**Exp-1**

Downloading and installing Hadoop on Ubuntu, Understanding different Hadoop modes, Startup scripts, Configuration files

# Aim:

To successfully install, configure, and run Hadoop on a local system using a single-node setup.

# Procedure:

1. **Install Java and SSH:**
   * Update your package lists and install OpenJDK 8 and SSH. sudo apt update

sudo apt install openjdk-8-jdk

java -version # Verify Java installation sudo apt install ssh

# Create Hadoop User:

* + Add a dedicated user for Hadoop and generate SSH keys for passwordless SSH. sudo adduser hadoop

su - hadoop # Switch to Hadoop user ssh-keygen -t rsa

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys chmod 640 ~/.ssh/authorized\_keys

ssh localhost # Test SSH connection to localhost

# Download and Install Hadoop:

* + Download the latest Hadoop version (3.3.6), extract the tarball, and move it to the desired location.

wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz tar -xvzf hadoop-3.3.6.tar.gz

mv hadoop-3.3.6 hadoop

# Configure Environment Variables:

* + **Update. bashrc to include Hadoop and Java paths.**

nano ~/.bashrc

# Add the following lines at the end

export JAVA\_HOME=/usr/lib/jvm/java-8-openjdk-amd64

export HADOOP\_HOME=$HOME/hadoop

export PATH=$PATH:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin source ~/.bashrc # Apply changes

# Edit Hadoop Configuration Files:

* + Modify configuration files to set up the necessary Hadoop directories and services.

# core-site.xml:

nano $HADOOP\_HOME/etc/hadoop/core-site.xml # Add between <configuration></configuration>:

<property>

<name>fs.defaultFS</name>

<value>hdfs://localhost:9000</value>

</property>

# hdfs-site.xml:

nano $HADOOP\_HOME/etc/hadoop/hdfs-site.xml

# Add:

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.namenode.name.dir</name>

<value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>

</property>

<property>

<name>dfs.datanode.data.dir</name>

<value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>

</property>

# mapred-site.xml:

cp $HADOOP\_HOME/etc/hadoop/mapred-site.xml.template

$HADOOP\_HOME/etc/hadoop/mapred-site.xml nano $HADOOP\_HOME/etc/hadoop/mapred-site.xml

# Add:

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

# o yarn-site.xml:

nano $HADOOP\_HOME/etc/hadoop/yarn-site.xml

# Add:

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

# Format the NameNode:

* + Format the HDFS NameNode. hdfs namenode -format

# Start Hadoop:

* + Start Hadoop services (NameNode, DataNode, ResourceManager, and NodeManager).

start-all.sh

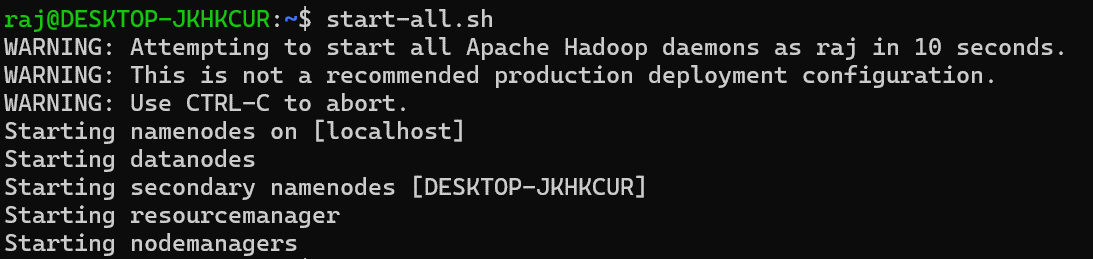
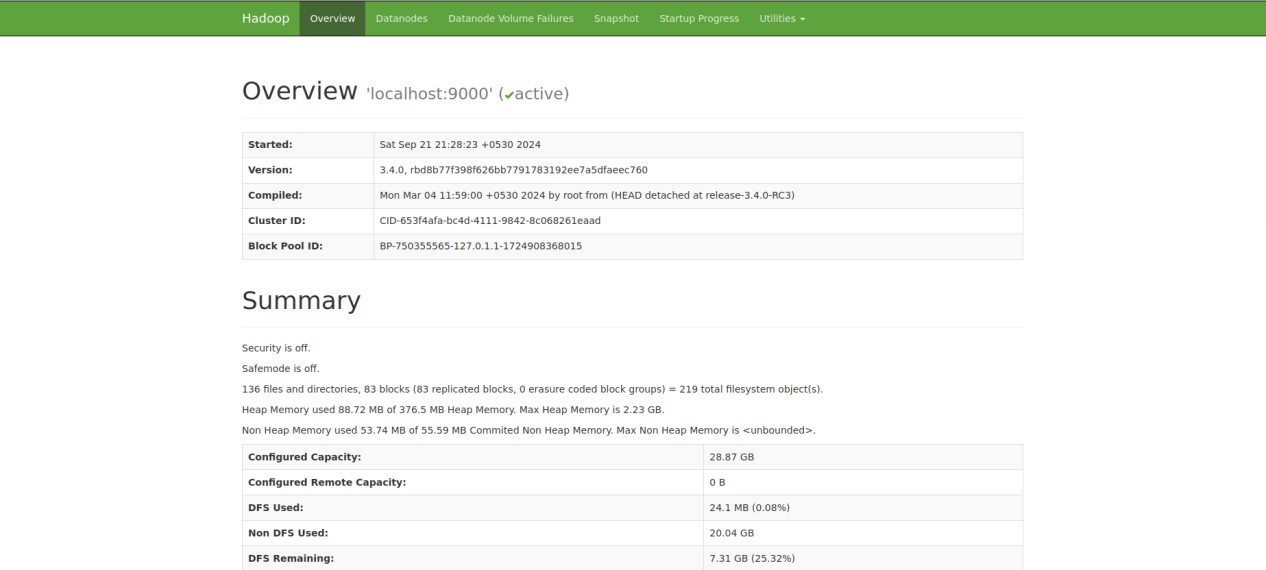
jps # Verify running services

# Access Web Interfaces:

* + Verify that Hadoop is running by accessing the following URLs:
    - **NameNode:** http://localhost:9870
    - **Resource Manager:** http://localhost:8088

# Stop Hadoop Cluster:

* + Stop all Hadoop services. stop-all.sh



# RESULT:

The step-by-step installation and configuration of Hadoop on Ubuntu system have been successfully completed.