

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

\$ nano ~/.bashrc

```

hadoop@kali: ~
File Actions Edit View Help
GNU nano 7.2 .bashrc

# Alias definitions.
# You may want to put all your additions into a separate file like
# ~/.bash_aliases, instead of adding them here directly.
# See /usr/share/doc/bash-doc/examples in the bash-doc package.

if [ -f ~/.bash_aliases ]; then
    . ~/.bash_aliases
fi

# enable programmable completion features (you don't need to enable
# this, if it's already enabled in /etc/bash.bashrc and /etc/profile
# sources /etc/bash.bashrc).
if ! shopt -oq posix; then
    if [ -f /usr/share/bash-completion/bash_completion ]; then
        . /usr/share/bash-completion/bash_completion
    elif [ -f /etc/bash_completion ]; then
        . /etc/bash_completion
    fi
fi

export JAVA_HOME=/usr/lib/jvm/jdk1.8.0_202
export PATH=$PATH:$JAVA_HOME/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=$HADOOP_HOME/etc/hadoop
export HADOOP_STREAMING=$HADOOP_HOME/share/hadoop/tools/lib/hadoop-streaming-3.4.0.jar
export HADOOP_LOG_DIR=$HADOOP_HOME/logs
export PDSH_RCMD_TYPE=ssh
export HADOOP_COMMON_LIB_NATIVE_DIR=~/.hadoop/lib/native
export HADOOP_OPTS="-Djava.library.path=~/.hadoop/lib/native"

```

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

```

kali-linux-2023.4-vmware-amd64 - VMware Workstation 17 Player (Non-commercial use only)
hadoop@kali: ~
File Actions Edit View Help
GNU nano 7.2 /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh

##
## Precedence rules:
## (yarn-env.sh|hdfs-env.sh) > hadoop-env.sh > hard-coded defaults
## (yarn_etc|HDFS_etc) > HADOOP_etc > 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 -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/jdk1.8.0_202
#
# 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 recommended that this variable not be set here but in
# /etc/profile.d or equivalent. Some options (such as
# --config) may reach strangely otherwise.
# export HADOOP_CONF_DIR=${HADOOP_HOME}/etc/hadoop

```

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

```

File Actions Edit View Help
GNU nano 7.2 /home/hac
<?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.defaultFS</name>
  <value>hdfs://localhost:9000</value> </property>
<property>
  <name>hadoop.proxyuser.dataflair.groups</name> <value>*</value>
  </property>
<property>
  <name>hadoop.proxyuser.dataflair.hosts</name> <value>*</value>
  </property>
<property>
  <name>hadoop.proxyuser.server.hosts</name> <value>*</value>
  </property>
<property>
  <name>hadoop.proxyuser.server.groups</name> <value>*</value>
  </property>
</configuration>

```

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

```

File Actions Edit View Help
GNU nano 7.2 /home/hadoop/hadoop/etc/hadoop
<?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/hadoopdata/hdfs/namenode</value>
</property>
<property>
  <name>dfs.datanode.data.dir</name>
  <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
</property>
</configuration>

```

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

```

hadoop@kali: ~
File Actions Edit View Help
GNU nano 7.2 /home/hadoop/hadoop/etc/hadoop/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>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@kali: ~
File Actions Edit View Help
GNU nano 7.2 /home/hadoop/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>
<property>
<name>yarn.nodemanager.env-whitelist</name>
<value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PREPENDED_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
</property>
</configuration>

```

\$ start-all.sh

```

hadoop@kali: ~
File Actions Edit View Help
(hadoop@kali)-[~]
└─$ 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]
Starting datanodes
Starting secondary namenodes [kali]
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-11 04:59:16,429 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform ... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers

```

\$ jps

```
(hadoop@kali)~$ jps
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
14436 NodeManager
16772 Jps
13830 SecondaryNameNode
14311 ResourceManager
13597 DataNode
13471 NameNode

(hadoop@kali)~$
```

localhost:9870

The screenshot shows a Kali Linux virtual machine running Apache Hadoop 3.3.6. The browser displays the Hadoop Overview page, which includes a table of status information.

Property	Value
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	Wed Aug 14 21:51:32 -0400 2024
Last Checkpoint	Never
Checkpoint Period	3600 seconds
Checkpoint Transactions	1000000

Below the table, there are sections for Checkpoint Image URI and Checkpoint Editlog URI, both showing the file path: `file:///tmp/hadoop-kali/dfs/namesecondary`.

At the bottom, the text "Hadoop, 2023." is displayed.

The screenshot shows a Kali Linux virtual machine with Hadoop 3.3.6 installed. The Hadoop web interface is open in a browser, displaying the 'All Applications' page for a cluster named 'hadoop'. The cluster is in a 'NEW' state. The scheduler metrics show a capacity scheduler with a memory allocation of 1024 MB and 1 vCore.

Cluster Metrics

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

Cluster Nodes Metrics

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

Scheduler Metrics

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

Showing 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
No data available in table															

Showing 0 to 0 of 0 entries