

Exp No: 1 Downloading and installing Hadoop, Understanding different Hadoop modes, Startup scripts, Configuration files.

\$ nano ~/.bashrc

```

GNU nano 7.2 /home/karthickragav/.bashrc
#
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk
export HADOOP_HOME=/home/karthickragav/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
#PIG settings
export PIG_HOME=~/pig
export PATH=$PATH:$PIG_HOME/bin
export PIG_CLASSPATH=$PIG_HOME/conf:$HADOOP_INSTALL/etc/hadoop/
export PIG_CONF_DIR=$PIG_HOME/conf
export PIG_CLASSPATH=$PIG_CONF_DIR:$PATH
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_STREAMER=~/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar
export HIVE_HOME=~/hive
export PATH=$PATH:$HIVE_HOME/bin

#set rc

```

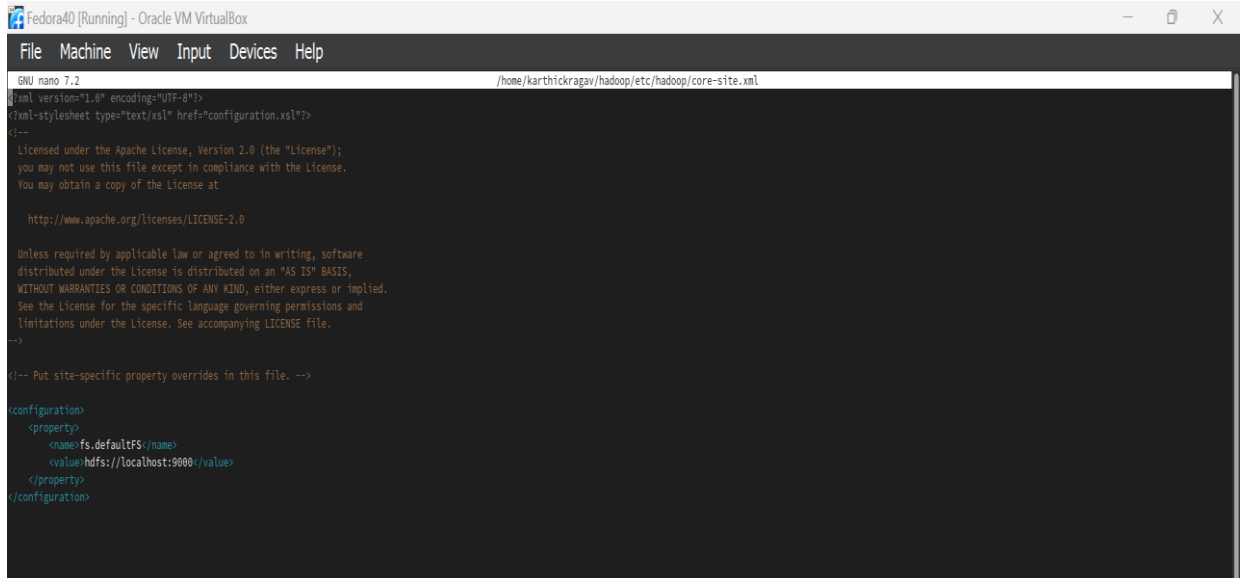
\$ nano \$HADOOP_HOME/etc/hadoop/hadoop-env.sh

```

GNU nano 7.2 /home/karthickragav/hadoop/etc/hadoop/hadoop-env.sh
# SETTINGS HERE WILL BE READ BY ALL HADOOP COMMANDS. THEREFORE,
# ONE CAN USE THIS FILE TO SET YARN, HDFS, AND MAPREDUCE
# CONFIGURATION OPTIONS INSTEAD OF xxx-env.sh.
#
# Precedence rules:
# (yarn-env.sh|hdfs-env.sh) > hadoop-env.sh > hard-coded defaults
# (YARN_xyz|HDFS_xyz) > HADOOP_xyz > hard-coded defaults
#
# Many of the options here are built from the perspective that users
# may want to provide OVERRIDING values on the command line.
# For example:
#
# JAVA_HOME=/usr/java/testing hdfs dfs -ls
#
# Therefore, the vast majority (BUT NOT ALL) of these defaults
# are configured for substitution and not append. If append
# is preferable, modify this file accordingly.
#
# Generic settings for HADOOP
##
#
# Technically, the only required environment variable is JAVA_HOME.
# All others are optional. However, the defaults are probably not
# preferred. Many sites configure these options outside of Hadoop,
# such as in /etc/profile.d
#
# The java implementation to use. By default, this environment
# variable is REQUIRED on all platforms except OS X:
export JAVA_HOME=/usr/lib/jvm/java-1.8.0-openjdk
#
# Location of Hadoop. By default, Hadoop will attempt to determine
# this location based upon its execution path.
export HADOOP_HOME=
#
# Location of Hadoop's configuration information, i.e., where this
# file is living. If this is not defined, Hadoop will attempt to
# locate it based upon its execution path.
#
# NOTE: It is recommend that this variable not be set here but in
# /etc/profile.d or equivalent. Some options (such as
# --config) may react strangely otherwise.
#
export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop
#
# The maximum amount of heap to use (Java -Xmx). If no unit
# is provided, it will be converted to MB. Daemons will
# prefer any line setting in their respective _conf variable.
# There is no default; the JVM will autoscale based upon machine
#

```

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



The screenshot shows a nano editor window titled "Fedora40 [Running] - Oracle VM VirtualBox". The file being edited is "/home/karthickragav/hadoop/etc/hadoop/core-site.xml". The content of the file is as follows:

```
GNU nano 7.2 /home/karthickragav/hadoop/etc/hadoop/core-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<!--
  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

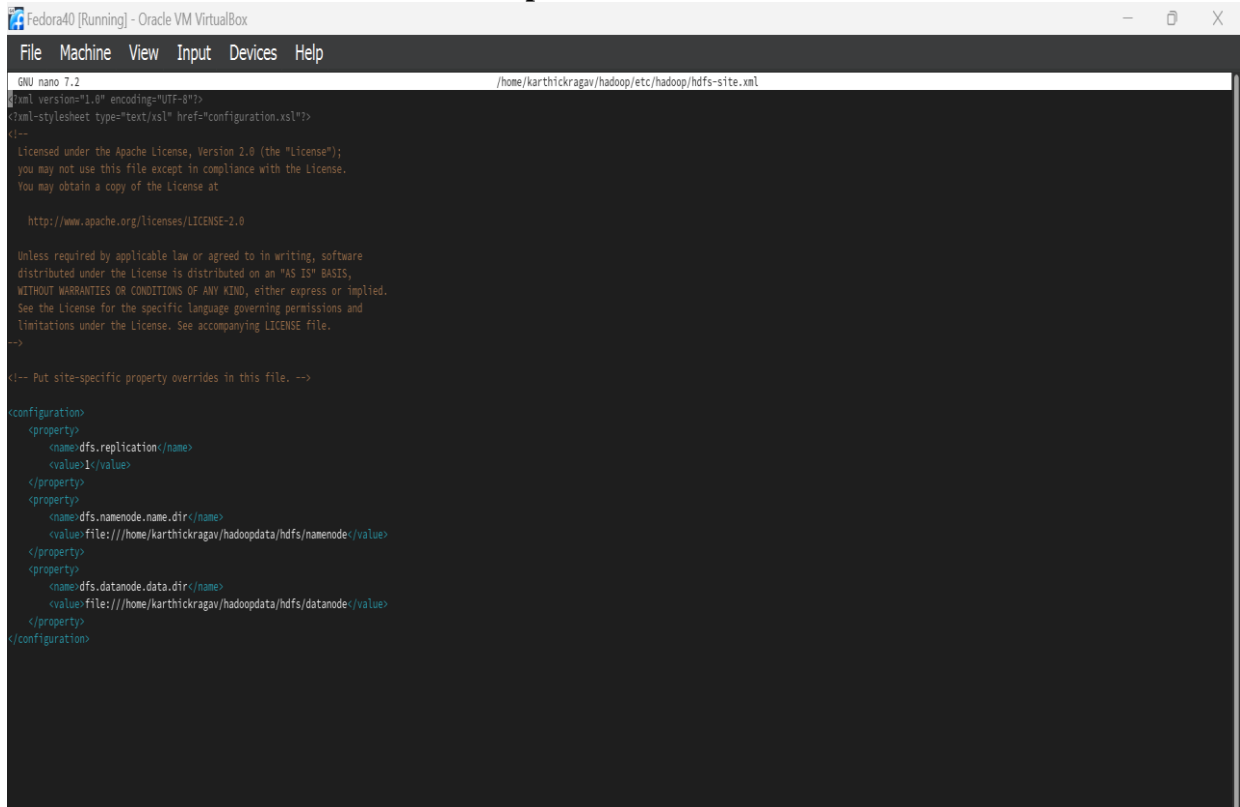
    http://www.apache.org/licenses/LICENSE-2.0

  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.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

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



The screenshot shows a nano editor window titled "Fedora40 [Running] - Oracle VM VirtualBox". The file being edited is "/home/karthickragav/hadoop/etc/hadoop/hdfs-site.xml". The content of the file is as follows:

```
GNU nano 7.2 /home/karthickragav/hadoop/etc/hadoop/hdfs-site.xml
<?xml version="1.0" encoding="UTF-8"?>
<!--
  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.namenode.name.dir</name>
    <value>file:///home/karthickragav/hadoopdata/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///home/karthickragav/hadoopdata/hdfs/datanode</value>
  </property>
</configuration>
```

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

```
GNU nano 7.2 /home/karthickragav/hadoop/etc/hadoop/mapred-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.
-->

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

<configuration>
  <property>
    <name>yarn.app.mapreduce.am.env</name>
    <value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>
  </property>
  <property>
    <name>mapreduce.map.env</name>
    <value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>
  </property>
  <property>
    <name>mapreduce.reduce.env</name>
    <value>HADOOP_MAPRED_HOME=$HADOOP_HOME/home/hadoop/hadoop/bin/hadoop</value>
  </property>
</configuration>
```

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

```
GNU nano 7.2 /home/karthickragav/hadoop/etc/hadoop/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>
</configuration>
```

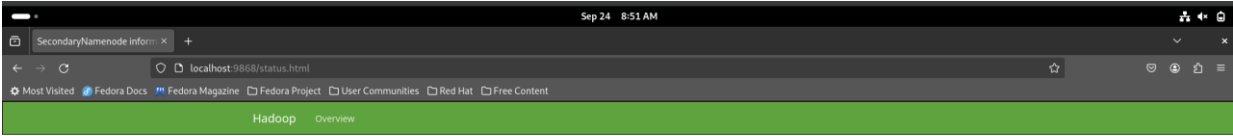
\$ start-all.sh

```
karthickragav@fedora:~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as karthickragav 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 [fedora]
Starting resourcemanager
Starting nodemanagers
karthickragav@fedora:~$
```

\$ jps



localhost:9870



Overview

| | |
|-------------------------|---|
| Version | 3.3.6 |
| Compiled | 2023-06-18T08:22Z by ubuntu from (HEAD detached at release-3.3.6-RC1) |
| NameNode Address | localhost:9000 |
| Started | Tue Sep 24 08:45:49 +0530 2024 |
| Last Checkpoint | Never |
| Checkpoint Period | 3600 seconds |
| Checkpoint Transactions | 1000000 |

Checkpoint Image URI

- file:///tmp/hadoop-karthickragavi/dfs/namesecondary

Checkpoint Editlog URI

- file:///tmp/hadoop-karthickragavi/dfs/namesecondary

Hadoop, 2023.

localhost:8088

