**Exp No: 9** **Date:**

# HADOOP

## SET UP A SINGLE HADOOP CLUSTER AND SHOW THE PROCESS USING WEB UI

**AIM:**

To set-up one node Hadoop cluster.

**PROCEDURE:**

1. System Update
2. Install Java
3. Add a dedicated Hadoop user
4. Install SSH and setup SSH certificates
5. Check if SSH works
6. Install Hadoop
7. Modify Hadoop config files
8. Format Hadoop filesystem
9. Start Hadoop
10. Check Hadoop through web UI
11. Stop Hadoop

**THEORY**

Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models. A Hadoop frame-worked application works in an environment that provides distributed storage and computation across clusters of computers. Hadoop is designed to scale up from a single server to thousands of machines, each offering local computation and storage.

**HADOOP ARCHITECTURE**

Hadoop framework includes following four modules:

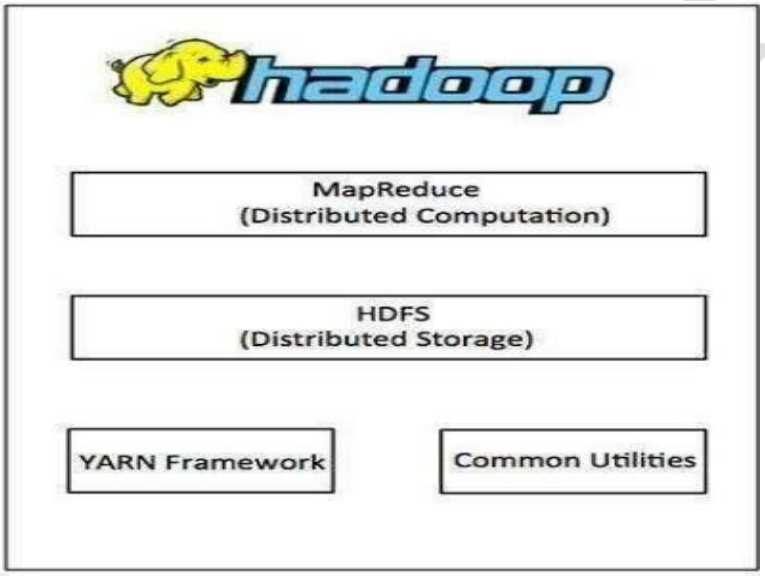
Hadoop Common: These are Java libraries and utilities required by other Hadoop modules. These libraries provide filesystem and OS level abstractions and contain the necessary Java files and scripts required to start Hadoop.

Hadoop YARN: This is a framework for job scheduling and cluster resource management.

Hadoop Distributed File System (HDFS): A distributed file system that provides high-throughput access to application data.

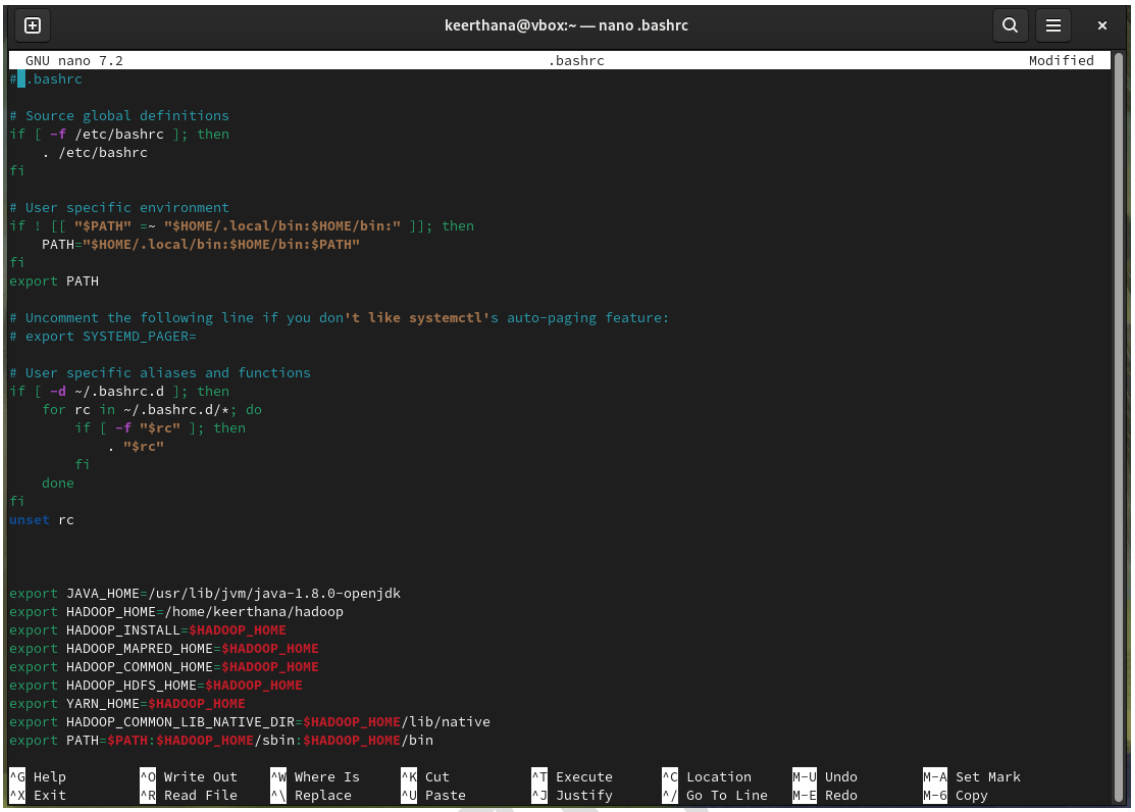
Hadoop MapReduce: This is a YARN-based system for parallel processing of large data sets.

We can use following diagram to depict these four components available in Hadoop framework.

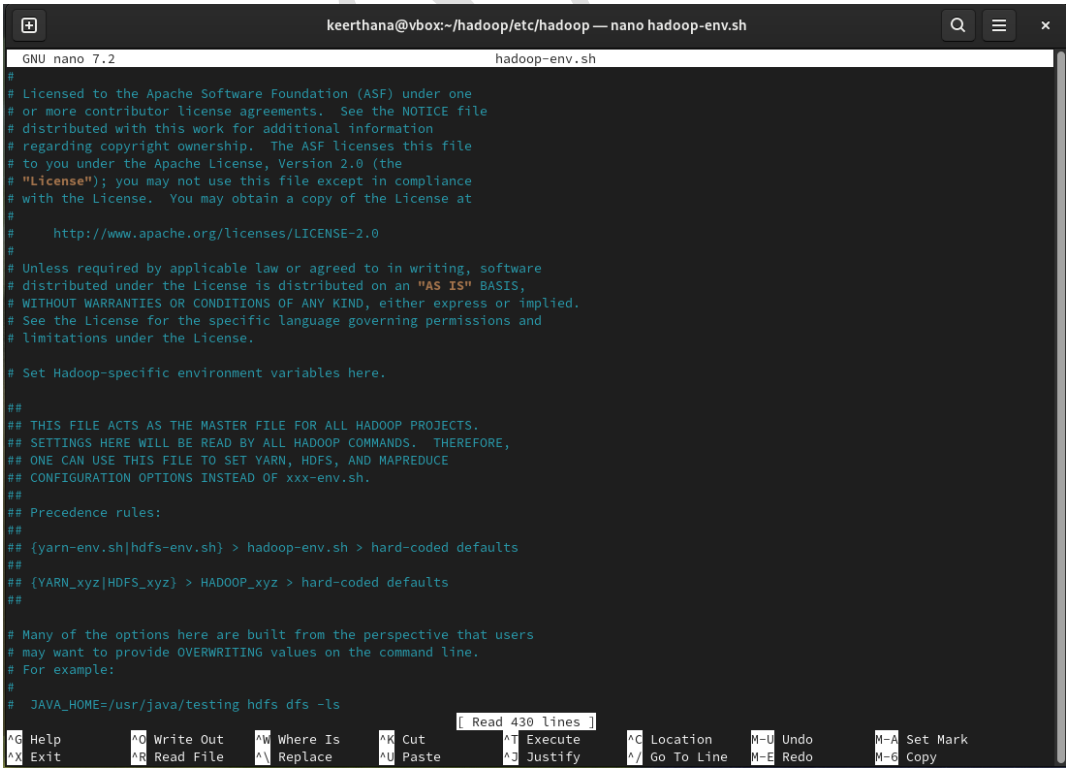


## PROCEDURE

## $ nano ~/.bashrc




## $ nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh



## $nano $HADOOP_HOME/etc/hadoop/core-site.xml

```
keerthana@vbox:~/hadoop/etc/hadoop — nano core-site.xml
  GNU nano 7.2                                    core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>
<name>fs.default.name</name>
<value>hdfs://localhost:9000</value>
</property>

<!--
<property>
<name>hadoop.tmp.dir</name>
<value>/home/cevher/apps/hadoop/tmp</value>
</property>
-->
<property>
<name>hadoop.proxyuser.keerthana.groups</name>
<value>*</value>
</property>
<property>
<name>hadoop.proxyuser.keerthana.hosts</name>
                                        [ Read 49 lines ]
^G Help      ^O Write Out   ^W Where Is   ^K Cut       ^T Execute     ^C Location   M-U Undo      M-A Set Mark
^X Exit      ^R Read File   ^\ Replace    ^U Paste     ^J Justify     ^/ Go To Line M-E Redo      M-6 Copy
```

## $nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml

```
keerthana@vbox:~/hadoop/etc/hadoop — nano hdfs-site.xml
  GNU nano 7.2                                    hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>

<property>
<name>dfs.replication</name>
<value>1</value>
</property>

<property>
<name>dfs.name.dir</name>
<value>file:///home/keerthana/hadoop/hadoopdata/hdfs/namenode</value>
</property>

<property>
<name>dfs.data.dir</name>
<value>file:///home/keerthana/hadoop/hadoopdata/hdfs/datanode</value>
</property>

</configuration>
                                        [ Read 36 lines ]
^G Help      ^O Write Out   ^W Where Is   ^K Cut       ^T Execute     ^C Location   M-U Undo      M-A Set Mark
^X Exit      ^R Read File   ^\ Replace    ^U Paste     ^J Justify     ^/ Go To Line M-E Redo      M-6 Copy
```

## $nano $HADOOP_HOME/etc/hadoop/mapred-site.xml

```
keerthana@vbox:~/hadoop/etc/hadoop — nano mapred-site.xml
  GNU nano 7.2                                    mapred-site.xml
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>mapred.job.tracker</name>
<value>localhost:9001</value>
</property>
<property>
<name>mapreduce.framework.name</name>
<value>yarn</value>
</property>
<property>
<name>mapreduce.application.classpath</name>
<value>$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/*:$HADOOP_MAPRED_HOME/share/hadoop/mapreduce/lib/*</value>
</property>
</configuration>

                                          [ Read 32 lines ]
^G Help      ^O Write Out   ^W Where Is    ^K Cut       ^T Execute    ^C Location   M-U Undo      M-A Set Mark
^X Exit      ^R Read File   ^\ Replace     ^U Paste     ^J Justify    ^/ Go To Line M-E Redo      M-6 Copy
```

## $nano $HADOOP_HOME/etc/hadoop/yarn-site.xml

```
keerthana@vbox:~/hadoop/etc/hadoop — nano yarn-site.xml
  GNU nano 7.2                                    yarn-site.xml
<?xml version="1.0"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->
<configuration>

<!-- Site specific YARN configuration properties -->

<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
<property>
<name>yarn.nodemanager.env-whitelist</name>
<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</>
</property>
</configuration>

                                          [ Read 27 lines ]
^G Help      ^O Write Out   ^W Where Is    ^K Cut       ^T Execute    ^C Location   M-U Undo      M-A Set Mark
^X Exit      ^R Read File   ^\ Replace     ^U Paste     ^J Justify    ^/ Go To Line M-E Redo      M-6 Copy
```

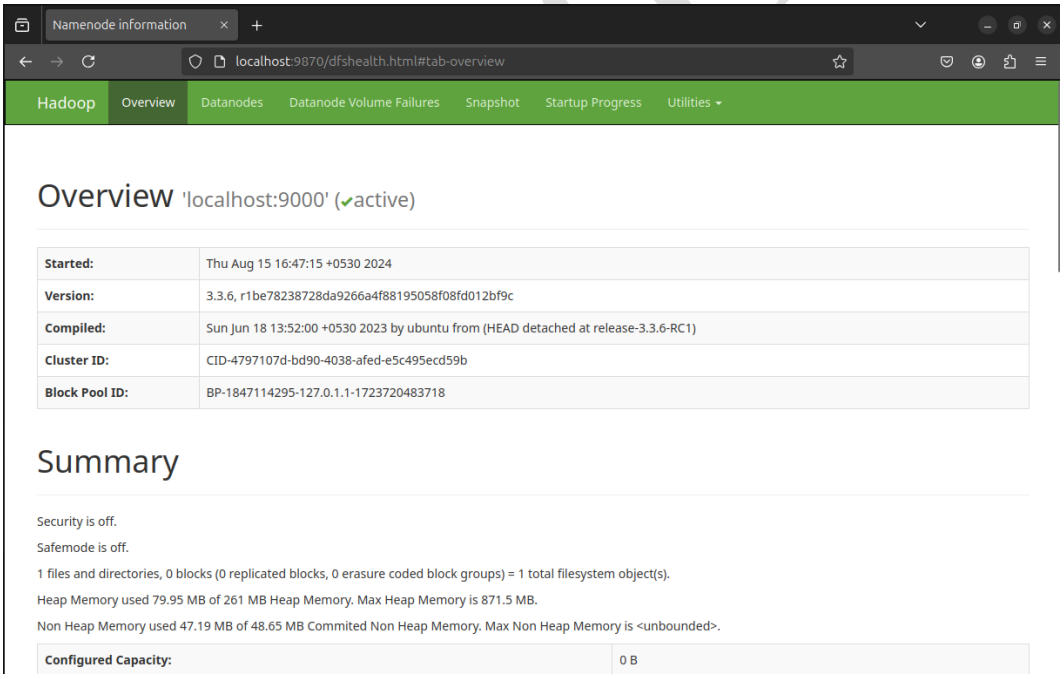
**$ start-all.sh**

```
keerthana@vbox:~/hadoop/etc/hadoop$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as keerthana in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [vbox]
Starting resourcemanager
Starting nodemanagers
keerthana@vbox:~/hadoop/etc/hadoop$
```

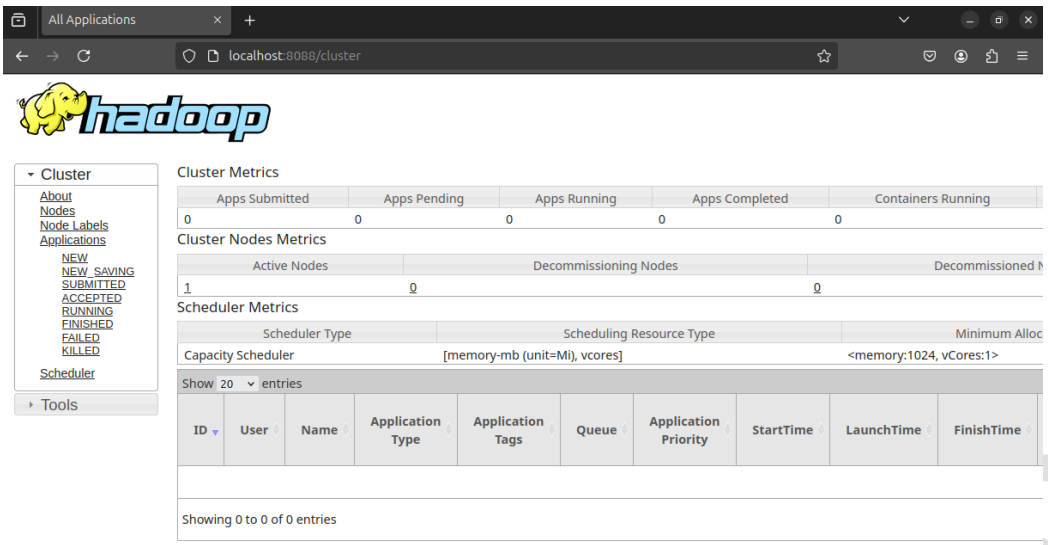
**$ jps**

```
keerthana@vbox:~$ jps
3392 NameNode
4674 Jps
4122 ResourceManager
3565 DataNode
3823 SecondaryNameNode
4255 NodeManager
keerthana@vbox:~$
```

**localhost:9870**



**localhost:8088**




**RESULT:**

Thus, Hadoop has been successfully installed.