National University
of computer and emerging sciences

BDA LAB PROJECT

Name: Jiyad Khan

Roll No: 19I-1771

Section: BS-DS (N)

Course's Name: Big Data Analytics

Course's Instructor: Sir Saad Naeem

Due Date: 28th, May 2021

Master Node

Requirement 1:

There need to be minimum 3 nodes configured for multi node setup.

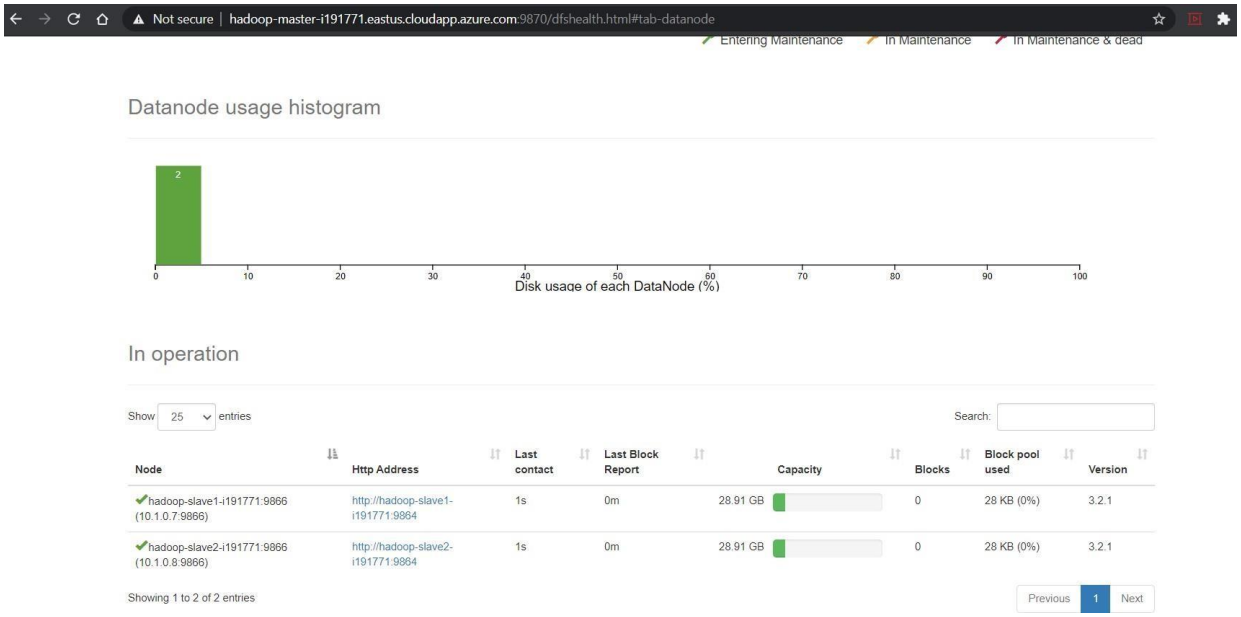
Numbers of the Machine	Name of the Machine	IP address of the Machine
Machine 1	hadoop-masters	10.1.0.6
Machine 2	hadoop-slave1-i191771	10.1.0.7
Machine 3	hadoop-slave2-i191771	10.1.0.8

Requirement 2:

The cluster is to be evaluated on the text file provided (“books_large_p1.txt” 2.52/2.3 GB).

Requirement 3: Data Node

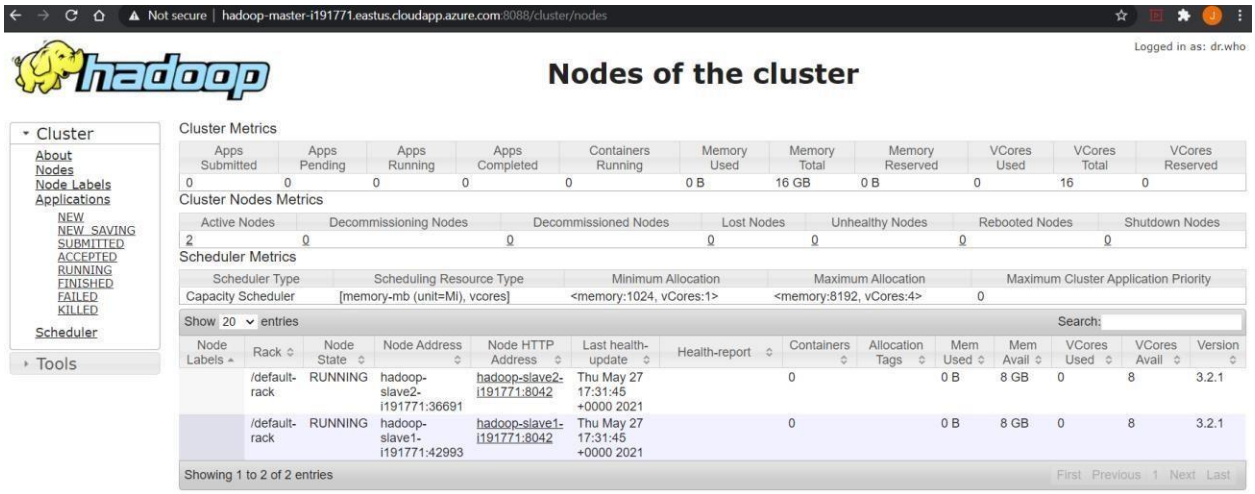
A Screenshot displaying all the data nodes that are up and sending their heartbeat to the Master Node.



Slaves Nodes

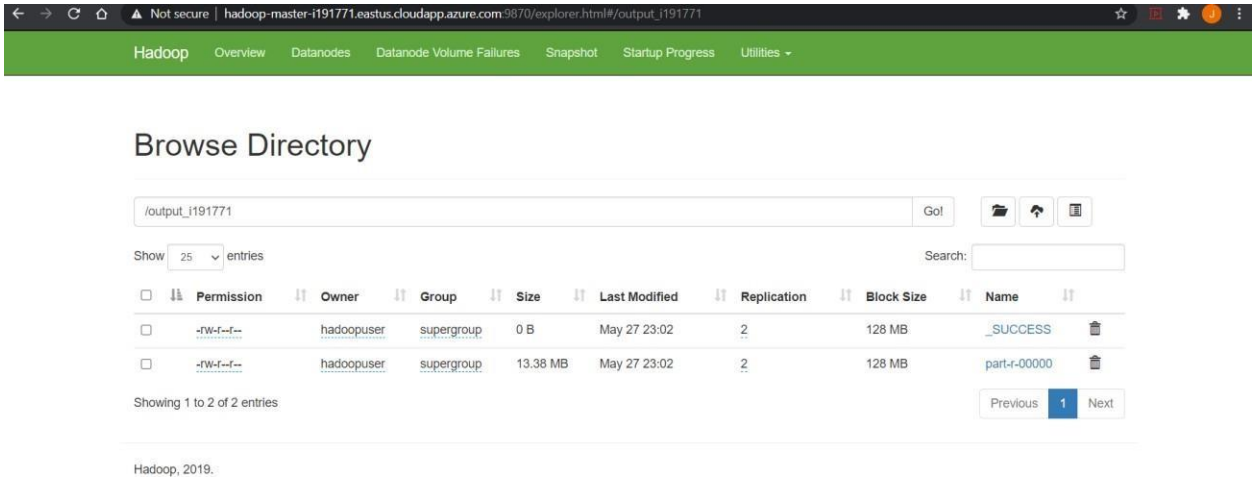
Requirement 1: Node States (here showing 32 GB Pooled Memory 8GB each)

A Screen shot showing the cluster metrics including total Pooled Memory and the number of Active Nodes.



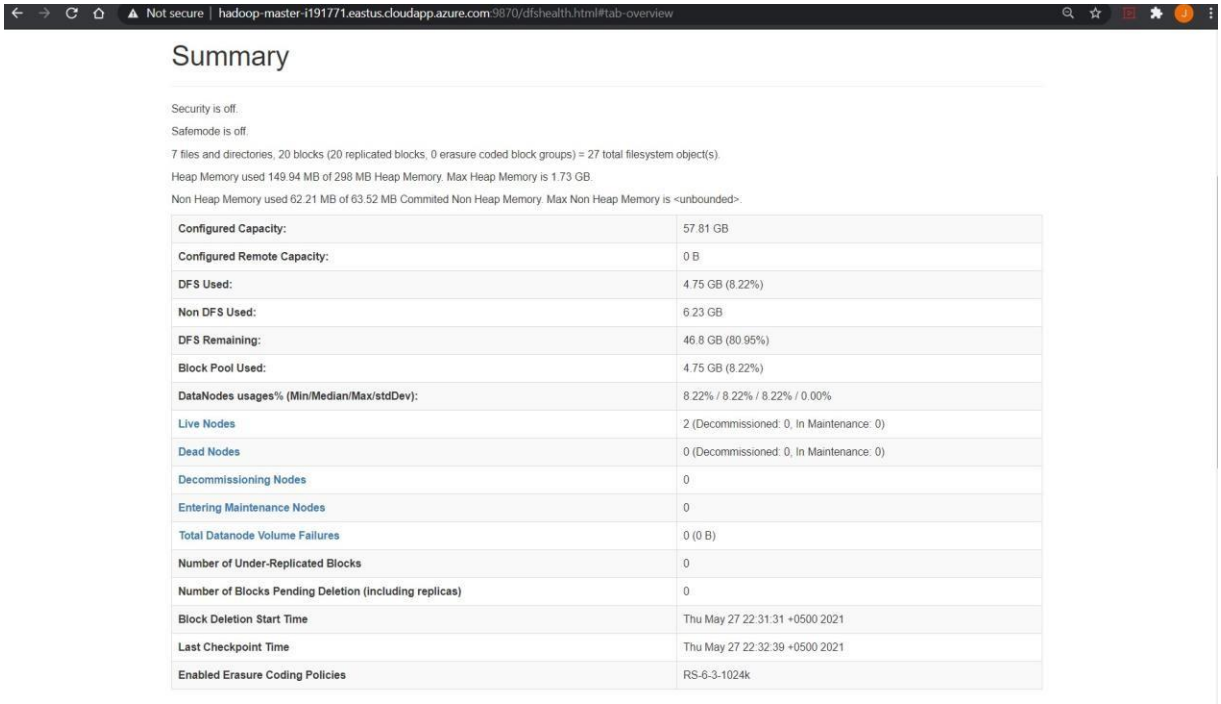
Requirement 2: Success on 2.5 GB Text file

A Screenshot displaying the directory of the cluster.



Requirement 3: Cluster Overview

A Screenshot showing live nodes, total configured capacity etc.



Summary

Security is off.

Safemode is off.

7 files and directories, 20 blocks (20 replicated blocks, 0 erasure coded block groups) = 27 total filesystem object(s).

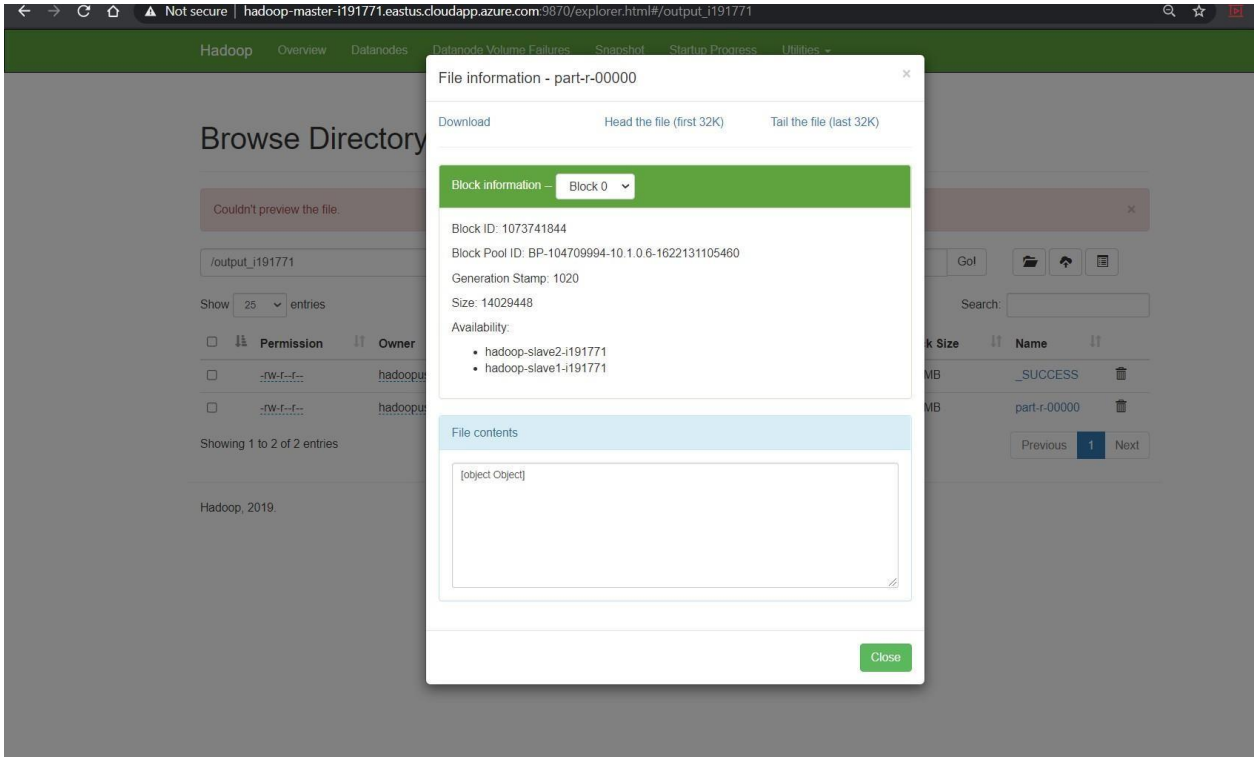
Heap Memory used 149.94 MB of 298 MB Heap Memory. Max Heap Memory is 1.73 GB.

Non Heap Memory used 62.21 MB of 63.52 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	57.81 GB
Configured Remote Capacity:	0 B
DFS Used:	4.75 GB (8.22%)
Non DFS Used:	6.23 GB
DFS Remaining:	46.8 GB (80.95%)
Block Pool Used:	4.75 GB (8.22%)
DataNodes usages% (Min/Median/Max/stdDev):	8.22% / 8.22% / 8.22% / 0.00%
Live Nodes	2 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Thu May 27 22:31:31 +0500 2021
Last Checkpoint Time	Thu May 27 22:32:39 +0500 2021
Enabled Erasure Coding Policies	RS-6-3-1024k

Requirement 4: GUI Results with Active Nodes

A Screenshot that shows the head of the results with “Availabilty” parameter showing how many slave nodes were involved in processing these results.



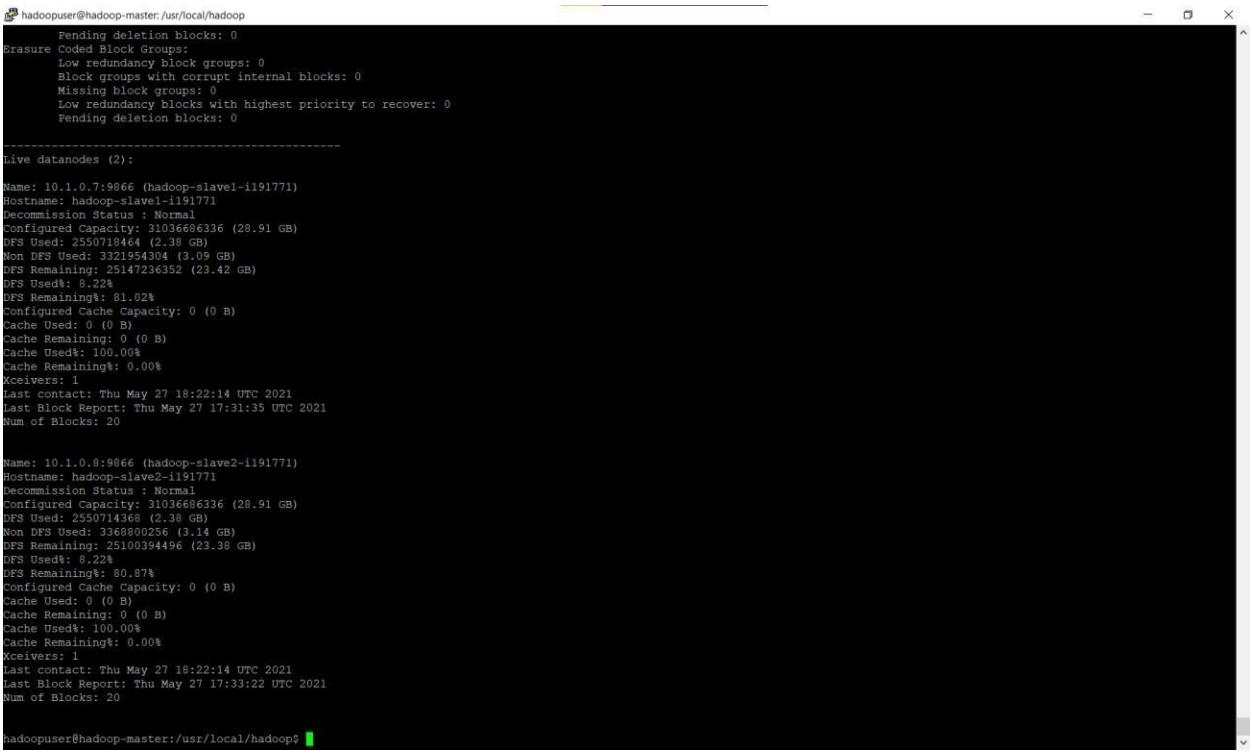
Requirement 5: CAT out the results in terminal

A Screenshot of the output using cat command running on “part-00000”.

Prove Task

Requirement 7:

- Prove that the 2.5 GB file was chunked, distributed, and processed on multiplenodes.
- Change the number of reducers to be equal to the number of Data Nodes.
- Prove how many mappers and reducers are configured on the cluster. By default,there is only one reducer which becomes the bottleneck of the whole process.



```
hadoopuser@hadoop-master: /usr/local/hadoop
        Pending deletion blocks: 0
Erasure Coded Block Groups:
        Low redundancy block groups: 0
        Block groups with corrupt internal blocks: 0
        Missing block groups: 0
        Low redundancy blocks with highest priority to recover: 0
        Pending deletion blocks: 0

-------------------------------------------------
Live datanodes (2):

Name: 10.1.0.7:9866 (hadoop-slave1-i191771)
Hostname: hadoop-slave1-i191771
Decommission Status : Normal
Configured Capacity: 31036686336 (28.91 GB)
DFS Used: 2550718464 (2.38 GB)
Non DFS Used: 3321954304 (3.09 GB)
DFS Remaining: 25147236352 (23.42 GB)
DFS Used%: 8.22%
DFS Remaining%: 81.02%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Thu May 27 18:22:14 UTC 2021
Last Block Report: Thu May 27 17:31:35 UTC 2021
Num of Blocks: 20


Name: 10.1.0.8:9866 (hadoop-slave2-i191771)
Hostname: hadoop-slave2-i191771
Decommission Status : Normal
Configured Capacity: 31036686336 (28.91 GB)
DFS Used: 2550714368 (2.38 GB)
Non DFS Used: 3368800256 (3.14 GB)
DFS Remaining: 25100394496 (23.38 GB)
DFS Used%: 8.22%
DFS Remaining%: 80.87%
Configured Cache Capacity: 0 (0 B)
Cache Used: 0 (0 B)
Cache Remaining: 0 (0 B)
Cache Used%: 100.00%
Cache Remaining%: 0.00%
Xceivers: 1
Last contact: Thu May 27 18:22:14 UTC 2021
Last Block Report: Thu May 27 17:33:22 UTC 2021
Num of Blocks: 20

hadoopuser@hadoop-master:/usr/local/hadoop$
```