2. Install, configure and run Hadoop and HDFS.

Step 1: Prerequisites

1. Install Java: Hadoop requires Java to run. Install it using:

bash

sudo apt update

sudo apt install openjdk-11-jdk -y

Verify the installation:

bash

java -version

2. Set JAVA\_HOME Environment Variable: Add the following line to your .bashrc or .zshrc file:

bash

export JAVA\_HOME=/usr/lib/jvm/java-11-openjdk-amd64

export PATH=$JAVA\_HOME/bin:$PATH

Apply changes:

bash

source ~/.bashrc

Step 2: Download Hadoop

1. Download the latest stable version of Hadoop from the official Apache Hadoop site: Apache Hadoop.

2. Extract the downloaded tarball:

bash

tar -xvzf hadoop-X.Y.Z.tar.gz

sudo mv hadoop-X.Y.Z /usr/local/hadoop

Step 3: Configure Hadoop

1. Update Environment Variables: Add the following to .bashrc:

bash

export HADOOP\_HOME=/usr/local/hadoop

export PATH=$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin:$PATH

export HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop

Apply changes:

bash

source ~/.bashrc

2. Edit Configuration Files: Navigate to the Hadoop configuration directory:

bash

cd $HADOOP\_CONF\_DIR

o Core Site (core-site.xml):

xml

<configuration>

<property>

<name>fs.defaultFS</name>

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

</property>

</configuration>

o HDFS Site (hdfs-site.xml):

xml

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

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

<value>file:///usr/local/hadoop/hdfs/namenode</value>

</property>

<property>

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

<value>file:///usr/local/hadoop/hdfs/datanode</value>

</property>

</configuration>

o MapReduce Site (mapred-site.xml): Rename and edit:

bash

cp mapred-site.xml.template mapred-site.xml

Content:

xml

<configuration>

<property>

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

<value>yarn</value>

</property>

</configuration>

o YARN Site (yarn-site.xml):

xml

<configuration>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

</configuration>

3. Format the NameNode:

bash

hdfs namenode -format

Step 4: Start Hadoop Services

1. Start HDFS:

bash

start-dfs.sh

Verify the NameNode and DataNode status via:

o http://localhost:9870 for the NameNode web UI.

2. Start YARN:

bash

start-yarn.sh

Verify YARN ResourceManager via:

o http://localhost:8088.

Step 5: Test Hadoop

1. Create a Directory in HDFS:

bash

hdfs dfs -mkdir /user

hdfs dfs -mkdir /user/yourusername

2. Copy a File to HDFS:

bash

hdfs dfs -put <local\_file\_path> /user/yourusername

3. List Files in HDFS:

bash

hdfs dfs -ls /user/yourusername

4. Run a Sample MapReduce Job:

bash

hadoop jar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-X.Y.Z.jar wordcount /user/yourusername/input /user/yourusername/output

5. Check the Output:

bash

hdfs dfs -cat /user/yourusername/output/part-r-00000