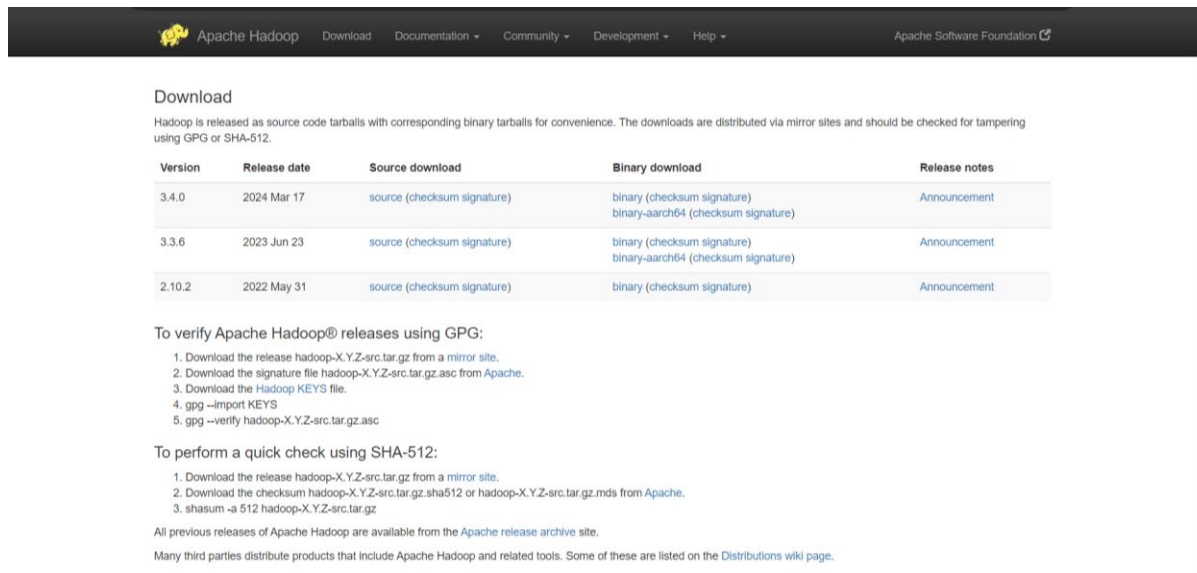


Exp. No : 1

Installation of Hadoop

1. Download Hadoop Binary file



The screenshot shows the Apache Hadoop website's download page. At the top, there is a navigation bar with links for Download, Documentation, Community, Development, and Help. The main heading is "Download". Below it, a paragraph states: "Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512."

Version	Release date	Source download	Binary download	Release notes
3.4.0	2024 Mar 17	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement

To verify Apache Hadoop® releases using GPG:

1. Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
2. Download the signature file `hadoop-X.Y.Z-src.tar.gz.asc` from Apache.
3. Download the [Hadoop KEYS](#) file.
4. `gpg --import KEYS`
5. `gpg --verify hadoop-X.Y.Z-src.tar.gz.asc`

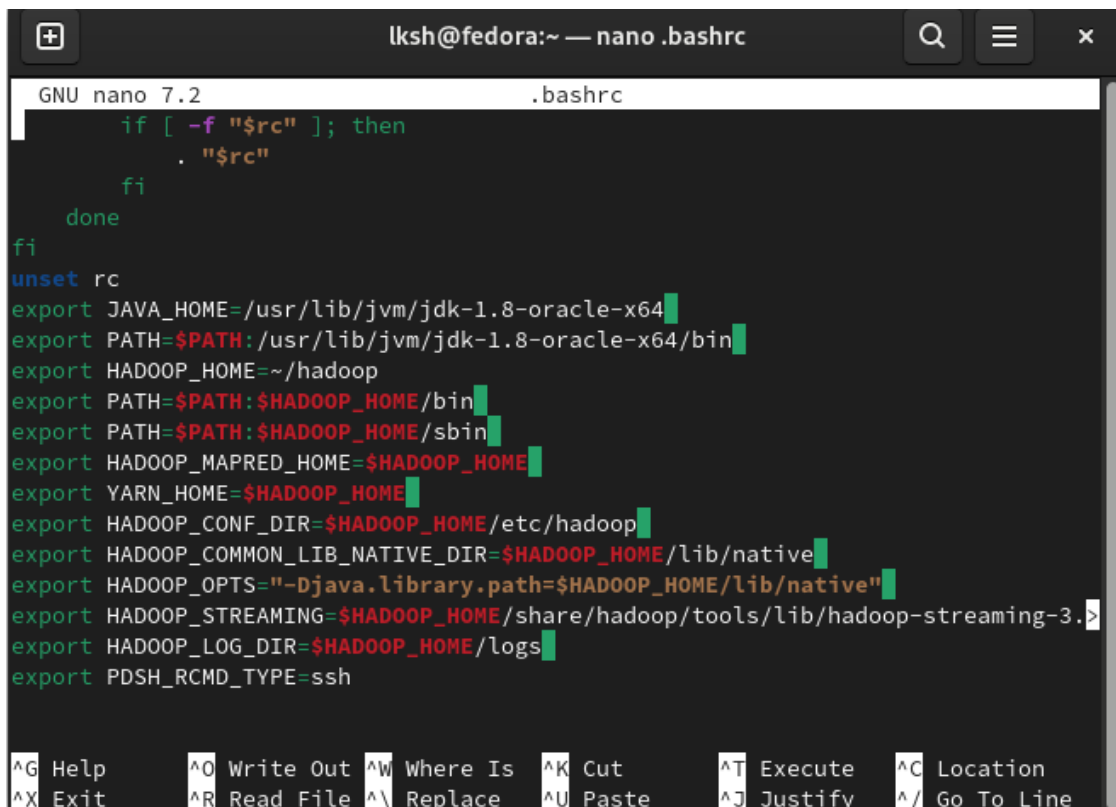
To perform a quick check using SHA-512:

1. Download the release `hadoop-X.Y.Z-src.tar.gz` from a [mirror site](#).
2. Download the checksum `hadoop-X.Y.Z-src.tar.gz.sha512` or `hadoop-X.Y.Z-src.tar.gz.mds` from Apache.
3. `shasum -a 512 hadoop-X.Y.Z-src.tar.gz`

All previous releases of Apache Hadoop are available from the [Apache release archive](#) site.

Many third parties distribute products that include Apache Hadoop and related tools. Some of these are listed on the [Distributions wiki](#) page.

2. Modifying .bashrc file

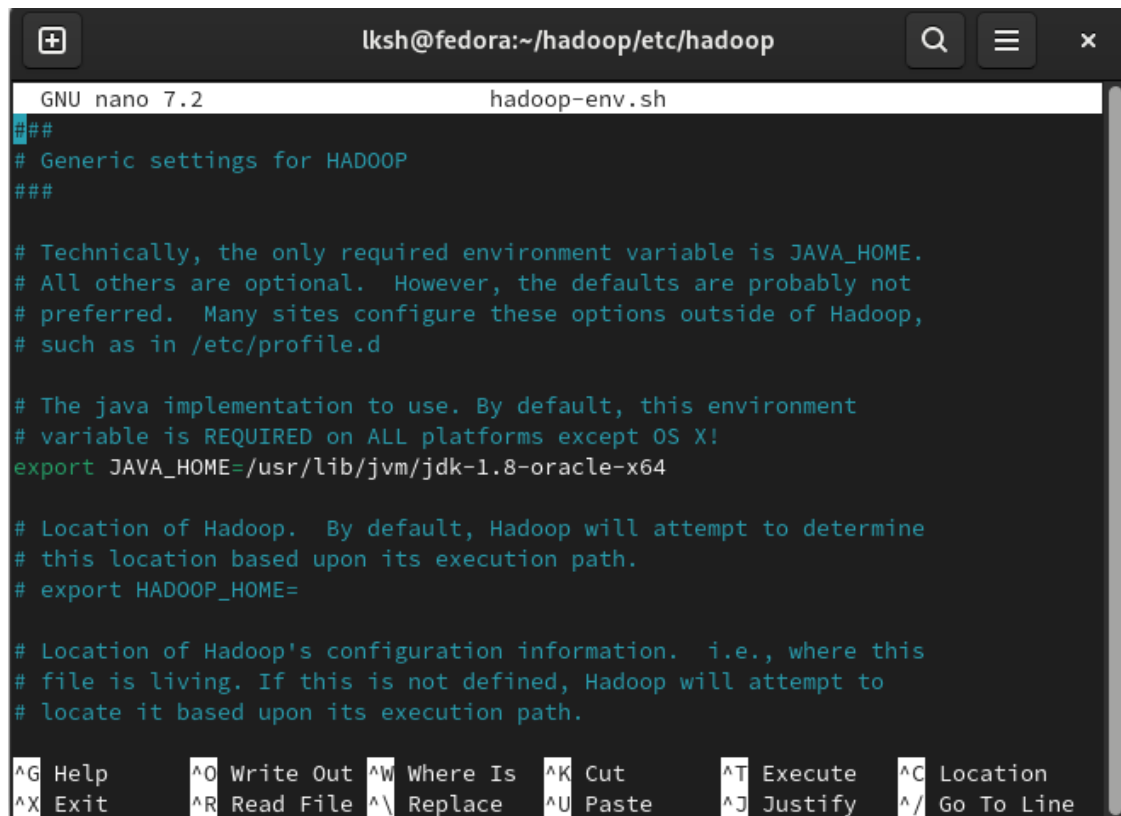


The screenshot shows a terminal window with the nano text editor editing the `.bashrc` file. The terminal title is `lksh@fedora:~ — nano .bashrc`. The editor shows the following content:

```
GNU nano 7.2 .bashrc
if [ -f "$rc" ]; then
    . "$rc"
fi
done
fi
unset rc
export JAVA_HOME=/usr/lib/jvm/jdk-1.8-oracle-x64
export PATH=$PATH:/usr/lib/jvm/jdk-1.8-oracle-x64/bin
export HADOOP_HOME=~/.hadoop
export PATH=$PATH:$HADOOP_HOME/bin
export PATH=$PATH:$HADOOP_HOME/sbin
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
export HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export PDSH_RCMD_TYPE=ssh
```

At the bottom of the terminal, there is a status bar showing various nano editor shortcuts: `^G Help`, `^O Write Out`, `^W Where Is`, `^K Cut`, `^T Execute`, `^C Location`, `^X Exit`, `^R Read File`, `^_ Replace`, `^U Paste`, `^J Justify`, and `^_ Go To Line`.

3. Updating JAVA_HOME in \$HADOOP_HOME/etc/hadoop/hadoop-env.sh



```

lksh@fedora:~/hadoop/etc/hadoop
GNU nano 7.2 hadoop-env.sh
##
# 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/jdk-1.8-oracle-x64

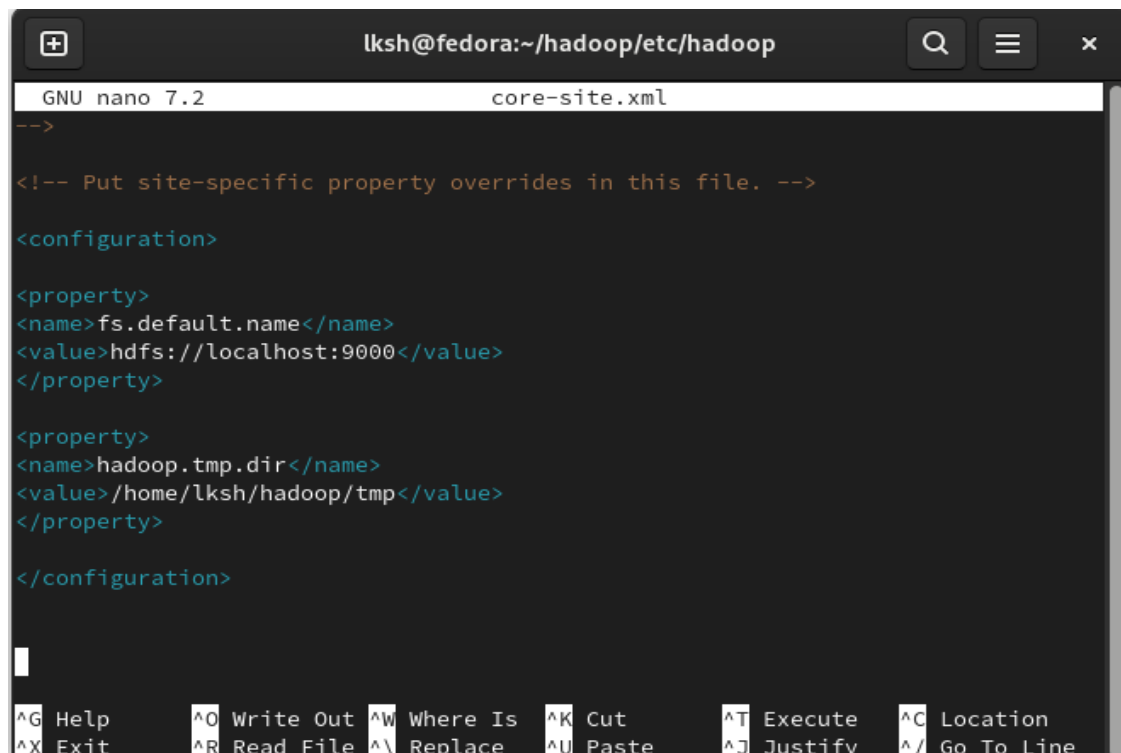
# Location of Hadoop.  By default, Hadoop will attempt to determine
# this location based upon its execution path.
# export HADOOP_HOME=

# 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.

^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^_ Go To Line

```

4. Modifying \$HADOOP_HOME/etc/hadoop/core-site.xml



```

lksh@fedora:~/hadoop/etc/hadoop
GNU nano 7.2 core-site.xml
-->

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

<configuration>

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

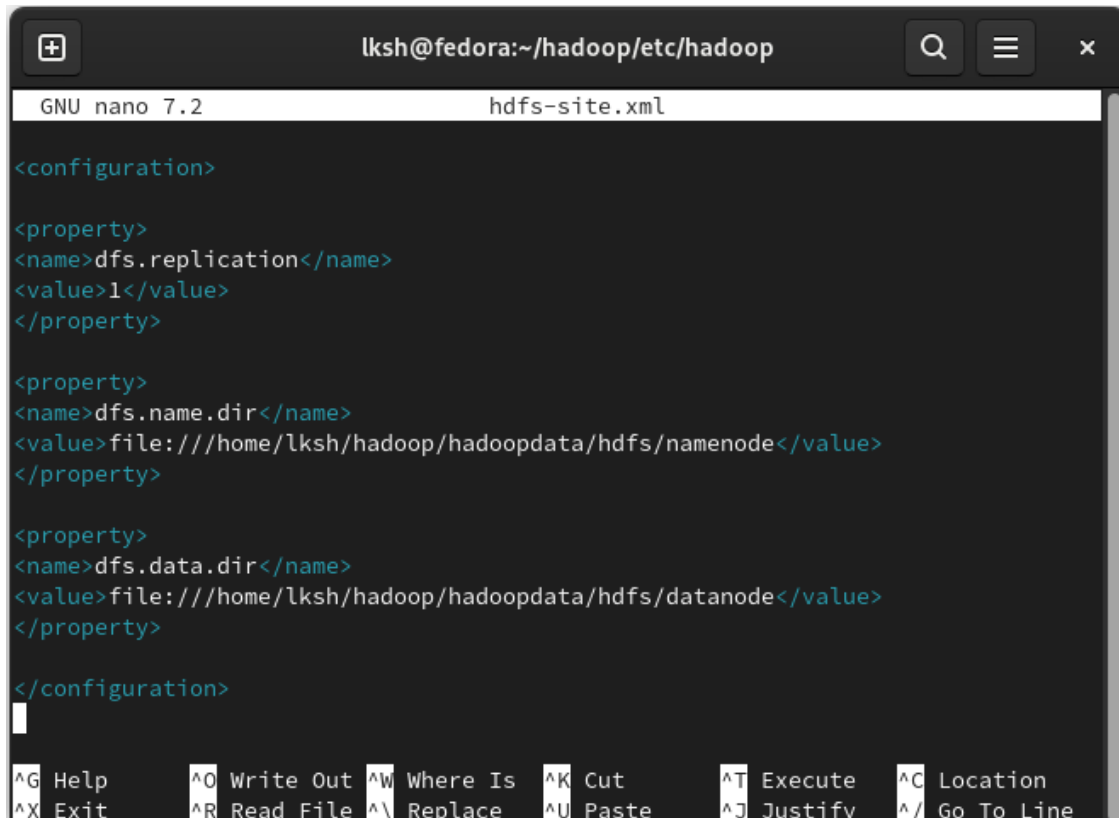
<property>
<name>hadoop.tmp.dir</name>
<value>/home/lksh/hadoop/tmp</value>
</property>

</configuration>

^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^_ Go To Line

```

5. Modifying \$HADOOP_HOME/etc/hadoop/hdfs-site.xml



```

lksh@fedora:~/hadoop/etc/hadoop
GNU nano 7.2 hdfs-site.xml

<configuration>

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

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

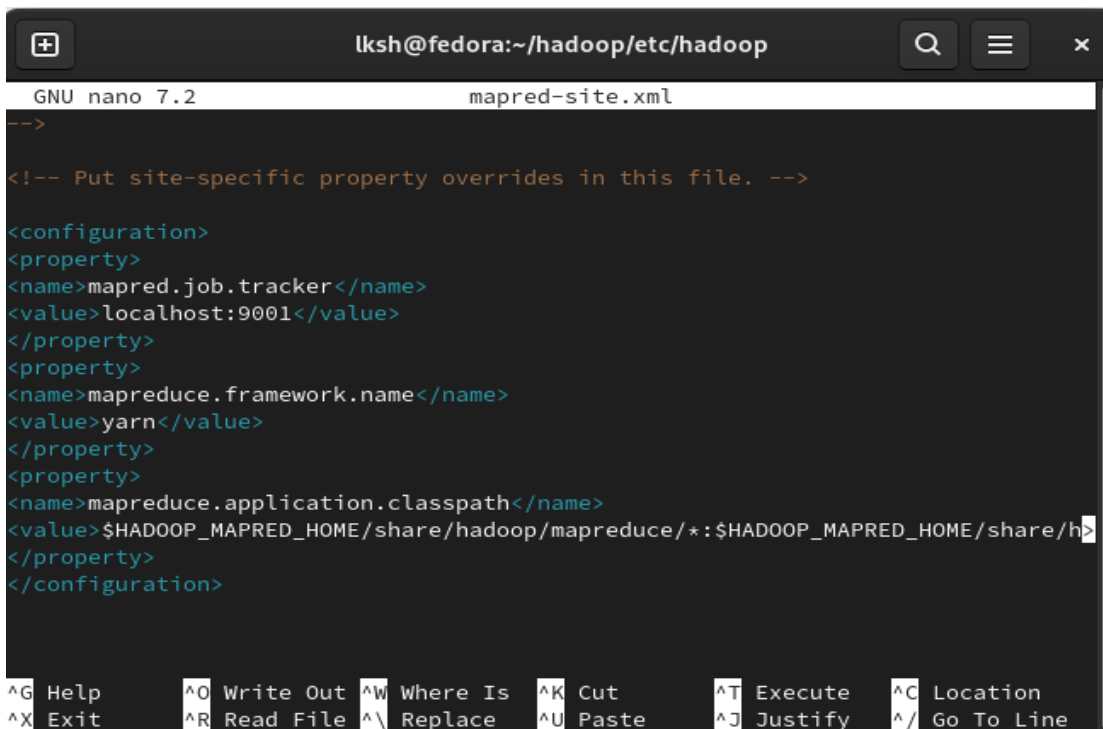
<property>
<name>dfs.data.dir</name>
<value>file:///home/lksh/hadoop/hadoopdata/hdfs/datanode</value>
</property>

</configuration>

```

[^]G Help [^]O Write Out [^]W Where Is [^]K Cut [^]T Execute [^]C Location
[^]X Exit [^]R Read File [^]\ Replace [^]U Paste [^]J Justify [^]/ Go To Line

6. Modifying \$HADOOP_HOME/etc/hadoop/mapred-site.xml



```

lksh@fedora:~/hadoop/etc/hadoop
GNU nano 7.2 mapred-site.xml
-->

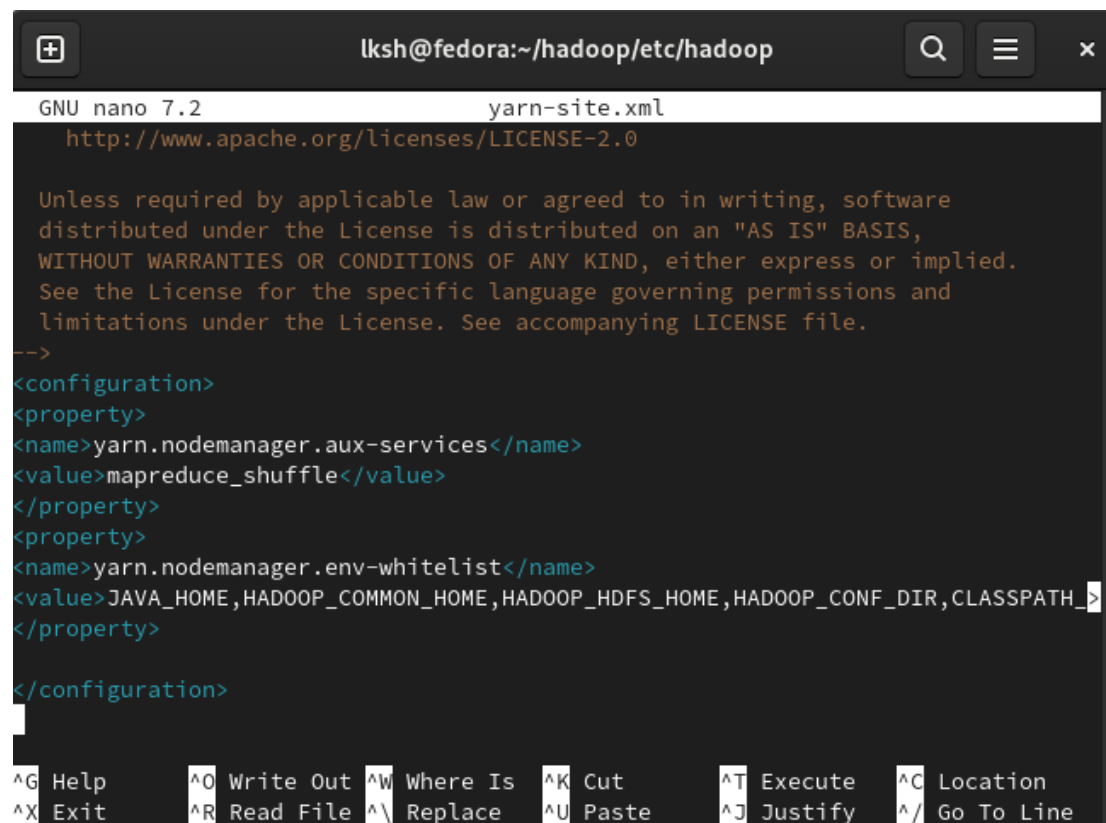
<!-- Put site-specific property overrides in this file. -->

<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/h
</property>
</configuration>

```

[^]G Help [^]O Write Out [^]W Where Is [^]K Cut [^]T Execute [^]C Location
[^]X Exit [^]R Read File [^]\ Replace [^]U Paste [^]J Justify [^]/ Go To Line

7. Modifying \$HADOOP_HOME/etc/hadoop/yarn-site.xml



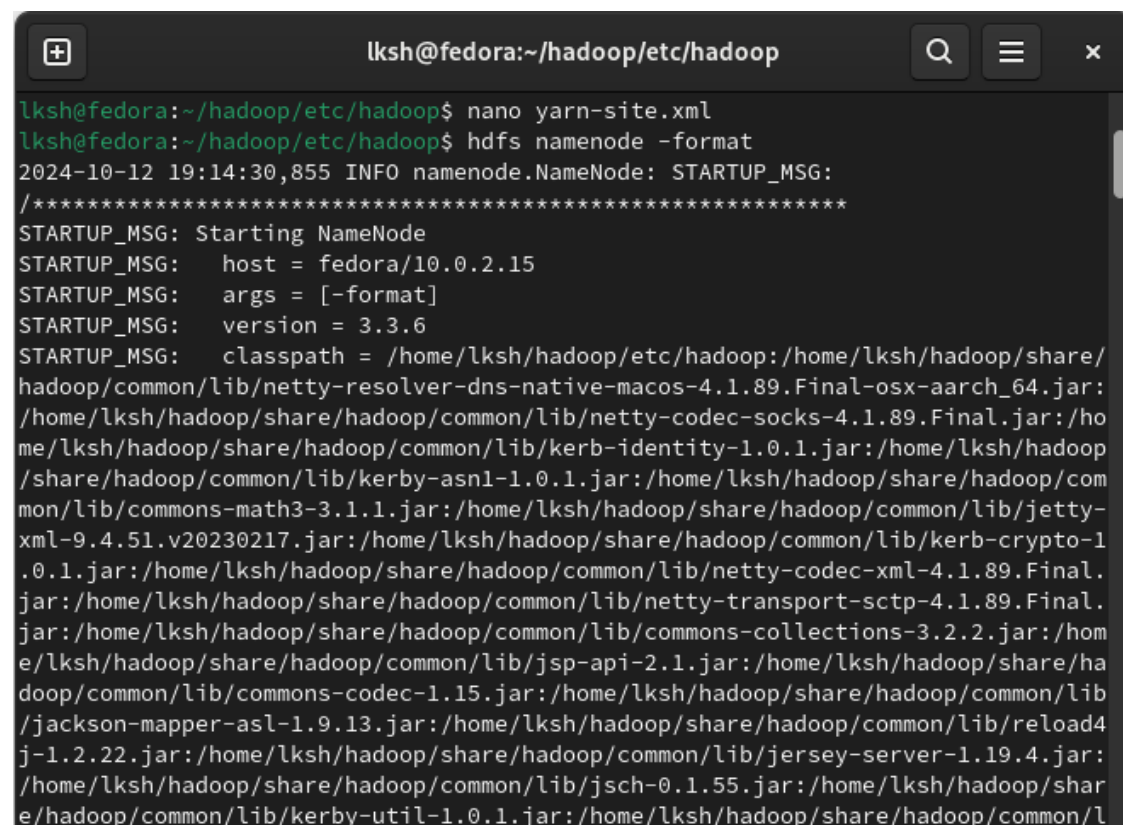
```

lksh@fedora:~/hadoop/etc/hadoop
GNU nano 7.2      yarn-site.xml
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_
</property>
</configuration>

^G Help      ^O Write Out ^W Where Is  ^K Cut       ^T Execute   ^C Location
^X Exit      ^R Read File ^\ Replace   ^U Paste     ^J Justify   ^_ Go To Line
  
```

8. Execute hdfs namenode -format



```

lksh@fedora:~/hadoop/etc/hadoop$ nano yarn-site.xml
lksh@fedora:~/hadoop/etc/hadoop$ hdfs namenode -format
2024-10-12 19:14:30,855 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = fedora/10.0.2.15
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.3.6
STARTUP_MSG:   classpath = /home/lksh/hadoop/etc/hadoop:/home/lksh/hadoop/share/
hadoop/common/lib/netty-resolver-dns-native-macos-4.1.89.Final-osx-aarch_64.jar:
/home/lksh/hadoop/share/hadoop/common/lib/netty-codec-socks-4.1.89.Final.jar:/ho
me/lksh/hadoop/share/hadoop/common/lib/kerb-identity-1.0.1.jar:/home/lksh/hadoop
/share/hadoop/common/lib/kerby-asn1-1.0.1.jar:/home/lksh/hadoop/share/hadoop/com
mon/lib/commons-math3-3.1.1.jar:/home/lksh/hadoop/share/hadoop/common/lib/jetty-
xml-9.4.51.v20230217.jar:/home/lksh/hadoop/share/hadoop/common/lib/kerb-crypto-1
.0.1.jar:/home/lksh/hadoop/share/hadoop/common/lib/netty-codec-xml-4.1.89.Final.
jar:/home/lksh/hadoop/share/hadoop/common/lib/netty-transport-sctp-4.1.89.Final.
jar:/home/lksh/hadoop/share/hadoop/common/lib/commons-collections-3.2.2.jar:/hom
e/lksh/hadoop/share/hadoop/common/lib/jsp-api-2.1.jar:/home/lksh/hadoop/share/ha
dooop/common/lib/commons-codec-1.15.jar:/home/lksh/hadoop/share/hadoop/common/lib
/jackson-mapper-asl-1.9.13.jar:/home/lksh/hadoop/share/hadoop/common/lib/reload4
j-1.2.22.jar:/home/lksh/hadoop/share/hadoop/common/lib/jersey-server-1.19.4.jar:
/home/lksh/hadoop/share/hadoop/common/lib/jsch-0.1.55.jar:/home/lksh/hadoop/shar
e/hadoop/common/lib/kerby-util-1.0.1.jar:/home/lksh/hadoop/share/hadoop/common/l
  
```

```
lksh@fedora:~/hadoop/etc/hadoop
.percentage set to 0. Disabling file IO profiling
2024-10-12 19:38:17,192 INFO blockmanagement.DatanodeManager: dfs.block.invalida
te.limit : configured=1000, counted=60, effected=1000
2024-10-12 19:38:17,195 INFO blockmanagement.DatanodeManager: dfs.namenode.datan
ode.registration.ip-hostname-check=true
2024-10-12 19:38:17,205 INFO blockmanagement.BlockManager: dfs.namenode.startup.
delay.block.deletion.sec is set to 000:00:00:00.000
2024-10-12 19:38:17,209 INFO blockmanagement.BlockManager: The block deletion wi
ll start around 2024 Oct 12 19:38:17
2024-10-12 19:38:17,224 INFO util.GSet: Computing capacity for map BlocksMap
2024-10-12 19:38:17,224 INFO util.GSet: VM type      = 64-bit
2024-10-12 19:38:17,240 INFO util.GSet: 2.0% max memory 475.6 MB = 9.5 MB
2024-10-12 19:38:17,241 INFO util.GSet: capacity     = 2^20 = 1048576 entries
2024-10-12 19:38:17,271 INFO blockmanagement.BlockManager: Storage policy satisf
ier is disabled
2024-10-12 19:38:17,276 INFO blockmanagement.BlockManager: dfs.block.access.toke
n.enable = false
2024-10-12 19:38:17,291 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.
safemode.threshold-pct = 0.999
2024-10-12 19:38:17,291 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.
safemode.min.datanodes = 0
2024-10-12 19:38:17,291 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.
safemode.extension = 30000
2024-10-12 19:38:17,292 INFO blockmanagement.BlockManager: defaultReplication
```

9. Execute start-dfs.sh

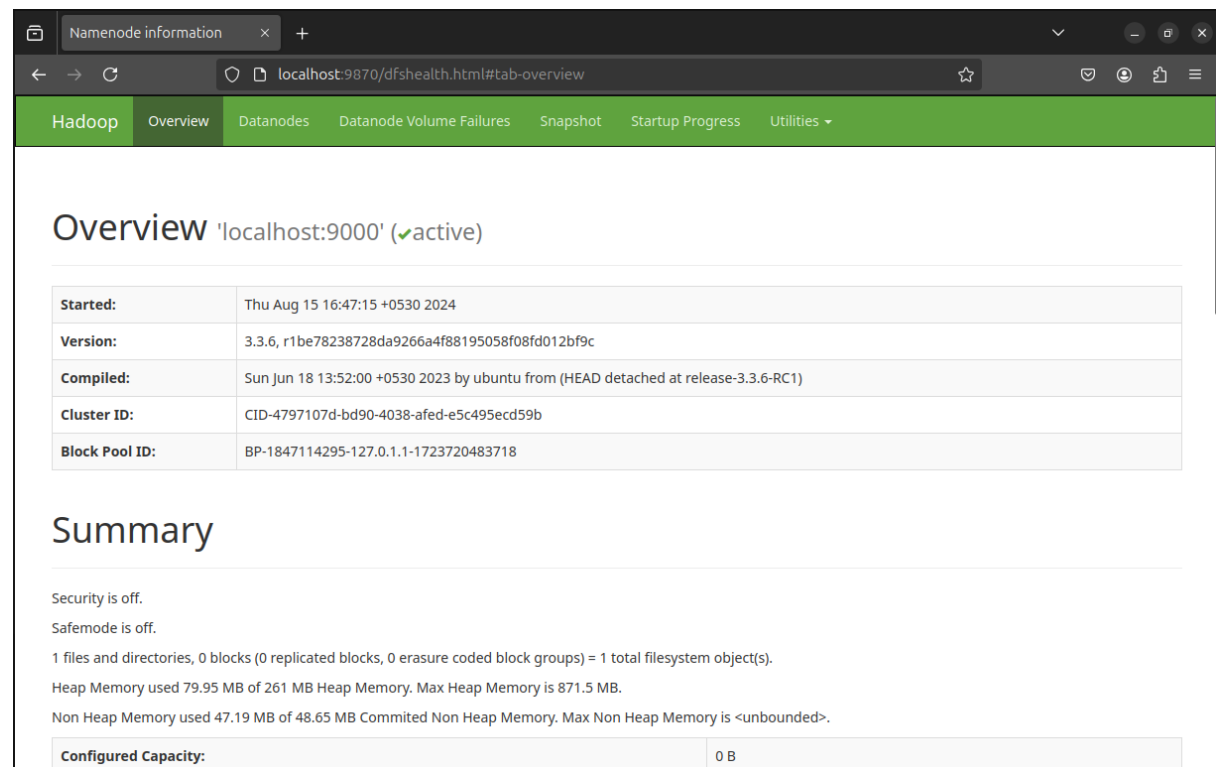
```
lksh@fedora:~/hadoop/etc/hadoop$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [fedora]
```

10. Execute start-yarn.sh

```
lksh@fedora:~/hadoop/etc/hadoop$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

11. Execute jps command

```
lksh@fedora:~/hadoop/etc/hadoop$ jps
3490 DataNode
4163 NodeManager
4599 Jps
3321 NameNode
4027 ResourceManager
3695 SecondaryNameNode
```

12. Open <http://localhost:9870>


Overview 'localhost:9000' (✓active)

Started:	Thu Aug 15 16:47:15 +0530 2024
Version:	3.3.6, r1be78238728da9266a4f88195058f08fd012bf9c
Compiled:	Sun Jun 18 13:52:00 +0530 2023 by ubuntu from (HEAD detached at release-3.3.6-RC1)
Cluster ID:	CID-4797107d-bd90-4038-afed-e5c495ecd59b
Block Pool ID:	BP-1847114295-127.0.1.1-1723720483718

Summary

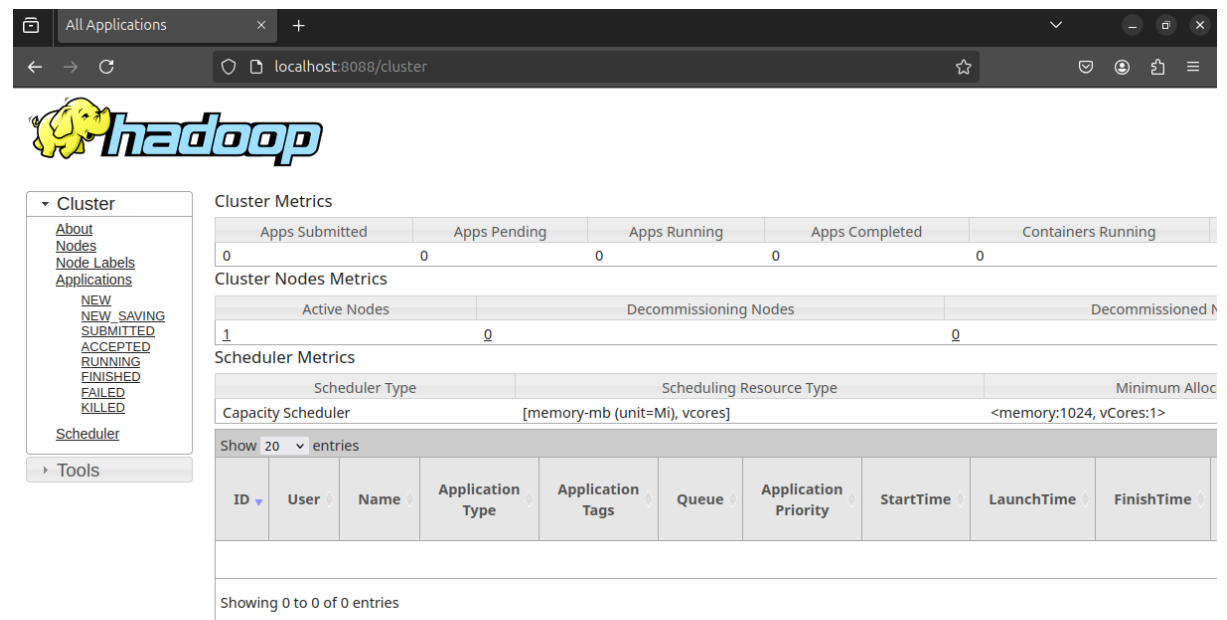
Security is off.
Safemode is off.

1 files and directories, 0 blocks (0 replicated blocks, 0 erasure coded block groups) = 1 total filesystem object(s).

Heap Memory used 79.95 MB of 261 MB Heap Memory. Max Heap Memory is 871.5 MB.

Non Heap Memory used 47.19 MB of 48.65 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	0 B
-----------------------------	-----

13. Open <http://localhost:8088>


Cluster

- About
- Nodes
- Node Labels
- Applications
 - NEW
 - NEW_SAVING
 - SUBMITTED
 - ACCEPTED
 - RUNNING
 - FINISHED
 - FAILED
 - KILLED
- Scheduler

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running
0	0	0	0	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes
1	0	0

Scheduler Metrics

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


Show 20 entries

ID	User	Name	Application Type	Application Tags	Queue	Application Priority	StartTime	LaunchTime	FinishTime
Showing 0 to 0 of 0 entries									

14. Open <http://localhost:8042>

NodeManager information

localhost:8042/node



ResourceManager

NodeManager

- Node Information
- List of Applications
- List of Containers

Tools

NodeManager information

Total Vmem allocated for Containers	16.80 GB
Vmem enforcement enabled	true
Total Pmem allocated for Container	8 GB
Pmem enforcement enabled	true
Total VCores allocated for Containers	8
Resource types	memory-mb (unit=Mj), vcores
NodeHealthyStatus	true
LastNodeHealthTime	Thu Aug 15 16:52:37 IST 2024
NodeHealthReport	
NodeManager started on	Thu Aug 15 16:48:34 IST 2024
NodeManager Version:	3.3.6 from 1be78238728da9266a4f88195058f08fd012bf9c by ubuntu source checksum d42eb795a5eadb0febf5e44a7f87a9 on 2023-06-18T08:31Z
Hadoop Version:	3.3.6 from 1be78238728da9266a4f88195058f08fd012bf9c by ubuntu source checksum 5652179ad55f76cb287d9c633bb53bbd on 2023-06-18T08:22Z