

Install, Configure and Run Hadoop and HDFS

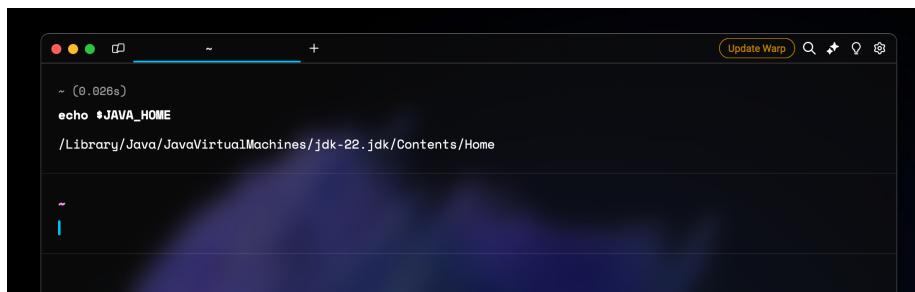
AIM:

To install, configure and run Hadoop and HDFS on the system.

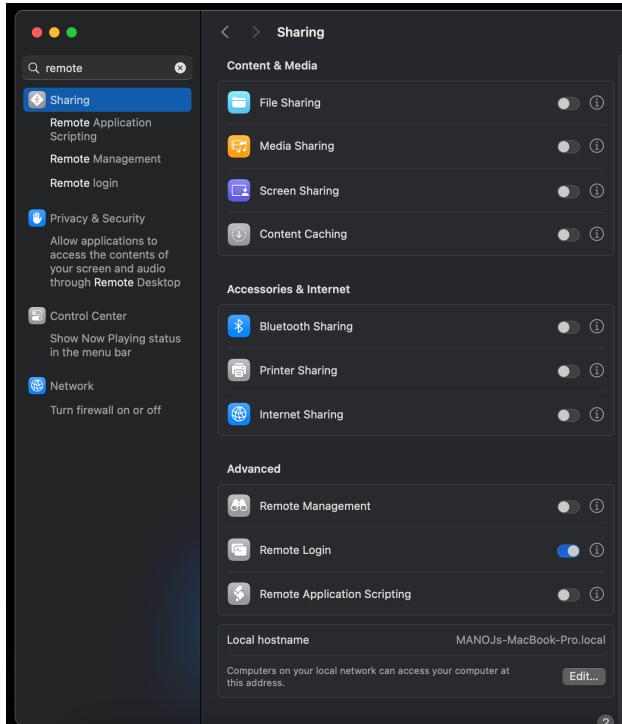
PROCEDURE:

1. Install JDK 8 and set path in the environment variables.
2. Enable SSH for Local Host.
3. Download Hadoop and configure the Hadoop-env.sh file.
4. Set the path of Hadoop in the variables for global access.
5. Set the Java Path inside Hadoop.
6. Configure the core-site.xml, mapred-site.xml, hdfs-site.xml, yarn-site.xml
7. Start Hadoop using start-all.sh command.
8. Open the browser and go to the URL localhost:9870 and verify that Hadoop has been installed correctly.

OUTPUT:



```
~ (0.026s)
echo $JAVA_HOME
/Library/Java/JavaVirtualMachines/jdk-22.jdk/Contents/Home
```



```
~ (3.81s)
ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
/Users/manoj/.ssh/id_rsa already exists.
Overwrite (y/n)? y
Your identification has been saved in /Users/manoj/.ssh/id_rsa
Your public key has been saved in /Users/manoj/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:4oXiteE27vidBGGwpMaeENvAD4cgEAVMA6ynSMU/61ug monoja@MANOJs-MacBook-Pro.local
The key's randomart image is:
+---[RSA 3072]----+
|*++o. +o ..+o.. |
| o o o o+..+= |
|. o + ..+o.o. |
|...* .. * o+.. |
|ooo . S o.o |
|o . o .. |
| + . .. |
| o .. . |
| E o+ |
+---[SHA256]-----+
```

```
~ (0.028s)
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys

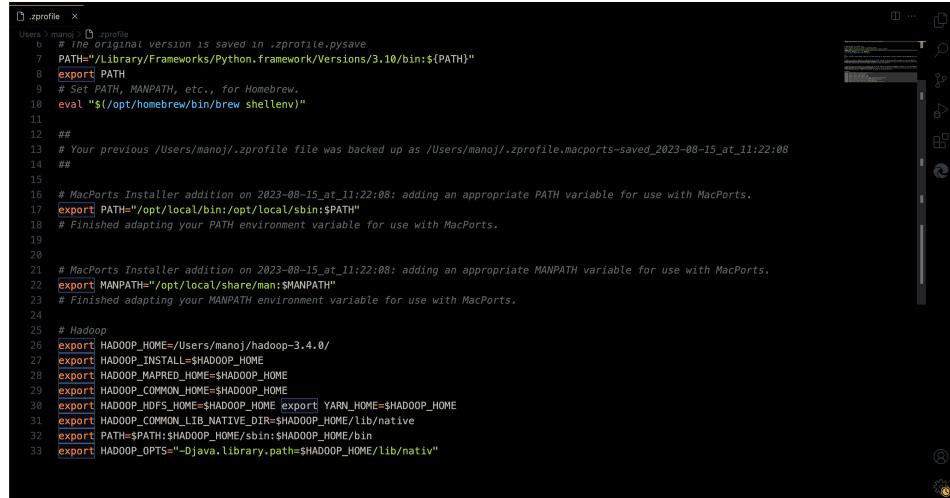
~ (0.021s)
chmod 0600 ~/.ssh/id_rsa.pub

~ (0.396s)
ssh localhost
```

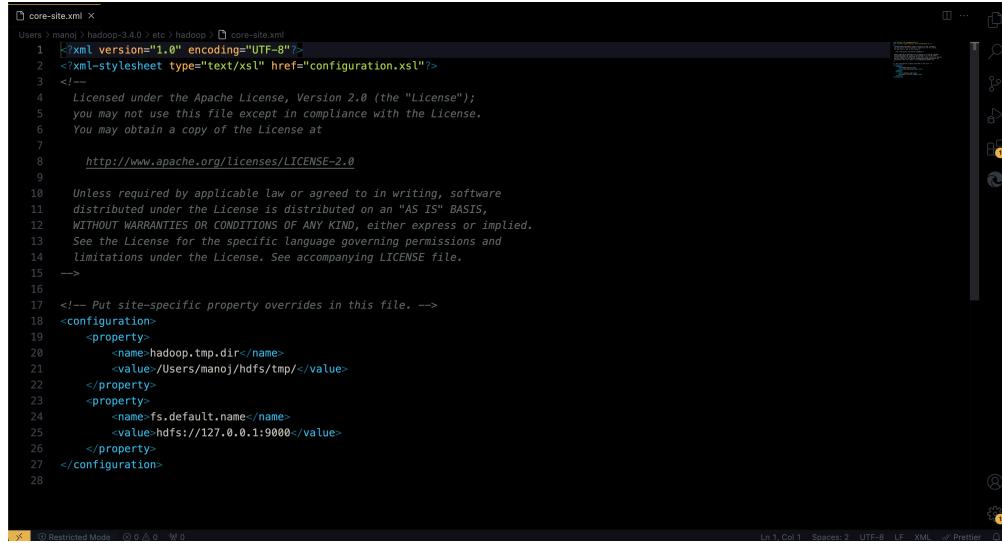
Download

Hadoop is released as source code tarballs with corresponding binary tarballs for convenience. The downloads are distributed via mirror sites and should be checked for tampering using GPG or SHA-512.

Version	Release date	Source download	Binary download	Release notes
3.4.0	2024 Mar 17	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
3.3.6	2023 Jun 23	source (checksum signature)	binary (checksum signature) binary-aarch64 (checksum signature)	Announcement
2.10.2	2022 May 31	source (checksum signature)	binary (checksum signature)	Announcement



```
./.zprofile
Users > manoj > ./.zprofile
6 # The original version is saved in .zprofile.pysave
7 PATH="/Library/Frameworks/Python.framework/Versions/3.10/bin:$PATH"
8 export PATH
9 # Set PATH, MANPATH, etc., for Homebrew,
10 eval "$(/opt/homebrew/bin/brew shellenv)"
11
12 ##
13 # Your previous /Users/manoj/.zprofile file was backed up as /Users/manoj/.zprofile.macports-saved_2023-08-15_at_11:22:00
14 ##
15
16 # MacPorts Installer addition on 2023-08-15_at_11:22:08: adding an appropriate PATH variable for use with MacPorts.
17 export PATH="/opt/local/bin:/opt/local/sbin:$PATH"
18 # Finished adapting your PATH environment variable for use with MacPorts.
19
20
21 # MacPorts Installer addition on 2023-08-15_at_11:22:08: adding an appropriate MANPATH variable for use with MacPorts.
22 export MANPATH="/opt/local/share/man:$MANPATH"
23 # Finished adapting your MANPATH environment variable for use with MacPorts.
24
25 # Hadoop
26 export HADOOP_HOME=/users/manoj/hadoop-3.4.0/
27 export HADOOP_INSTALL=$HADOOP_HOME
28 export HADOOP_MAPRED_HOME=$HADOOP_HOME
29 export HADOOP_COMMON_HOME=$HADOOP_HOME
30 export HADOOP_HDFS_HOME=$HADOOP_HOME export YARN_HOME=$HADOOP_HOME
31 export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
32 export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
33 export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
```



```
core-site.xml
Users > manoj > hadoop-3.4.0 > etc > hadoop > ./core-site.xml
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
3 <!--
4 Licensed under the Apache License, Version 2.0 (the "License");
5 you may not use this file except in compliance with the License.
6 You may obtain a copy of the License at
7
8 http://www.apache.org/licenses/LICENSE-2.0
9
10 Unless required by applicable law or agreed to in writing, software
11 distributed under the License is distributed on an "AS IS" BASIS,
12 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
13 See the License for the specific language governing permissions and
14 limitations under the License. See accompanying LICENSE file.
15 -->
16
17 <!-- Put site-specific property overrides in this file. -->
18 <configuration>
19   <property>
20     <name>hadoop.tmp.dir</name>
21     <value>/Users/manoj/hdfs/tmp</value>
22   </property>
23   <property>
24     <name>fs.default.name</name>
25     <value>hdfs://127.0.0.1:9000</value>
26   </property>
27 </configuration>
28
```

hdfs-site.xml

```
Users > manoj > hadoop-3.4.0 > etc > hadoop > hdfs-site.xml
-->
19   <configuration>
20     <property>
21       <name>dfs.data.dir</name>
22       <value>/Users/manoj/hdfs/namenode</value>
23     </property>
24     <property>
25       <name>dfs.data.dir</name>
26       <value>/Users/manoj/hdfs/datanode</value>
27     </property>
28     <property>
29       <name>dfs.replication</name>
30       <value>1</value>
31     </property>
32   </configuration>
33
```

mapred-site.xml

```
Users > manoj > hadoop-3.4.0 > etc > hadoop > mapred-site.xml
-->
16
17  <!-- Put site-specific property overrides in this file. -->
18  <configuration>
19    <property>
20      <name>mapreduce.framework.name</name>
21      <value>yarn</value>
22    </property>
23  </configuration>
24
```

```
Users > manoj > hadoop-3.4.0 > etc > hadoop > yarn-site.xml

15  <configuration>
16    <property>
17      <name>yarn.nodemanager.aux-services</name>
18      <value>mapreduce_shuffle</value>
19    </property>
20    <property>
21      <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
22      <value>org.apache.hadoop.mapred.ShuffleHandler</value>
23    </property>
24    <property>
25      <name>yarn.resourcemanager.hostname</name>
26      <value>127.0.0.1</value>
27    </property>
28    <property>
29      <name>yarn.acl.enable</name>
30      <value>0</value>
31    </property>
32    <property>
33      <name>yarn.nodemanager.env-whitelist</name>
34      <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRE
35      </value>
36    </property>
37  </configuration>
38
```

```
MANOJs-MacBook-Pro:~
```

```
~ (8.434s)
hdfs namenode -format
2024-08-03 12:03:45,604 INFO util.GSet: capacity      = 2^17 = 131072 entries
Re-format filesystem in Storage Directory root= /Users/manoj/hdfs/tmp/dfs/name; location= null ? (Y or N) y
2024-08-03 12:03:52,517 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1827826417-127.0.0.1-1722668324
93
2024-08-03 12:03:52,518 INFO common.Storage: Will remove files: [/Users/manoj/hdfs/tmp/dfs/name/current/fsimage_0000000000000000, /Users/manoj/hdfs/tmp/dfs/name/current/VERSION, /Users/manoj/hdfs/tmp/dfs/name/current/fsimage_0000000000000000.md5, /Users/manoj/hdfs/tmp/dfs/name/current/seen_txid, /Users/manoj/hdfs/tmp/dfs/name/current edits_inprogress_0000000000000001]
2024-08-03 12:03:52,545 INFO common.Storage: Storage directory /Users/manoj/hdfs/tmp/dfs/name has been successfully formatted.
2024-08-03 12:03:52,564 INFO namenode.FSImageFormatProtobuf: Saving image file /Users/manoj/hdfs/tmp/dfs/name /current/fsimage.ckpt_0000000000000000 using no compression
2024-08-03 12:03:52,626 INFO namenode.FSImageFormatProtobuf: Image file /Users/manoj/hdfs/tmp/dfs/name /current/fsimage.ckpt_0000000000000000 of size 400 bytes saved in 0 seconds .
2024-08-03 12:03:52,635 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2024-08-03 12:03:52,638 INFO namenode.blockmanagement.DatanodeManager: Slow peers collection thread shutdown
2024-08-03 12:03:52,653 INFO namenode.FSNamesystem: Stopping services started for active state
2024-08-03 12:03:52,653 INFO namenode.FSNamesystem: Stopping services started for standby state
2024-08-03 12:03:52,656 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2024-08-03 12:03:52,656 INFO namenode.NameNode: SHUTDOWN_MSG:
*****SHUTDOWN_MSG: Shutting down NameNode at MANOJs-MacBook-Pro.local/127.0.0.1*****
*****
```

The screenshot shows a web browser window with the URL <http://localhost:9876/dfshealth.html#tab-overview>. The page has a green header bar with tabs for Hadoop, Overview, Datanodes, Datanode Volume Failures, Snapshot, Startup Progress, and Utilities. The Overview tab is selected. Below the header is the title "Overview 'localhost:9000' (✓active)". A table provides system information:

Started:	Sat Aug 03 12:05:12 +0530 2024
Version:	3.4.0, rdb8b771398f626bb7791783192ee7a5dfaec760
Compiled:	Mon Mar 04 12:05:00 +0530 2024 by root from (HEAD detached at release-3.4.0-RC3)
Cluster ID:	CID-31d05629-40de-437e-a83c-116679f8ab9c
Block Pool ID:	BP-1827826417-127.0.0.1-1722666832493

The "Summary" section indicates Security is off, SafeMode is off, and lists file and heap memory usage. Below this is another table:

Configured Capacity:	0 B
Configured Remote Capacity:	0 B
DFS Used:	0 B (100%)
Non DFS Used:	0 B

RESULT:

Thus, to install, configure and to run Hadoop and HDFS has been done successfully.