

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

```
$ nano ~/.bashrc
```

```

fedora40 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
GNU nano 2.2 /home/karthickragav/.bashrc Modified
#
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
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_STREAMING=/usr/share/hadoop/tools/lib/hadoop-streaming-3.3.6.jar
export HIVE_HOME=/usr/share/hadoop/hive
export PATH=$PATH:$HIVE_HOME/bin

unset rc

```

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

```
Fedora40 [Running] - Oracle VM VirtualBox
File Machine View Input Devices Help
GNU nano 7.2 /home/karthickragav/hadoop/etc/hadoop/hadoop-env.sh Modified

# 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=

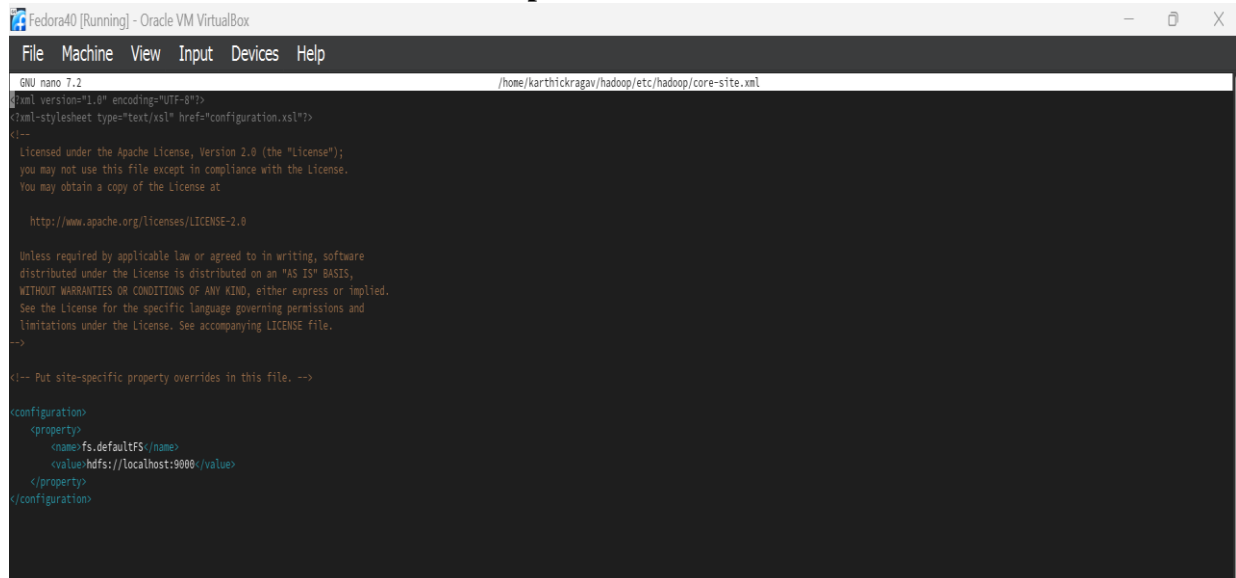
Location of Hadoop's configuration information.  i.e., where this
file is living.  If this is not defined, Hadoop will attempt to
locate it based upon its execution path.

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 Xmx setting in their respective _OPT 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

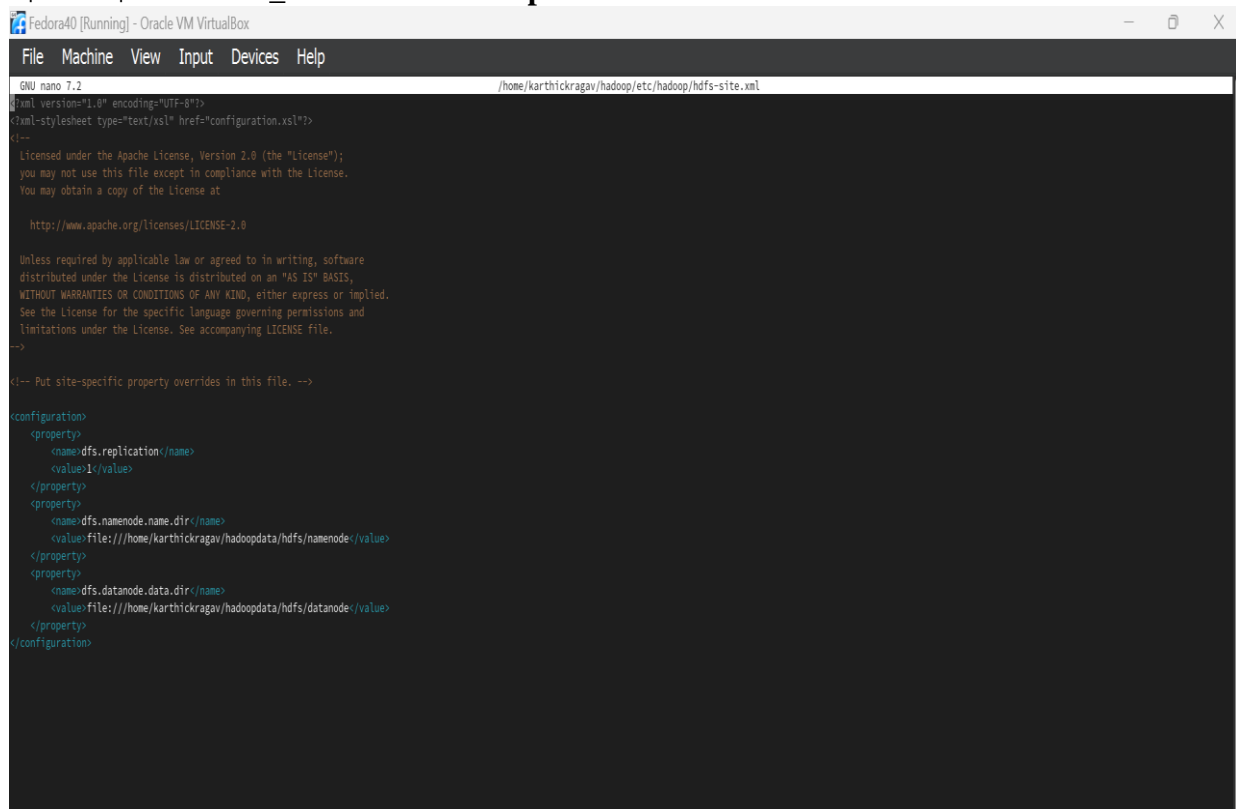
    http://www.apache.org/licenses/LICENSE-2.0

  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>
```

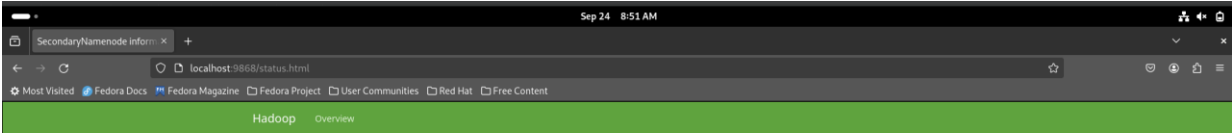
\$ start-all.sh

```
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

Sep 24 8:52 AM

All Applications

localhost:8088/cluster

Most Visited Fedora Docs Fedora Magazine Fedora Project User Communities Red Hat Free Content

All Applications

Cluster

- About
- Nodes
- Node Labels
- Applications

NEW NEW SAVING SUBMITTED ACCEPTED RUNNING FINISHED FAILED KILLED Scheduler Tools

Cluster Metrics

Apps Submitted		Apps Pending		Apps Running		Apps Completed		Containers Running		Used Resources		Total Resources		Reser	
0	0	0	0	0	0	0	0	<memory:0 B, vCores:0>	<memory:0 B, vCores:0>	<memory:8 GB, vCores:8>	<memory:0 B, vCores:0>				

Cluster Nodes Metrics

Active Nodes		Decommissioning Nodes		Decommissioned Nodes		Lost Nodes		Unhealthy Nodes	
1	0	0	0	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type		Scheduling Resource Type		Minimum Allocation		Maximum Allocation	
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0	0	0	0

Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime	State	FinalStatus	Running Containers	Allocated CPU VCores	Allocated Memory MB	Allocated GPUs	Reserved CPU VCores
No data available in table																

Showing 0 to 0 of 0 entries