ↻ Not secure | ec2-3-234-214-88.compute-1.amazonaws.com:8088/cluster/app/application_1644316532039_0002 ☆ ⓘ Update ⓘ

Logged in as: dr who



Application application_1644316532039_0002

Cluster

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

Tools

Application Overview

User:	hadoop
Name:	nullpointerexceptioncount
Application Type:	MAPREDUCE
Application Tags:	
Application Priority:	0 (Higher Integer value indicates higher priority)
YarnApplicationState:	FINISHED
Queue:	default
FinalStatus Reported by AM:	SUCCEEDED
Started:	Tue Feb 08 11:06:09 +0000 2022
Launched:	Tue Feb 08 11:06:10 +0000 2022
Finished:	Tue Feb 08 11:06:46 +0000 2022
Elapsed:	36sec
Tracking URL:	History
Log Aggregation Status:	SUCCEEDED
Application Timeout (Remaining Time):	Unlimited
Diagnostics:	
Unmanaged Application:	false
Application Node Label expression:	<Not set>
AM container Node Label expression:	<DEFAULT_PARTITION>

Application Metrics

Total Resource Preempted:	<memory:0, vCores:0>
Total Number of Non-AM Containers Preempted:	0
Total Number of AM Containers Preempted:	0
Resource Preempted from Current Attempt:	<memory:0, vCores:0>
Number of Non-AM Containers Preempted from Current Attempt:	0
Aggregate Resource Allocation:	511061 MB-seconds, 166 vcore-seconds
Aggregate Preempted Resource Allocation:	0 MB-seconds, 0 vcore-seconds

Show 20 ▾ entries

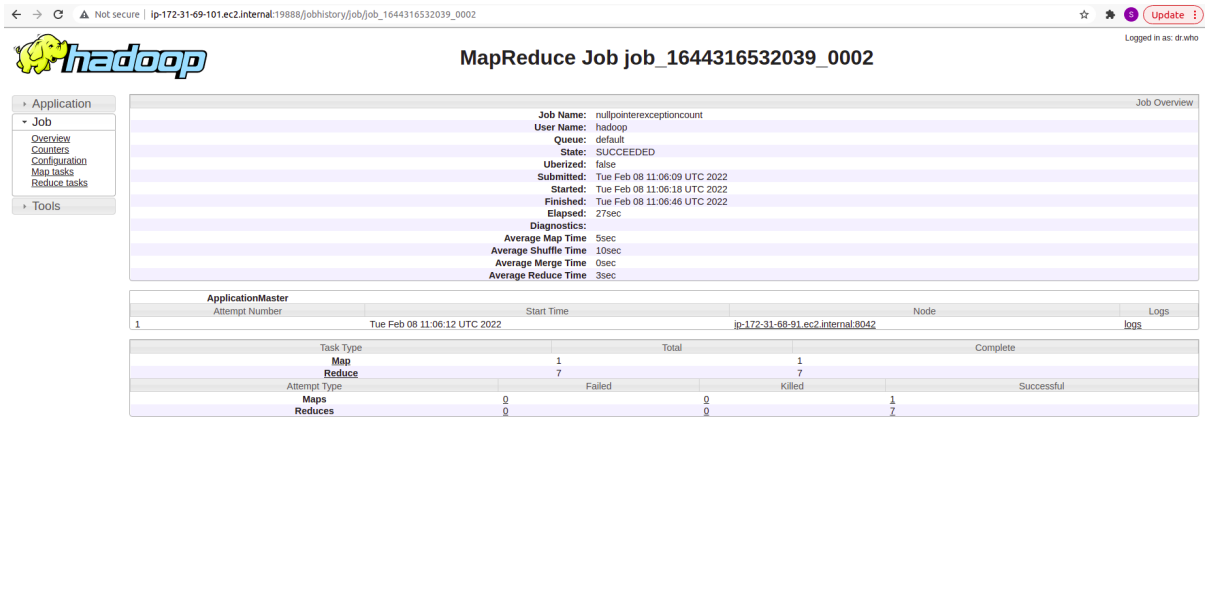
Attempt ID	Started	Node	Logs	Nodes blacklisted by the app	Nodes blacklisted by the system
appattempt_1644316532039_0002_000001	Tue Feb 8 16:36:09 +0550 2022	http://ip-172-31-68-91.ec2.internal:8042	Logs	0	0

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

Step 14: click history and you can see the number of Map and number of Reduce .

Eg: Map : 1
Reducer : 7

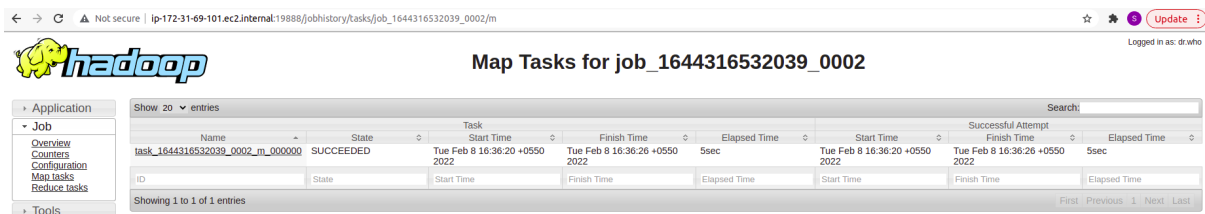


The screenshot shows the Hadoop MapReduce Job Overview page for job _1644316532039_0002. The page includes a sidebar with navigation links (Application, Job, Overview, Counters, Configuration, Map tasks, Reduce tasks, Tools) and a main content area. The main content area displays job details such as Job Name, User Name, Queue, State, Uberized, Submitted, Started, Finished, Elapsed, and Diagnostics. Below this, there is a table showing the ApplicationMaster attempt number, start time, node, and logs. The table also shows the task type (Map, Reduce) and the number of tasks (1 Map, 7 Reduce). The table is filtered by 'Task Type' and 'Attempt type'.

Task Type	Total	Complete
Map	1	1
Reduce	7	7

Attempt type	Maps	Failed	Killed	Successful
Maps	0	0	0	1
Reduces	0	0	0	7

Step 15 : click “Map tasks”



The screenshot shows the Hadoop Map Tasks for job _1644316532039_0002. The page includes a sidebar with navigation links (Application, Job, Overview, Counters, Configuration, Map tasks, Reduce tasks, Tools) and a main content area. The main content area displays a table of map tasks with columns for Name, State, Task, Start Time, Finish Time, Elapsed Time, and Successful Attempt. The table shows one entry for task _1644316532039_0002_m_000000, which is in the SUCCEEDED state. The table is filtered by 'Show 20 entries' and 'Search'.

Name	State	Task	Start Time	Finish Time	Elapsed Time	Successful Attempt
task_1644316532039_0002_m_000000	SUCCEEDED		Tue Feb 8 16:36:20 +0550 2022	Tue Feb 8 16:36:26 +0550 2022	5sec	Tue Feb 8 16:36:20 +0550 2022

When you click attempt you can see the status

← → 🔒 Not secure | ip-172-31-69-101.ec2.internal:19888/jobhistory/task_1644316532039_0002_m_000000

☆ 🌟 🔄 Update ⓘ Logged in as: dr.who

hadoop Attempts for task_1644316532039_0002_m_000000

Application
Job
Task
Task Overview Counters
Tools

Show 20 ▾ entries

Attempt

State

Status

Node

Logs

Start Time

Finish Time

Elapsed Time

Note

attempt_1644316532039_0002_m_000000_0	SUCCEEDED	map	default-rack/ip-172-31-73-227.ec2.internal:8042	logs	Tue Feb 8 16:36:20 +0550 2022	Tue Feb 8 16:36:26 +0550 2022	5sec	
---------------------------------------	-----------	-----	---	------	-------------------------------	-------------------------------	------	--

Attempt

State

Status

Node

Logs

Start Time

Finish Time

Elapsed Time

Note

Showing 1 to 1 of 1 entries

First Previous 1 Next Last

Step 15 : click “Reducer tasks “ can see the status as “SUCCEEDED” or not

← → 🔒 Not secure | ip-172-31-69-101.ec2.internal:19888/jobhistory/tasks/job_1644316532039_0002/r

☆ 🌟 🔄 Update ⓘ Logged in as: dr.who

hadoop Reduce Tasks for job_1644316532039_0002

Application
Job
Task Overview Counters Configuration Map tasks Reduce tasks
Tools

Show 20 ▾ entries

Task

Successful Attempt

Name	State	Start Time	Finish Time	Elapsed Time	Start Time	Shuffle Finish Time	Merge Finish Time	Finish Time	Elapsed Time Shuffle	Elapsed Time Merge	Elapsed Time Reduce	Elapsed Time
task_1644316532039_0002_r_000000	SUCCEEDED	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:42 +0550 2022	13sec	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:39 +0550 2022	Tue Feb 8 16:36:39 +0550 2022	Tue Feb 8 16:36:42 +0550 2022	9sec	0sec	3sec	13sec
task_1644316532039_0002_r_000001	SUCCEEDED	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:42 +0550 2022	12sec	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:39 +0550 2022	Tue Feb 8 16:36:39 +0550 2022	Tue Feb 8 16:36:42 +0550 2022	9sec	0sec	3sec	12sec
task_1644316532039_0002_r_000002	SUCCEEDED	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:45 +0550 2022	15sec	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:45 +0550 2022	12sec	0sec	3sec	15sec
task_1644316532039_0002_r_000003	SUCCEEDED	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:45 +0550 2022	15sec	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:45 +0550 2022	12sec	0sec	3sec	15sec
task_1644316532039_0002_r_000004	SUCCEEDED	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:42 +0550 2022	13sec	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:39 +0550 2022	Tue Feb 8 16:36:39 +0550 2022	Tue Feb 8 16:36:42 +0550 2022	9sec	0sec	3sec	13sec
task_1644316532039_0002_r_000005	SUCCEEDED	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:44 +0550 2022	15sec	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:44 +0550 2022	12sec	0sec	3sec	15sec
task_1644316532039_0002_r_000006	SUCCEEDED	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:45 +0550 2022	15sec	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:45 +0550 2022	12sec	0sec	3sec	15sec

ID

State

Start Time

Finish Time

Elapsed Time

Start Time

Shuffle Time

Merge Time

Finish Time

Elapsed Shuffle

Elapsed Merge

Elapsed Reduce

Elapsed Time

Note

Showing 1 to 7 of 7 entries

First Previous 1 Next Last

Step 16 : click reducer name and you can see as follows

← → 🔒 Not secure | ip-172-31-69-101.ec2.internal:19888/jobhistory/task_1644316532039_0002_r_000006

☆ 🌟 🔄 Update ⓘ Logged in as: dr.who

hadoop Attempts for task_1644316532039_0002_r_000006

Application
Job
Task
Task Overview Counters
Tools

Show 20 ▾ entries

Attempt

State

Status

Node

Logs

Start Time

Shuffle Finish Time

Merge Finish Time

Finish Time

Elapsed Time Shuffle

Elapsed Time Merge

Elapsed Time Reduce

Elapsed Time

Note

attempt_1644316532039_0002_r_000006_0	SUCCEEDED	reduce > reduce	default-rack/ip-172-31-73-227.ec2.internal:8042	logs	Tue Feb 8 16:36:29 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:41 +0550 2022	Tue Feb 8 16:36:45 +0550 2022	12sec	0sec	3sec	15sec
---------------------------------------	-----------	-----------------	---	------	-------------------------------	-------------------------------	-------------------------------	-------------------------------	-------	------	------	-------

Attempt

State

Status

Node

Logs

Start Time

Shuffle Time

Merge Time

Finish Time

Elapsed Shuffle

Elapsed Merge

Elapsed Reduce

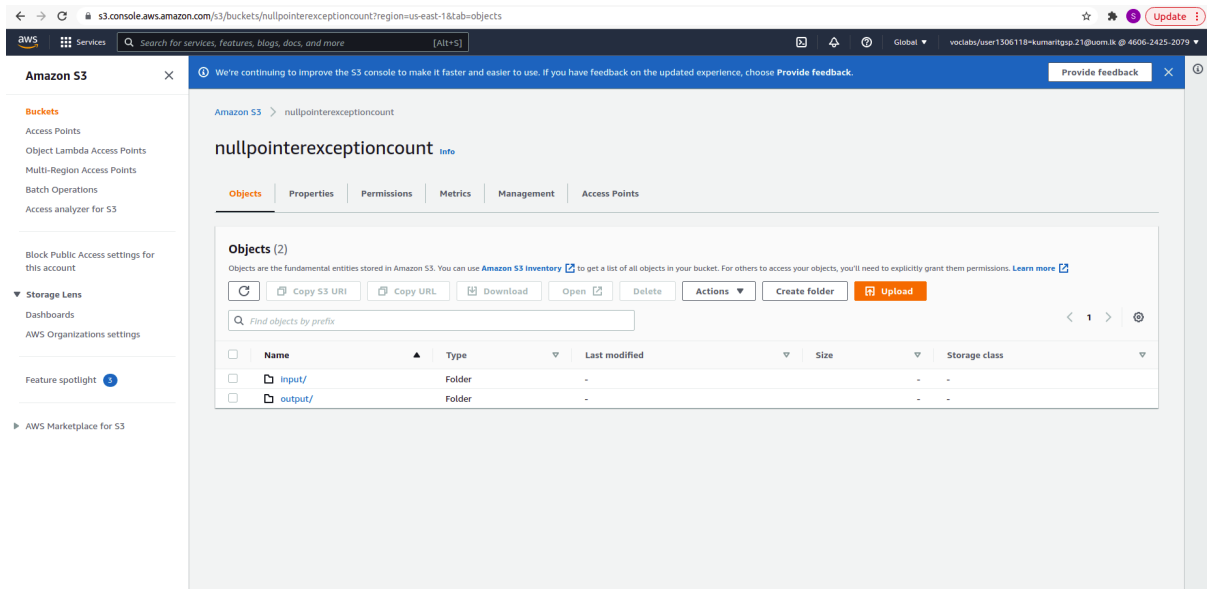
Elapsed Time

Note

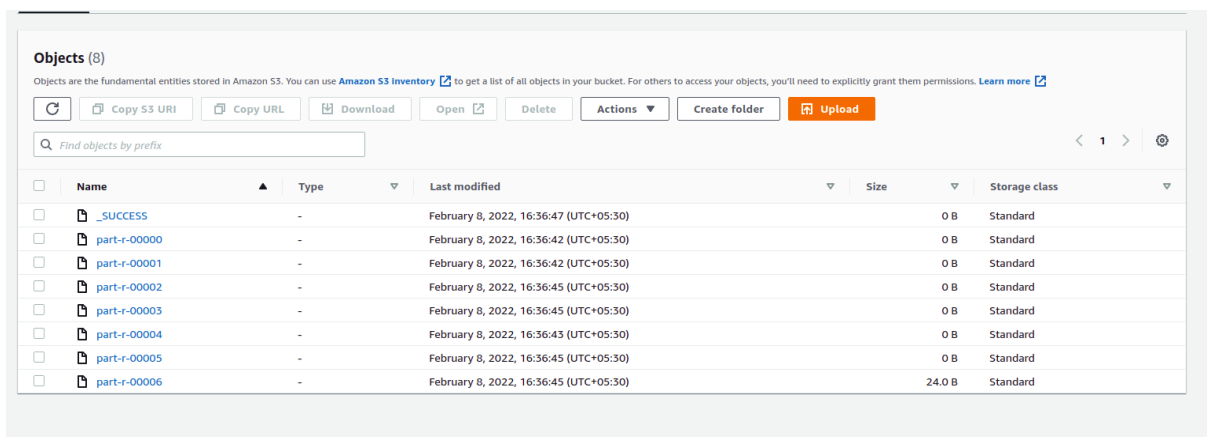
Showing 1 to 1 of 1 entries

First Previous 1 Next Last

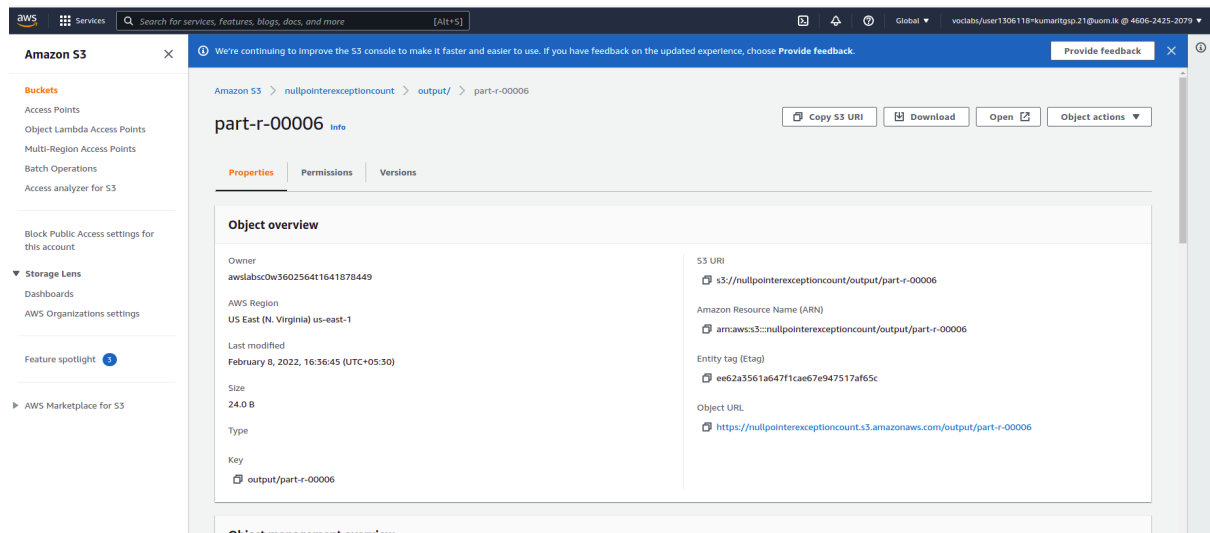
Step 17: goto S3 location and you will see there is “output” directory which was our output destination when we were creating step



Step 18: goto the output directory and you will see the file structure as follows. There are a _SUCCESS file and 7 reducers .



Step 19 : click reducer and you can see the output. In here i clicked “part-r-00006Info” and you will see as follows.



Step 20 : click “open” you can see how many null pointer exceptions are in the log file

