

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

\$ nano ~/.bashrc

```

kamal@kamal-VirtualBox: ~
GNU nano 6.2 /home/kamal/.bashrc
# ~/.bashrc: executed by bash(1) for non-login shells.
# see /usr/share/doc/bash/examples/startup-files (in the package bash-doc)
# for examples

# If not running interactively, don't do anything
case $- in
  *(*) ;;
  *) return;;
esac

# don't put duplicate lines or lines starting with space in the history.
# See bash(1) for more options
HISTCONTROL=ignoreboth

# append to the history file, don't overwrite it
shopt -s histappend

# for setting history length see HISTSIZE and HISTFILESIZE in bash(1)
HISTSIZE=1000
HISTFILESIZE=2000

# check the window size after each command and, if necessary,
# update the values of LINES and COLUMNS.
shopt -s checkwinsize

# If set, the pattern "*" used in a pathname expansion context will
# match all files and zero or more directories and subdirectories.
#shopt -s globstar

# make less more friendly for non-text input files, see lesspipe(1)
[ -x /usr/bin/lesspipe ] && eval "$(SHELL=/bin/sh lesspipe)"

# set variable identifying the chroot you work in (used in the prompt below)
if [ -z "${debian_chroot:-}" ] && [ -r /etc/debian_chroot ]; then
    debian_chroot=$(cat /etc/debian_chroot)
fi

```

\$ nano \$HADOOP\_HOME/etc/hadoop/hadoop-env.sh

```

kamal@kamal-VirtualBox: ~
GNU nano 6.2 /home/kamal/hadoop-3.4.0/etc/hadoop/hadoop-env.sh

# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
# to you 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.

# Set Hadoop-specific environment variables here.

##
## THIS FILE ACTS AS THE MASTER FILE FOR ALL HADOOP PROJECTS.
## 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 OVERWRITING values on the command line.
# For example:

```

## \$nano \$HADOOP\_HOME/etc/hadoop/core-site.xml

```
kamal@kamal-VirtualBox: ~
GNU nano 6.2 /home/kamal/hadoop-3.4.0/etc/hadoop/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.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

```
kamal@kamal-VirtualBox: ~
GNU nano 6.2 /home/kamal/hadoop-3.4.0/etc/hadoop/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/kamal/hadoopdata/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///home/kamal/hadoopdata/hdfs/datanode</value>
  </property>
</configuration>
```

## \$nano \$HADOOP\_HOME/etc/hadoop/mapred-site.xml

```

kamal@kamal-VirtualBox: ~
GNU nano 6.2 /home/kamal/hadoop-3.4.0/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

```

kamal@kamal-VirtualBox: ~
GNU nano 6.2 /home/kamal/hadoop-3.4.0/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>
  <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
  </property>
</configuration>

```

## \$ start-all.sh

```
kamal@kamal-VirtualBox: ~  
kamal@kamal-VirtualBox: $ nano ~/.bashrc  
kamal@kamal-VirtualBox: $ nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh  
kamal@kamal-VirtualBox: $ nano $HADOOP_HOME/etc/hadoop/core-site.xml  
kamal@kamal-VirtualBox: $ nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml  
kamal@kamal-VirtualBox: $ nano $HADOOP_HOME/etc/hadoop/mapred-site.xml  
kamal@kamal-VirtualBox: $ nano $HADOOP_HOME/etc/hadoop/yarn-site.xml  
kamal@kamal-VirtualBox: $ start-all.sh  
WARNING: Attempting to start all Apache Hadoop daemons as kamal 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 [kamal-VirtualBox]  
Starting resourcemanager  
Starting nodemanagers  
kamal@kamal-VirtualBox: $ █
```

## \$ jps

```
kamal@kamal-VirtualBox: $ jps  
8192 NodeManager  
6928 NameNode  
8592 Jps  
7226 DataNode  
8075 ResourceManager  
7660 SecondaryNameNode  
kamal@kamal-VirtualBox: $
```

localhost:9870

Firefox Web Browser Oct 12 13:32

Namenode information x +

localhost:9870/dfshealth.html#tab-overview

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

## Overview 'localhost:9000' (✓active)

Started:	Sat Oct 12 13:30:59 +0530 2024
Version:	3.4.0, rbd8b77f398f626bb7791783192ee7a5dfaee760
Compiled:	Mon Mar 04 12:05:00 +0530 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-32c23d02-25bf-43cb-9ed8-5e8d3c7e42aa
Block Pool ID:	BP-160558784-127.0.1.1-1725263007915

## Summary

Security is off.

Safemode is off.

75 files and directories, 33 blocks (33 replicated blocks, 0 erasure coded block groups) = 108 total filesystem object(s).

Heap Memory used 65.56 MB of 187 MB Heap Memory. Max Heap Memory is 436 MB.


Non Heap Memory used 53.37 MB of 54.78 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	24.34 GB
----------------------	----------

localhost:8088

All Applications

localhost:8088/cluster



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
0	0	0	0	0

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decon
1	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maxi
Capacity Scheduler	[memory-mb (unit=Mi), vcores]	<memory:1024, vCores:1>	<memory:819:

Show 20 entries

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