**Lab Sheet: Single-Node Hadoop Cluster with Docker**

Name: K. D. Hamitha Apsara

Index: 19APSE4270

Semester: 3rd Year 2nd Semester

Due Date: 18th of November 2024

**Assignment**

Sabaragamuwa University of Sri Lanka

Faculty of Computing

Department of Software Engineering

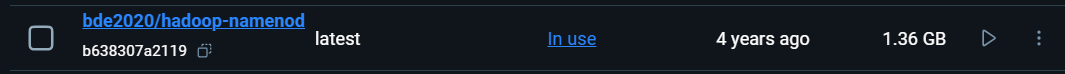
**SE6103 - Parallel and Distributed Systems**

1. Confirming Pre-requisites

Command: *docker--version*

Output: 

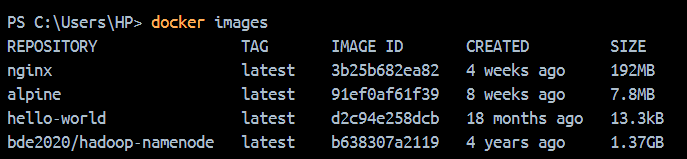
1. Step 1: Pull the Hadoop Docker image

Command: *docker pull bde2020/hadoop-namenode:latest*

Output:

1. Verify the Download

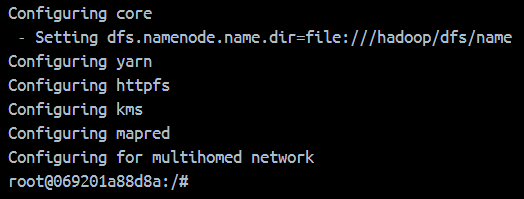
Command: docker images

Output: 

1. Step 2: Start the Hadoop Container

Command:

*docker run-it--name hadoop-cluster-p 9870:9870-p 8088:8088-p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash*

Output:

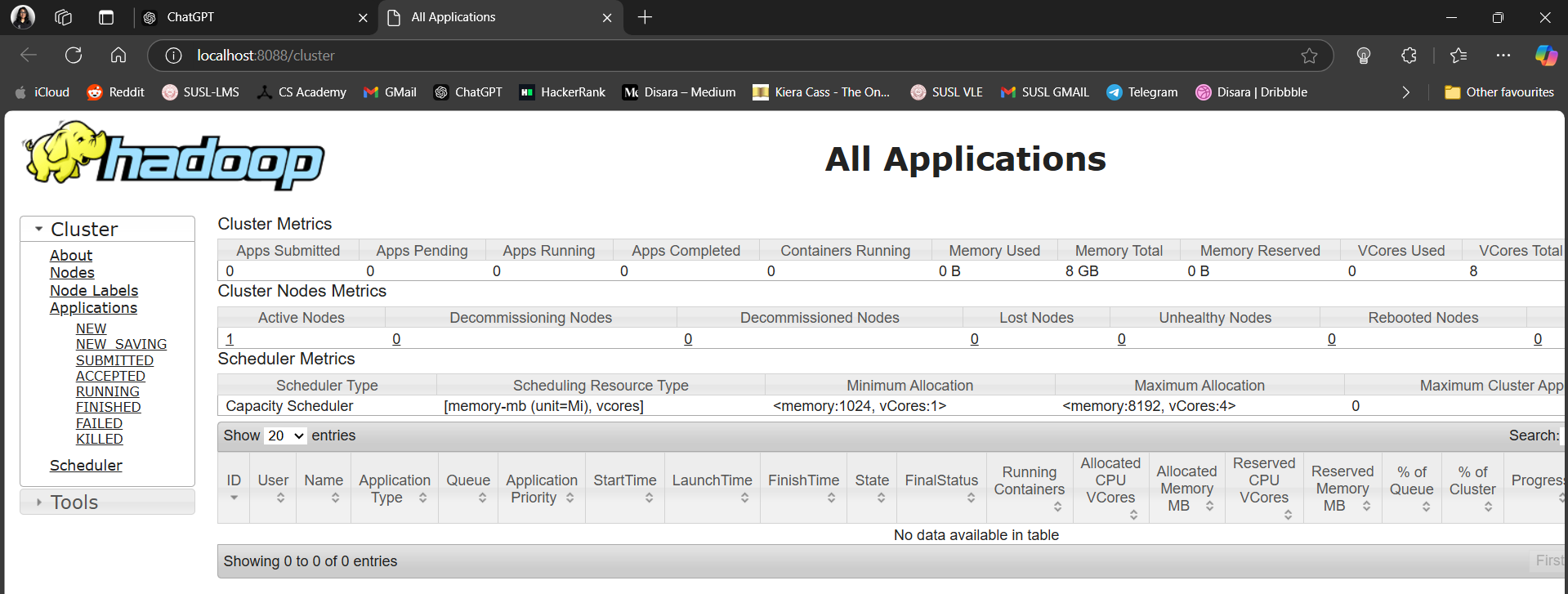
1. Start Hadoop Services

Commands to run each service separately:

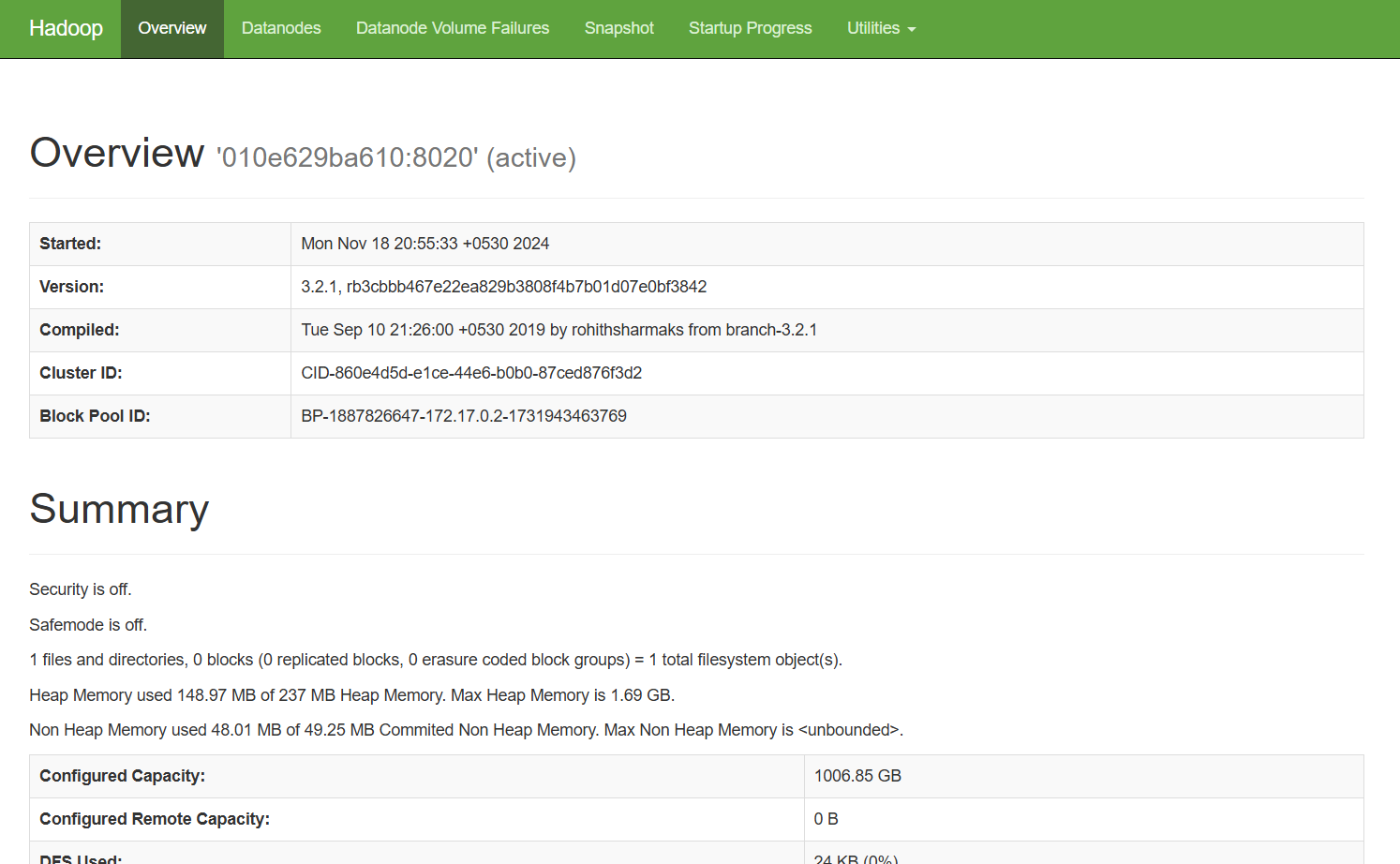
* */opt/hadoop-3.2.1/bin/hdfs --daemon start namenode*
* */opt/hadoop-3.2.1/bin/hdfs --daemon start datanode*
* */opt/hadoop-3.2.1/bin/yarn --daemon start resourcemanager*
* */opt/hadoop-3.2.1/bin/yarn --daemon start nodemanager*

1. Step 3: Access Hadoop Web Interfaces

* HDFS Web Interface (Resource Manager):



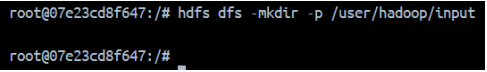
* YARN Web Interface (NameNode Web UI):



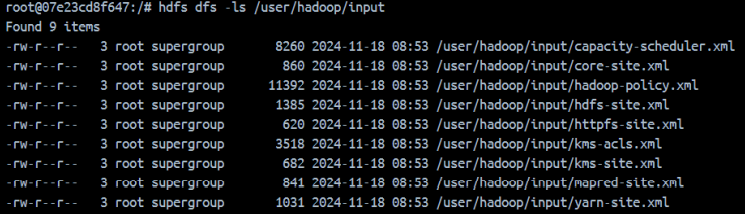
Step 4: Running a Sample MapReduce Job

Upload Sample Data to HDFS

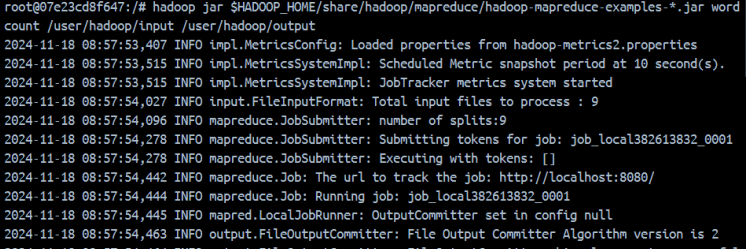
Command and Output:



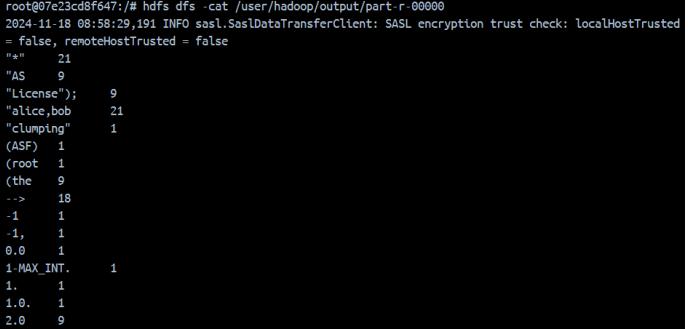
Files uploaded successfully:

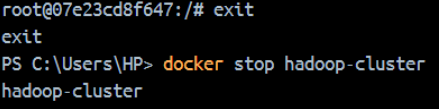


7.2 Run the WordCount Job

****

7.3. Check the output:

****

****8. Step 5: Exiting the Container

8.1. Stop the container:

8.2. Restart the container: