**Lab Sheet 01**

Sabaragamuwa University of Sri Lanka  
Faculty of Computing  
Department of Software Engineering  
SE6103 - Parallel and Distributed Systems

Name : M.R.S.C. Abewardhana  
Index No : 19APSE4268  
Academic Period : VI Semester

**Check docker version**

* docker version
* 

**Pull and verify the Hadoop image**

* docker pull bde2020/hadoop-namenode:latest
* A screen shot of a computer

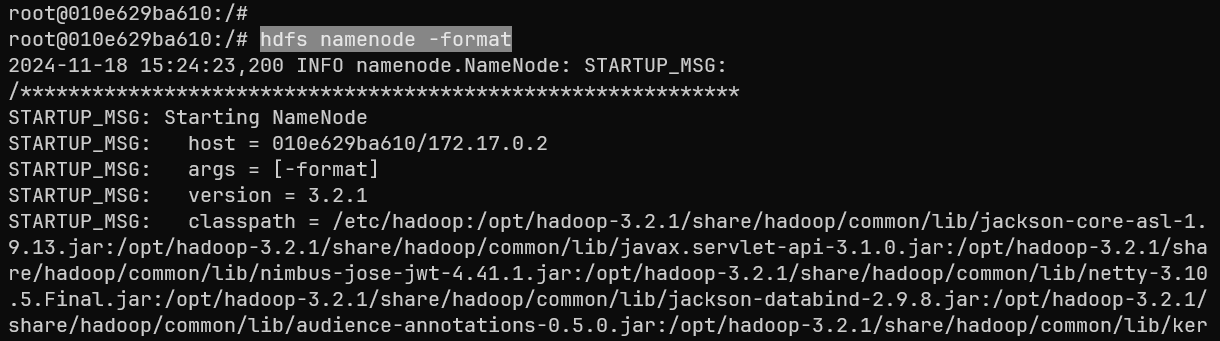
  Description automatically generated

**Run the docker container**

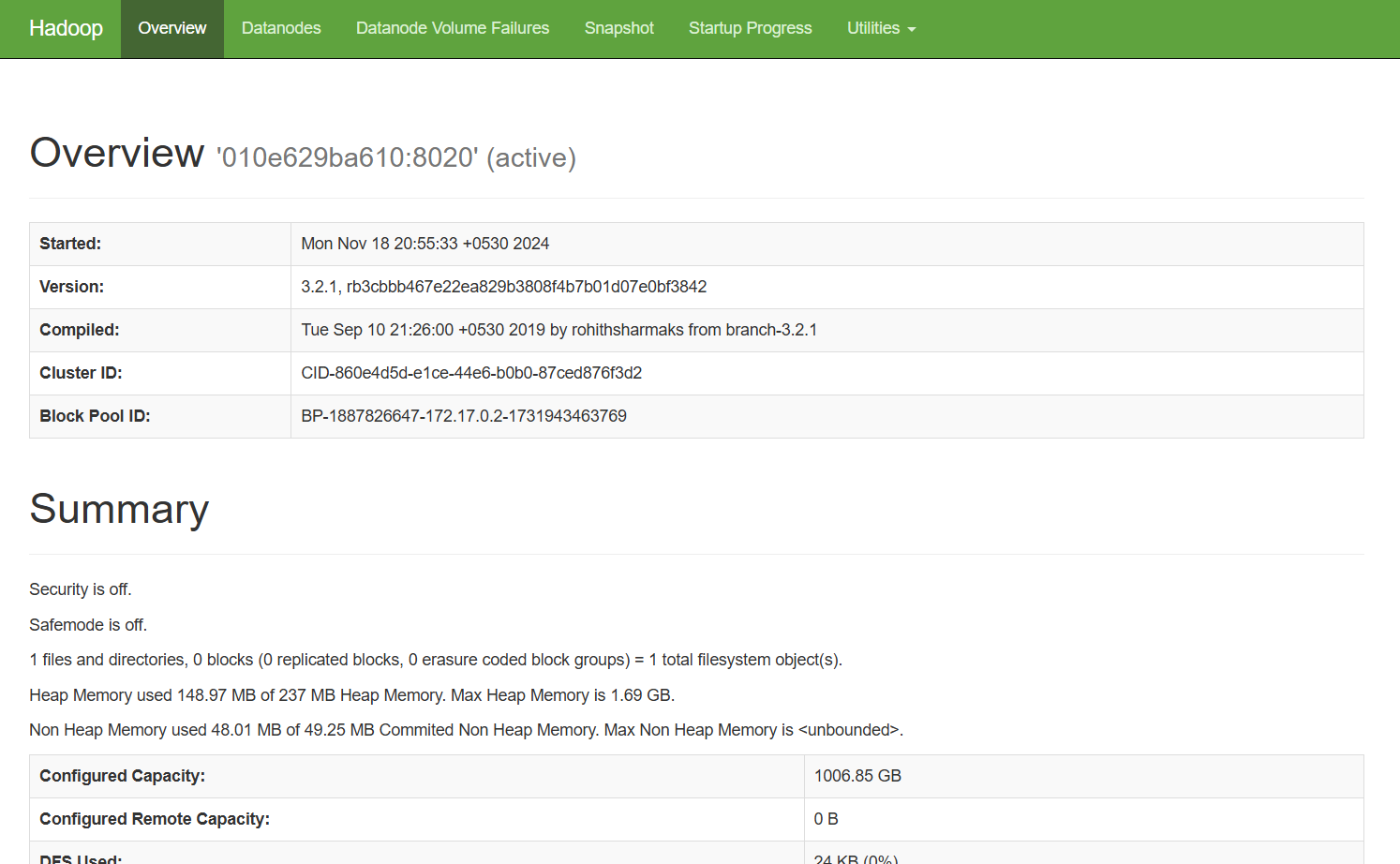
* docker run -it --name hadoop-cluster -p 9870:9870 -p 8088:8088 -p 50070:50070 bde2020/hadoop-namenode:latest /bin/bash
* A computer screen with white text

  Description automatically generated

**Configure the Hadoop file system**

* hdfs namenode -format
* hdfs namenode &
* hdfs datanode &
* 

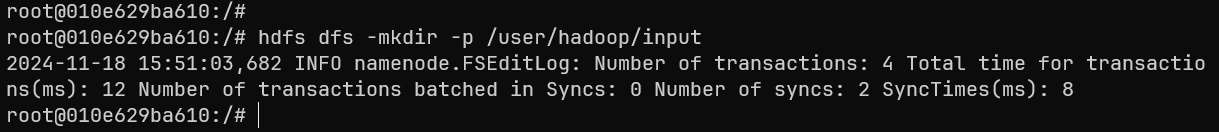
**Check the Hadoop health using local port**



**Start the node manager and resource manager**

* yarn nodemanager &
* yarn resourcemanager &

**Upload sample data to HDFS**

* hdfs dfs -mkdir -p /user/hadoop/input
* 
* hdfs dfs -put $HADOOP\_HOME/etc/hadoop/\*.xml /user/hadoop/input
* A screen shot of a computer

  Description automatically generated

**Run the word counter job**

* hadoop jar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-\*.jar wordcount /user/hadoop/input /user/hadoop/output
* A screenshot of a computer program

  Description automatically generated

**Check the output**

* hdfs dfs -cat /user/hadoop/output/part-r-00000
* A computer screen with white text

  Description automatically generated

**Stop the cluster**

* docker stop hadoop-cluster
* 