

HADOOP 3.4.1 STEP BY STEP INSTALLATION ON UBUNTU 24.04.1 LTS

Hardware specification:

OS: Ubuntu 24.04.1 LTS

Memory: 4 GB

CPU: 2 cores

Storage: 25 GB

Prerequisites:

Java Development Kit (JDK)

SSH key pair

Hadoop version 3.4.1

STEP 1: Install JDK

apt install openjdk-8-jdk

```
root@kk:~# apt install openjdk-8-jdk
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  adwaita-icon-theme alsa-topology-conf alsa-ucm-conf at-spi2-common
  fonts-dejavu-extra gsettings-desktop-schemas gtk-update-icon-cache
  libatk-bridge2.0-0t64 libatk-wrapper-java libatk-wrapper-java-jni
root@kk:~# java -version
openjdk version "1.8.0_432"
OpenJDK Runtime Environment (build 1.8.0_432-8u432-ga-us1-0ubuntu2~24.04-ga)
OpenJDK 64-Bit Server VM (build 25.432-bga, mixed mode)
root@kk:~#
```

STEP 2: Generate SSH key pair from “hadoop” user

```
root@kk:~# su hadoop
hadoop@kk:/root$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoop/.ssh/id_rsa):
Created directory '/home/hadoop/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/hadoop/.ssh/id_rsa
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:ezHyiSXLbj4Ikf5/pvZYpCpWYh/UMN6yEA9lI/GVP8 hadoop@kk
The key's randomart image is:
+---[RSA 3072]-----+
|      o.. ..      |
|      . o o .      |
|      . = o .      |
|      . Bo. .      |
|      +SB*+. E|
|      ..O@*=      |
|      o*B=.      |
|      . *oo o+      |
|      o +oo=o.      |
+-----[SHA256]-----+
hadoop@kk:/root$
```

```
hadoop@kk:/$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
hadoop@kk:/$ chmod 640 ~/.ssh/authorized_keys
hadoop@kk:/$
```

STEP 3: Install Hadoop

```
# wget https://dlcdn.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz
```

```
hadoop@kk:~$ wget https://dlcdn.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz
--2025-01-21 11:09:23-- https://dlcdn.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz
Resolving dlcdn.apache.org (dlcdn.apache.org)... 2a04:4e42::644, 151.101.2.132
Connecting to dlcdn.apache.org (dlcdn.apache.org)|2a04:4e42::644|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 974002355 (929M) [application/x-gzip]
Saving to: 'hadoop-3.4.1.tar.gz'

hadoop-3.4.1.tar.gz      100%[=====>] 928.88M  1.15MB/s   in 7m 46s

2025-01-21 11:17:10 (1.99 MB/s) - 'hadoop-3.4.1.tar.gz' saved [974002355/974002355]

hadoop@kk:~$
```

Extract folder and rename

```
# tar -xvzf hadoop-3.4.1.tar.gz
```

```
# mv hadoop-3.4.1 hadoop
```

Setup Hadoop and Java Environment variables

```
# nano ~/.bashrc
```

```
export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
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 HADOOP_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 JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64
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 HADOOP_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"
#####
hadoop@kk:~$
```

Load the configuration in the current environment

```
# source ~/.bashrc
```

```
hadoop@kk:~$ source ~/.bashrc
hadoop@kk:~$
```

STEP 3: Configure Hadoop environment variable file

Edit the `$HADOOP_HOME/etc/hadoop/hadoop-env.sh` file to configure `JAVA_HOME`

```
hadoop@kk:~$ nano $HADOOP_HOME/etc/hadoop/hadoop-env.sh
```

Add `export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64`

```
GNU nano 7.2 /home/hadoop/hadoop/etc/hadoop/hadoop-env.sh *

# 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-8-openjdk-amd64

# The language environment in which Hadoop runs.  Use the English
# environment to ensure that logs are printed as expected.
```

Create directory for **namenode** and **datanode**

```
hadoop@kk:~/hadoop$ mkdir -p ~/hadoopdata/hdfs/{namenode,datanode}
```

Edit the `$HADOOP_HOME/etc/hadoop/core-site.xml` file and update with your system hostname

```
hadoop@kk:~/hadoop$ nano $HADOOP_HOME/etc/hadoop/core-site.xml
```

Add below line

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

```
GNU nano 7.2 /home/hadoop/hadoop/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>
</configuration>
```

Edit the **\$HADOOP_HOME/etc/hadoop/hdfs-site.xml** file and change the namenode and datanode directory paths

```
hadoop@kk:~/hadoop$ nano $HADOOP_HOME/etc/hadoop/hdfs-site.xml
```

Add below line

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.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>
```

```
GNU nano 7.2 /home/hadoop/hadoop/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. -->

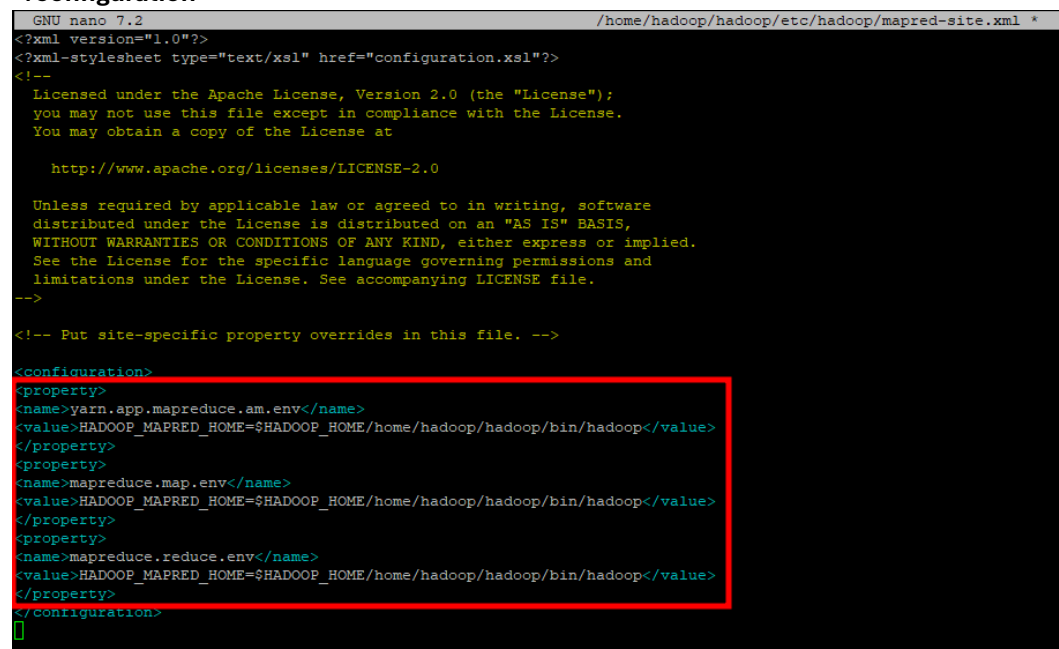
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///home/hadoop/hadoopdata/hdfs/datanode</value>
  </property>
</configuration>
```

Edit the **\$HADOOP_HOME/etc/hadoop/mapred-site.xml** file

```
hadoop@kk:~/hadoop$ nano $HADOOP_HOME/etc/hadoop/mapred-site.xml
```

Add below line

```
<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>
```



```
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>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>
```

Edit the **\$HADOOP_HOME/etc/hadoop/yarn-site.xml** file

```
hadoop@kk:~/hadoop$ nano $HADOOP_HOME/etc/hadoop/yarn-site.xml
```

Add below line

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

```

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>
<property>
<name>yarn.nodemanager.aux-services</name>
<value>mapreduce_shuffle</value>
</property>
</configuration>

```

STEP 4: Start Hadoop Cluster

Format the Hadoop Namenode

```

hadoop@kk:~/hadoop$ hdfs namenode -format
WARNING: /home/hadoop/hadoop/logs does not exist. Creating.
2025-01-21 12:19:55,118 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = kk/127.0.1.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 3.4.1
STARTUP_MSG:   classpath = /home/hadoop/hadoop/etc/hadoop:/home/hadoop/hadoop/

```

Initiate the Hadoop cluster with **start-all.sh**

```

hadoop@kk:~/hadoop$ 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: Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.
Starting datanodes
Starting secondary namenodes [kk]
kk: Warning: Permanently added 'kk' (ED25519) to the list of known hosts.
Starting resourcemanager
Starting nodemanagers
hadoop@kk:~/hadoop$

```

Verify running status of Hadoop services

jps

```

hadoop@kk:~/hadoop$ jps
5605 ResourceManager
5081 NameNode
5434 SecondaryNameNode
5214 DataNode
6063 Jps
5743 NodeManager
hadoop@kk:~/hadoop$

```

STEP 5: Access Dashboard

Cluster

AboutNodesNode LabelsApplicationsNEWNEW SAVINGSUBMITTEDACCEPTEDRUNNINGFINISHEDPAUSEDKILLED

SchedulerTools

Nodes of the cluster

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical VCores Used %
0	0	0	0		<memory:0 B, vCores:0>	<memory:8 GB, vCores:8>	<memory:0 B, vCores:0>	51	0

Cluster Nodes Metrics

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

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority	Scheduler Busy %	RM Dispatcher EventQueue Size	Scheduler Dispatcher EventQueue Size
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0	0	0	0

Show 20 entries

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Allocation Tags	Mem Used	Mem Avail	Phys Mem Used %	VCores Used	VCores Avail	Phys VCores Used %	Version
/default-rack		RUNNING	kk:38791	kk:8042	Tue Jan 21 12:23:46 +0000 2025		0		0 B	8 GB	51	0	8	1	3.4.1

Showing 1 to 1 of 1 entries

DATANODE

HadoopOverviewUtilities

DataNode on kk:9866

Cluster ID:C/D-8ab0bd97-0f68-43db-80bb-33b3cc2e4f4f

Started:Tue Jan 21 18:09:33 +0545 2025

Version:3.4.1, H4d782530834895833b8f06a083222b79422849b1

Block Pools

Namenode Address	Namenode HA State	Block Pool ID	Actor State	Last Heartbeat Sent	Last Heartbeat Response	Last Block Report	Last Block Report Size (Max Size)
localhost:9000	active	BP-1654681001-127.0.1.1-1737401806112	RUNNING	2s	2s	37 minutes	0 B (128 MB)

Volume Information

Directory	StorageType	Capacity Used	Capacity Left	Capacity Reserved	Reserved Space for Replicas	Blocks
/home/hadoop/hadoopdata/hdfs/datanode	DISK	28 KB	12.28 GB	0 B	0 B	0

Hadoop, 2024.

NAMENODE

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 113.9 MB of 240 MB Heap Memory. Max Heap Memory is 800.5 MB.

Non Heap Memory used 58.4 MB of 59.71 MB Committed Non Heap Memory. Max Non Heap Memory is <unbounded>.

Configured Capacity:	22.89 GB
Configured Remote Capacity:	0 B
DFS Used:	28 KB (0%)
Non DFS Used:	9.3 GB
DFS Remaining:	12.28 GB (54.02%)
Block Pool Used:	28 KB (0%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 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	0
Number of Blocks Pending Deletion (including replicas)	0
Block Deletion Start Time	Tue Jan 21 18:08:30 +0545 2025
Last Checkpoint Time	Tue Jan 21 18:04:56 +0545 2025
Last HA Transition Time	Never
Enabled Erasure Coding Policies	RS-6-3-1024k

STEP 6: Create directories in HDFS filesystem

```
# hdfs dfs -mkdir /kaushal
```

```
hadoop@kk:/$ hdfs dfs -mkdir /kaushal
```





```
hadoop@kk:~/hadoop$ hdfs dfs -ls /  
Found 1 items  
drwxr-xr-x - hadoop supergroup 0 2025-01-21 13:03 /kaushal  
hadoop@kk:~/hadoop$ bin/hdfs dfs -touchz /kaushal/hello.txt  
hadoop@kk:~/hadoop$
```

Browse file


Not secure | 192.168.1.201:9870/explorer.html#/kaushal

Hadoop Overview Datanodes Datanode Volume Failures Snapshot Startup Progress Utilities

Browse Directory

Show entries Search:

<input type="checkbox"/>	Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name	
<input type="checkbox"/>	-rw-r--r--	hadoop	supergroup	0 B	Jan 21 18:58	1	128 MB	hello.txt	

Showing 1 to 1 of 1 entries Previous **1** Next

Hadoop, 2024.

