# Installing Java and configuring environment variables

* **sudo apt install default-jdk default-jre -y**
* **java -version**

daya@daya-Latitude-E7270 2 java -version openjdk version "11.0.21" 2023-10-17

OpenJDK Runtime Environment (build 11.0.21+9-post -Ubuntu- Bubuntu118.04)

OpenJDK 64-Bit Server VM (build

[11.0.21+9-post-Ubuntu-@ubuntu118.04,](mailto:11.0.21%2B9-post-Ubuntu-@ubuntu118.04) mixed mode, sharing)

# Create a user for Hadoop and configure SSH

**>sudo adduser hadoop**

**>sudo usermod -aG sudo hadoop**

**>sudo su -hadoop**

**>sudo apt install openssh-server openssh-client -y**

**>ssh-keygen -t rsa**

**>cat~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys**

**>sudo chmod 640 ~/.ssh/authorized\_keys**

**>ssh localhost**

Welcome to Ubuntu 18.04.6 LTS (GNU/Linux 4.15.0-220-generic x86\_64)

* Documentation: https://help.ubuntu.com
* Management: https://landscape.canonical.com
* Support: https://ubuntu.com/advantage

Expanded Security Maintenance for Infrastructure is enabled.

72 updates can be applied immediately.

17 of these updates are ESM Infra security updates.

2 of these updates are ESM Apps security updates.

To see these additional updates run: apt list --upgradable

New release 20.04.5 LTS' available.

Run 'do-release-upgrade to upgrade to it.

Last login: Sat Dec 9 16:42:50 2023 from 127.0.0.1

# Download and install Apache Hadoop on Ubuntu

* **sudo su hadoop**
* **wget** [**https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz**](https://dlcdn.apache.org/hadoop/common/hadoop-3.3.6/hadoop-3.3.6.tar.gz)
* **tar -xvzf hadoop-3.3.4.tar.gz**
* **sudo mv hadoop-3.3.4 /usr/local/hadoop**
* **sudo mkdir /usr/local/hadoop/logs**
* **sudo chown -R hadoop:hadoop /usr/local/hadoop**

# Configuration Of Hadoop

* **sudo nano -/.bashrc**

export HADOOP\_HOME=/usr/local/hadoop export HADOOP INSTALL SHADOOP\_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"

* **source -/.bashrc**

# Configuration Of Java Environment Variables

* **sudo nano $HADOOP\_HOME/etc/hadoop/hadoop-env.sh** export JAVA\_HOME=/usr/lib/jvm/java-11-openjdk-amd64 export HADOOP CLASSPATH SHADOOP\_HOME/lib/\*.jar"
* **cd /usr/local/hadoop/lib**
* **sudo wget** [**https://jcenter.bintray.com/javax/activation/javax.activation-**](https://jcenter.bintray.com/javax/activation/javax.activation-)**api/1.2.0/javax.activation-api-1.2.0.jar**

**>hadoop version**

hadoop@daya-Latitude-E7270:-$ hadoop version Hadoop 3.3.6

Source code repository https://github.com/apache/hadoop.git -r 1be78238728da9266a4f88195058f08fd012bf9c

Compiled by ubuntu on 2023-66-18T08:22Z Compiled on platform linux-x86\_64 Compiled with protoc 3.7.1

From source with checksum 5652179ad55f76cb287d9c633bb53bbd This command was run using

/usr/local/hadoop/share/hadoop/common/hadoop-common-3.3.6.jar

**>sudo nano $HADOOP\_HOME/etc/hadoop/core-site.xml**

<property>

<name>fs.default.name</name>

<value>hdfs://0.0.0.0:9000</value>

<description>The default file system URI</description>

</property>

**>sudo mkdir -p/home/hadoop/hdfs/(namenode, datanode]**

**>sudo chown -R hadoop: hadoop/home/hadoop/hdfs**

**>sudo nano $HADOOP\_HOME/etc/hadoop/hdfs-site.xml**

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

<property>

<name>dfs.name.dir</name>

<value File:///home/hadoop/hdfs/namenode</value>

</property>

<property>

<name>dfs.data.dir</name>

<value>file:///home/hadoop/hdfs/datanode</value>

</property>

**>sudo nano $HADOOP\_HOME/etc/hadoop/mapred-site.xml**

<property>

<name>mapreduce.framework.name</name>

<value>yarn</value>

</property>

**>sudo nano $HADOOP\_HOME/etc/hadoop/yarn-site.xml**

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

**>hdfs namenode -format**

**>start-all.sh**

WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 16 seconds.

WARNING: This is not a recommended production deployment configuration.

WARNING: Use CTRL-C to abort. Starting namenodes on [0.0.0.0] Starting datanodes

Starting secondary namenodes [daya-Latitude-E7270] Starting resourcemanager

Starting nodemanagers

* **hadoop@Desktop-g5dIsio:-$ start-all.sh**

WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.

WARNING: Use CTRL-C to abort,

WARNING : This is not a recommended production deployment configuration.

Starting namenodes on [0.0.0.0]

Starting secondary namenodes [daya-Latitude-E7270] Starting datanodes

Starting resourcemanager Starting nodemanagers

**>hadoop@Desktop-g5dfsio:-$jps**

22705 SecondaryNameNode

22370 DataNode

22979 Resource Manager

22154 NameNode

23196 NodeManager 24063.Jps

**>hadoop@Desktop-g5dfsio: $ hdfs dfs -ls/**

Found 1 items

Drwxr-xr-x -hadoop supergroup 0 2023-12-09 16:52/hello

**>hadoop@Desktop-g5dfsio: 5 hdfs dfs -mkdir /sample**

**>hadoop@Desktop-g5dfsio: $ hdfs dfs-ls/**

Found 2 items

drwxr-xr-x -hadoop supergroup 0 2023-12-09 16:59/sample

drwxr-xr-x -hadoop supergroup 0 2023-12-09 16:52

/hello

**>hadoop@Desktop-g5dfsio:-5 hdfs dfs -touchz/test.txt**

**>hadoop@Desktop-g5dfsio:-$ hdfs dfs-ls/**

Found 3 items

drwxr-xr-x -hadoop supergroup 0 2023-12-09 16:59/sample

drwxr-xr-x -hadoop supergroup 0 2023-12-09 16:52 hell0

-rw-r-- 1 hadoop supergroup 0 2023-12-09 17:01 /test.txt

**>hadoop@Desktop-g5dfsio:-$ hdfs dfs-put/Desktop/copy.txt /hello**

**>hadoop@Desktop-g5dfsio:-5 hdfs dfs-ls/bello**

Found 1 items

-rw-r-- 1 hadoop supergroup 13 2023-12-09 17:06 /hello/copy.txt

**>hadoop@Desktop-g5dfsio:-5 hdfs dfs-cat/hello/copy.txt**

hello

**>hadoop@Desktop-g5dfsior-5 bdfs dis-get /test.txt /Desktop**

**>hadoopt Desktop-g5dfslo: Is/Desktop**

copy.txt test.txt

**>hadoop@Desktop-g5dfsio:-S hdfs dfs -moveFromLocal/Desktop/test2.txt**

**/sample**

**>hadoop@Desktop-g5dfsio:-$ hdfs dis-Is/sample**

Found 1 items

-rw-r--r-- 1 hadoop supergroup 7 2023-12-09 17:14/sample/test2.txt

**>hadoop@Desktop-g5dfsio: 5 hdfs dfs-cp/hello/copy.txt/sample**

**>hadoop@Desktop-g5dfsio:-$ hdfs dfs -ls /sample**

Found 2 items

-rw-r-- 1 hadoop supergroup 13 2023-12-09 17:17/sample/copy.txt

-rw-r-- 1 hadoop supergroup 7 2023-12-09 17:14/sample/test2.txt

**>hadoop@Desktop-g5dfsio: $ hdfs dfs -mv/sample/test2.txt/**

**>hadoop@Desktop-g5dfsio:-$ hdfs dfs -ls/**

Found 4 items

drwxr-xr-x-hadoop supergroup 0 2023-12-09 17:19/sample

drwxr-xr-x-hadoop supergroup 0 2023-12-09 17:06/hello

-rw-r-- 1 hadoop supergroup 0 2023-12-09 17:01/test.txt

-rw-r-- 1 hadoop supergroup 7 2023-12-09 17:14/test2.txt

**>hadoop@Desktop-g5dfsio: $ hdfs dfs-rmr /test2.txt**

rmr: DEPRECATED: Please use ‘-rm -r' instead. Deleted/test2.txt

**>hadoop@Desktop-g5dfsio:-$ hdfs dfs -ls/**

Found 3 items

drwxr-xr-x hadoop supergroup 0 2023-12-09 17:19/sample

drwxr-xr-x -hadoop supergroup 0 2023-12-09 17:06 /hello

-rw-r-- 1 hadoop supergroup 0 2023-12-09 17:01/test.txt

**>hadoop@Desktop-g5dfsio:-5 hdfs dfs-du/**

13 13/sample

13 13 /hello

00 /test.tx

**>hadoop@Desktop-gidfsio:-$ hdfs dis-dus/**

dus: DEPRECATED: Please use 'du-s' instead.

26 26/

**>hadoop@Desktop-g5dfsio:-$ hdfs dis-stat/sample**

2023-12-09 11:49:25

**>hadoop@Desktop-g5dfsio:-5 hdfs dis-setrep -R-w6/test.txt**

Replication 6 set: Mestist Waiting for test.txt... done