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

$ nano ~/.bashrc


```
GNU nano 7.2                                                                /home/hadoop/.bashrc
# .bashrc

# Source global definitions
if [ -f /etc/bashrc ]; then
    . /etc/bashrc
fi

# User specific environment
if ! [[ "$PATH" =~ "$HOME/.local/bin:$HOME/bin:" ]]; then
    PATH="$HOME/.local/bin:$HOME/bin:$PATH"
fi
export PATH

# Uncomment the following line if you don't like systemctl's auto-paging feature:
# export SYSTEMD_PAGER=

# User specific aliases and functions
if [ -d ~/.bashrc.d ]; then
    for rc in ~/.bashrc.d/*; do
        if [ -f "$rc" ]; then
            . "$rc"
        fi
    done
fi
#Hadoop Related Options
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-11.0.24.0.8-2.fc40.x86_64
export HADOOP_HOME=/home/hadoop/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 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 PIG_HOME=/home/hadoop/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:$PATH
export HIVE_HOME=/home/jeciyazhini/hive
export PATH=$PATH:$HIVE_HOME/bin
export HADOOP_USER_CLASSPATH_FIRST=true
```

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


```
GNU nano 7.2                                                                hadoop-env.sh
# 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-11-openjdk-11.0.24.0.8-2.fc40.x86_64

# The language environment in which Hadoop runs. Use the English
# environment to ensure that logs are printed as expected.
export LANG=en_US.UTF-8

# 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
# memory size.
# export HADOOP_HEAPSIZE_MAX=

# The minimum amount of heap to use (Java -Xms).  If no unit
# is provided, it will be converted to MB.  Daemons will
# prefer any Xms setting in their respective _OPT variable.
# There is no default; the JVM will autoscale based upon machine
# memory size.
# export HADOOP_HEAPSIZE_MIN=

# Enable extra debugging of Hadoop's JAAS binding, used to set up
# Kerberos security.
# export HADOOP_JAAS_DEBUG=true

# Extra Java runtime options for all Hadoop commands. We don't support
# IPv6 yet/still, so by default the preference is set to IPv4.
```

$nano $HADOOP_HOME/etc/hadoop/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/hadoop/hadoop/tmp</value>
</property>
</configuration>
```
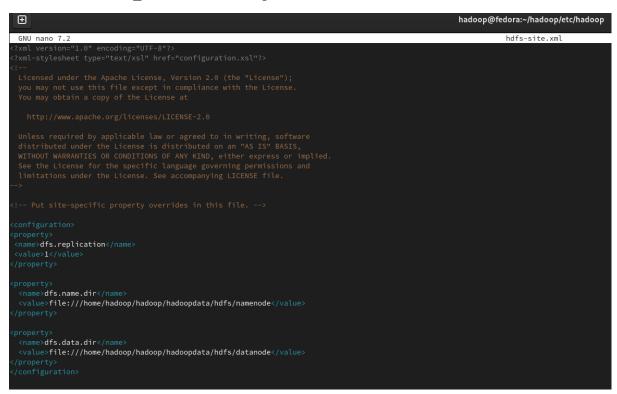
$nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml

```
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/hadoop/hadoop/hadoopdata/hdfs/namenode</value>
</property>

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

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

```
⊞                                                    hadoop@fedora:~/hadoop/etc/hadoop

  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>
```

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

```
⊞                                                    hadoop@fedora:~/hadoop/etc/hadoop

  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>
<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,HADOOP_HOME</value>
</property>
</configuration>
```
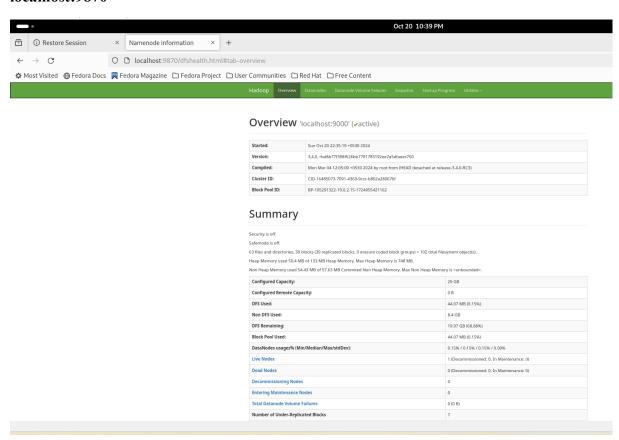
$ start-all.sh

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

$ jps

```
hadoop@fedora:~/hadoop/sbin$ jps
4593 Jps
3764 SecondaryNameNode
3416 DataNode
3210 NameNode
4026 ResourceManager
4175 NodeManager
hadoop@fedora:~/hadoop/sbin$
```

localhost:9870

Oct 20 10:39 PM

Restore Session ✕ | Namenode information ✕ | +

← → C | localhost:9870/dfshealth.html#tab-overview

✿ Most Visited ⊕ Fedora Docs ▣ Fedora Magazine ▢ Fedora Project ▢ User Communities ▢ Red Hat ▢ Free Content

Hadoop   Overview   Datanodes   Datanode Volume Failures   Snapshot   Startup Progress   Utilities ▾

Overview 'localhost:9000' (✓active)

Started:	Sun Oct 20 22:35:19 +0530 2024
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaeec760
Compiled:	Mon Mar 04 12:05:00 +0530 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-16485073-7091-4369-9ccc-b802a280676f
Block Pool ID:	BP-105291322-10.0.2.15-1724955421162

Summary

Security is off.

Safemode is off.

63 files and directories, 39 blocks (39 replicated blocks, 0 erasure coded block groups) = 102 total filesystem object(s).

Heap Memory used 50.4 MB of 133 MB Heap Memory. Max Heap Memory is 748 MB.

Non Heap Memory used 54.43 MB of 57.63 MB Commited Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	29 GB
Configured Remote Capacity:	0 B
DFS Used:	44.07 MB (0.15%)
Non DFS Used:	8.4 GB
DFS Remaining:	19.97 GB (68.88%)
Block Pool Used:	44.07 MB (0.15%)
DataNodes usages% (Min/Median/Max/stdDev):	0.15% / 0.15% / 0.15% / 0.00%
Live Nodes	1 (Decommissioned: 0, In Maintenance: 0)
Dead Nodes	0 (Decommissioned: 0, In Maintenance: 0)
Decommissioning Nodes	0
Entering Maintenance Nodes	0
Total Datanode Volume Failures	0 (0 B)
Number of Under-Replicated Blocks	7

localhost:8088