Hadoop Installation

Step: 1 Install Java

sudo apt update
sudo apt install openjdk-11-jdk -y

Step 1: Install Hadoop via Homebrew

Update Homebrew: (On macos)

```
brew update
```

---

sudo adduser hadoopuser
sudo usermod -aG sudo hadoopuser
su - hadoopuser

sudo apt install openssh-server -y
ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 600 ~/.ssh/authorized_keys

ssh localhost
If prompted, type "yes"

cd ~
wget https://downloads.apache.org/hadoop/common/hadoop-3.4.1/hadoop-3.4.1.tar.gz
tar -xzvf hadoop-3.4.1.tar.gz

mv hadoop-3.4.1 hadoop

nano ~/.bashrc

```
# Hadoop Variables
export HADOOP_HOME=$HOME/hadoop
export HADOOP_INSTALL=$HADOOP_HOME
export HADOOP_MAPRED_HOME=$HADOOP_HOME
export HADOOP_COMMON_HOME=$HADOOP_HOME
export HADOOP_HDFS_HOME=$HADOOP_HOME
export YARN_HOME=$HADOOP_HOME
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export JAVA_HOME=/usr/lib/jvm/java-11-openjdk-amd64
export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
```

Install Hadoop: (On macos)

```
brew install hadoop
```

Note: As of now, Homebrew installs the latest stable Hadoop (3.4.1 or close). You can confirm with:

```
hadoop version
```

Step 2: Configure Environment Variables

Add to your ~/.zshrc or ~/.bash_profile:

```
# Hadoop
```

```
export HADOOP_HOME=/opt/homebrew/Cellar/hadoop/3.4.1/libexec
export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop
export JAVA_HOME=$(/usr/libexec/java_home)
```

Apply changes:

```
source ~/.zshrc
```

Step 3: Edit Configuration Files

- **1. hadoop-env.sh**

```
nano $HADOOP_CONF_DIR/hadoop-env.sh
```

Add or update

```
export JAVA_HOME=$(/usr/libexec/java_home)
```

- **2. core-site.xml**

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

- **3. hdfs-site.xml**

```xml
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:///opt/homebrew/var/hadoop/hdfs/namenode</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:///opt/homebrew/var/hadoop/hdfs/datanode</value>
  </property>
</configuration>
```

◆ **4. `mapred-site.xml`**

```
cp $HADOOP_CONF_DIR/mapred-site.xml.template
$HADOOP_CONF_DIR/mapred-site.xml
```

Then edit:

```xml
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

◆ 5. **yarn-site.xml**

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

Step 4: Format HDFS

```
hdfs namenode -format
```

▸ Step 5: Start Hadoop Services

```
start-dfs.sh
start-yarn.sh
```

🧪 Step 6: Verify It's Running

Check running Java processes:

```
jps
```

Expected:

```
NameNode
DataNode
```

```
ResourceManager
NodeManager
```

Access UIs:

- **HDFS UI** → http://localhost:9870
- **YARN UI** → http://localhost:8088

📂 Step 7: Test HDFS

```
hdfs dfs -mkdir /user
hdfs dfs -mkdir /user/yourname
echo "Hello Hadoop" > hello.txt
hdfs dfs -put hello.txt /user/yourname/
hdfs dfs -cat /user/yourname/hello.txt
```

🟧 Stop Services

```
stop-yarn.sh
stop-dfs.sh
```

Logs from my Installation:-[Incase of any failure check the below logs and fix the issue]

vi /opt/homebrew/Cellar/hadoop/3.4.1/libexec/etc/hadoop/mapred-site.xml

cat /opt/homebrew/Cellar/hadoop/3.4.1/libexec/etc/hadoop/mapred-site.xml

```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.
```

```
<!-- Put site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

---

cat /opt/homebrew/Cellar/hadoop/3.4.1/libexec/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>
```

---

vi /opt/homebrew/Cellar/hadoop/3.4.1/libexec/etc/hadoop/yarn-site.xml
cat /opt/homebrew/Cellar/hadoop/3.4.1/libexec/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>
</configuration>
```

---

hdfs namenode -format

```
WARNING: /opt/homebrew/var/hadoop does not exist. Creating.
2025-08-04 20:41:10,542 INFO namenode.NameNode: STARTUP_MSG:
```

```
/************************************************************
```

---

start-dfs.sh

Starting namenodes on [localhost]
localhost: ssh: connect to host localhost port 22: Connection refused
Starting datanodes
localhost: ssh: connect to host localhost port 22: Connection refused
Starting secondary namenodes [Dhandapanis-MacBook-Pro.local]
Dhandapanis-MacBook-Pro.local: ssh: connect to host dhandapanis-macbook-pro.local port 22: Connection refused
2025-08-04 20:47:16,591 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

---

sudo systemsetup -setremotelogin on
Password:
setremotelogin: remote login is already On.

---

sudo systemsetup -getremotelogin
Password:
Remote Login: On

---

ssh-keygen -t rsa -P "" -f ~/.ssh/id_rsa

Generating public/private rsa key pair.
Your identification has been saved in /Users/dhandapanidhandapaniyedappalli/.ssh/id_rsa
Your public key has been saved in /Users/dhandapanidhandapaniyedappalli/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:1hOtgo1gsF8QwqSI3L8YTsHfUMttQ84ISzVrTV+6GNo dhandapanidhandapaniyedappalli@Dhandapanis-MacBook-Pro.local
The key's randomart image is:
```
+---[RSA 3072]----+
|o+ o+.+ o  .   |
|+.++..= % ..o   |
|+..+oo.* O.o.   |
|  o+o+++.+o.   |
|  o.oooSE+.    |
| o o .. .  |
| o .      |
|          |
|          |
+----[SHA256]-----+
```

---

cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
chmod 600 ~/.ssh/authorized_keys

---

ssh localhost
The authenticity of host 'localhost (::1)' can't be established.
ED25519 key fingerprint is SHA256:Cm2PogazSAUvoml0qtZs6yVPLCxC6ACeLC5u1X0auHg.
This key is not known by any other names.

Are you sure you want to continue connecting (yes/no/[fingerprint])? yes
Warning: Permanently added 'localhost' (ED25519) to the list of known hosts.
ssh_dispatch_run_fatal: Connection to ::1 port 22: Broken pipe

---

start-yarn.sh

Starting resourcemanager
ERROR: Cannot set priority of resourcemanager process 15920
Starting nodemanagers
localhost: WARNING: /opt/homebrew/Cellar/hadoop/3.4.1/libexec/logs does not exist. Creating.
localhost: ERROR: Cannot set priority of nodemanager process 16021

If your installation is successful then you should get the below message when you execute hadoop —version:-

hadoop version

Hadoop 3.4.1
Source code repository https://github.com/apache/hadoop.git -r 4d7825309348956336b8f06a08322b78422849b1
Compiled by mthakur on 2024-10-09T14:57Z
Compiled on platform linux-x86_64
Compiled with protoc 3.23.4
From source with checksum 7292fe9dba5e2e44e3a9f763fce3e680
This command was run using /opt/homebrew/Cellar/hadoop/3.4.1/libexec/share/hadoop/common/hadoop-common-3.4.1.jar