|  |  |
| --- | --- |
| EXP NO: 1(B) | Understanding different Hadoop modes Startup scripts, Configuration files. |
|  |
| DATE: |  |

**AIM: -**

**BACKGROUND THEORY: -**

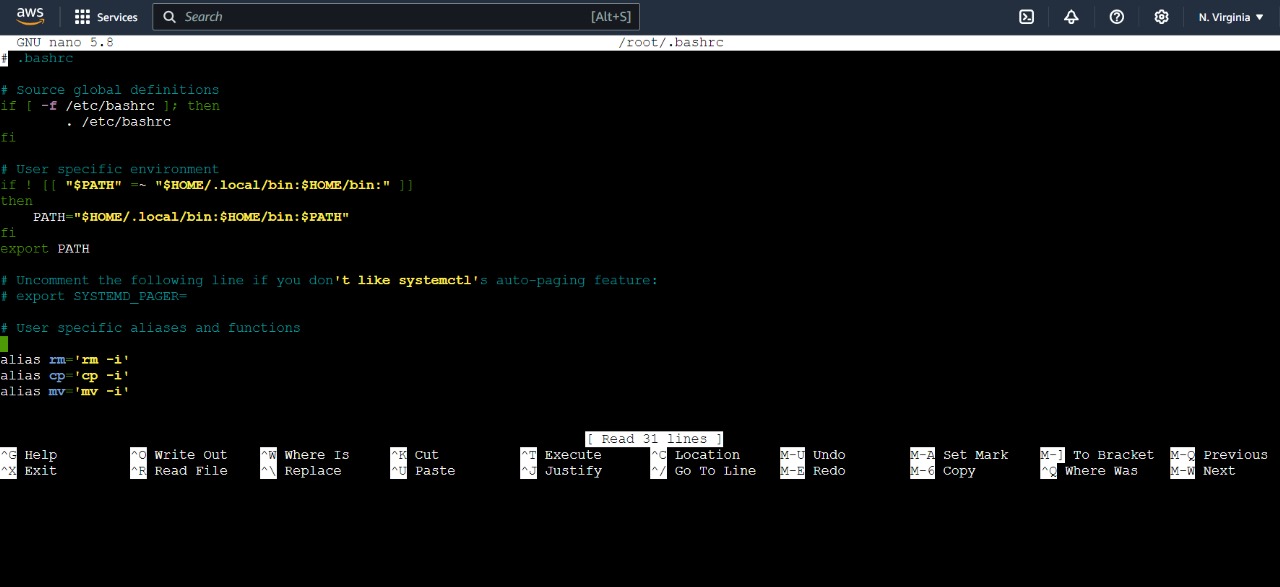
# **PROCEDURE: -**

* Switch to superuser mode using sudo su.
* Update all packages with sudo yum update -y.
* Install Java with sudo yum install java-11-openjdk-devel -y or sudo yum install java-1.8.0-amazon-corretto-devel -y.
* Verify the Java installation using java -version.
* Navigate to /usr/local/ using cd /usr/local/.
* Download Hadoop with sudo wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz.
* Extract the Hadoop archive using sudo tar -xvzf hadoop-3.3.6.tar.gz.
* Rename the extracted folder with sudo mv hadoop-3.3.6 hadoop.
* Edit the .bashrc file to add Hadoop and Java environment variables.
* Reload the .bashrc file with source ~/.bashrc and verify Hadoop installation using hadoop version.
* Edit core-site.xml to set the default file system and temporary directory.
* Configure hdfs-site.xml to set replication factor, NameNode, and DataNode directories.
* Create mapred-site.xml and configure the MapReduce framework to use YARN.
* Edit yarn-site.xml to set up ResourceManager and NodeManager services.
* Create the necessary Hadoop directories for the NameNode, DataNode, and temporary storage.
* Format the Hadoop HDFS filesystem using hdfs namenode -format.
* Start the HDFS services with start-dfs.sh.
* Start YARN services with start-yarn.sh.
* Check running processes with jps to ensure services are active.
* Access Hadoop web interfaces: NameNode at http://localhost:9870/ and ResourceManager at http://localhost:8088/.

# **CODING: -**

* sudo su
* sudo  yum update -y
* sudo yum install java-11-openjdk-devel -y **or** sudo yum install java-1.8.0-amazon-corretto-devel -y
* java -version
* cd /usr/local/
* sudo wget https://downloads.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz
* sudo tar -xvzf hadoop-3.3.6.tar.gz
* sudo mv hadoop-3.3.6 hadoop
* sudo nano ~/.bashrc
  + # Hadoop variables
  + export HADOOP\_HOME=/usr/local/hadoop
  + export PATH=$PATH:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin
  + # Java variables
  + export JAVA\_HOME=$(readlink -f /usr/bin/java | sed "s:/bin/java::")
  + export PATH=$PATH:$JAVA\_HOME/bin
* source ~/.bashrc
* hadoop version
* Configure core-site.xml
  + sudo nano $HADOOP\_HOME/etc/hadoop/core-site.xml
  + <configuration>
  + <property>
  + <name>fs.defaultFS</name>
  + <value>hdfs://localhost:9000</value>
  + </property>
  + <property>
  + <name>hadoop.tmp.dir</name>
  + <value>/usr/local/hadoop/tmp</value>
  + <description>Temporary directory for Hadoop</description>
  + </property>
  + </configuration>
* Configure hdfs-site.xml
  + sudo nano $HADOOP\_HOME/etc/hadoop/hdfs-site.xml
  + <configuration>
  + <property>
  + <name>dfs.replication</name>
  + <value>1</value> <!-- Since this is a single-node setup -->
  + </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>
* Create the mapred-site.xml[If the mapred-site.xml.template is not present]
  + sudo nano /usr/local/hadoop/etc/hadoop/mapred-site.xml
  + <configuration>
  + <property>
  + <name>mapreduce.framework.name</name>
  + <value>yarn</value>
  + </property>
  + </configuration>
* Configure yarn-site.xml
  + sudo nano $HADOOP\_HOME/etc/hadoop/yarn-site.xml
  + <configuration>
  + <property>
  + <name>yarn.nodemanager.aux-services</name>
  + <value>mapreduce\_shuffle</value>
  + </property>
  + <property>
  + <name>yarn.resourcemanager.resource-tracker.address</name>
  + <value>localhost:8025</value>
  + </property>
  + <property>
  + <name>yarn.resourcemanager.scheduler.address</name>
  + <value>localhost:8030</value>
  + </property>
  + <property>
  + <name>yarn.resourcemanager.address</name>
  + <value>localhost:8050</value>
  + </property>
  + </configuration>
* Set Up Hadoop Directories
  + sudo mkdir -p /usr/local/hadoop/hdfs/namenode
  + sudo mkdir -p /usr/local/hadoop/hdfs/datanode
  + sudo mkdir -p /usr/local/hadoop/tmp
* Format the HDFS Filesystem
  + hdfs namenode -format
* Start Hadoop Services(As Non - Root User)
  + start-dfs.sh
  + start-yarn.sh
  + Jps
* Access the Hadoop Web Interfaces
  + **NameNode**: http://localhost:9870/ (shows the HDFS overview)
  + **ResourceManager**: http://localhost:8088/ (shows the YARN overview)
  + **LocalHost - 127.0.0.1 or public DNS :- 3.117.182.16\**

# **OUTPUT: -**



# 