

Vicente De Leon

Big Data Applications

Indiana University

Hadoop Installation Mac M1

Installation source: <https://medium.com/p/ac7c3c5a6ac9>

- My text file regarding Hadoop Installation will also be submitted in final Assignment 4 submission.
- Having JAVA 1.8 for the Intro to Spark - PySpark (Data Science On-Ramp course Summer 2023) was a huge plus.

1.

```
4 A) open terminal and type (to check if Java is installed). I already had this step thanks to the Apache HBase installation from Apache Spark  
coursework: java -version
```

2.

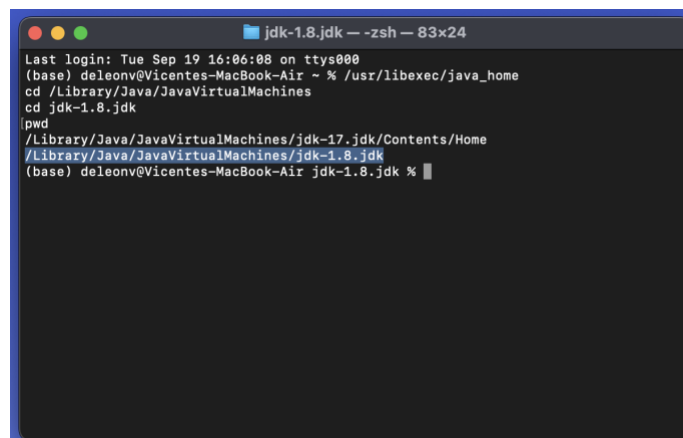
```
6 B) Enable SSG to Local host in the system setting (please follow tutorial):  
7 Run:  
8 cd  
9 pwd  
10 ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa (run this only one time in Home Directory to create security key for SSH)  
11 cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys (exit Home Directory and run this in new terminal)  
12 chmod 0600 ~/.ssh/id_rsa.pub  
13 ssh localhost  
14 exit
```

3. Just like the HBase in PySpark, we must download the tar file as well:

```
16 C) Go to: https://www.apache.org/dyn/closer.cgi/hadoop/common/hadoop-3.2.3/hadoop-3.2.3.tar.gz?source=post_page-----  
ac7c3c5a6ac9----- (download tar file)
```

4.

```
18 D) Configure Hadoop by updating JAVA_HOME Path and adjusting settings in core-site.xml, hdfs-site.xml, mapred-site.xml, and yarn-site.xml  
19 configure environment variables using: .zprofile  
20  
21 D.1) FIRST, open terminal and run the following:  
22 /usr/libexec/java_home  
23 cd /Library/Java/JavaVirtualMachines  
24 cd jdk-1.8.jdk  
25 pwd  
26  
27 and from output, copy the path- Library/Java/JavaVirtualMachines/jdk-1.8.jdk
```



```
jdk-1.8.jdk - zsh - 83x24  
Last login: Tue Sep 19 16:06:08 on ttys000  
(base) deleonv@Vicentes-MacBook-Air ~ % /usr/libexec/java_home  
cd /Library/Java/JavaVirtualMachines  
cd jdk-1.8.jdk  
pwd  
/Library/Java/JavaVirtualMachines/jdk-1.8.jdk/Contents/Home  
(base) deleonv@Vicentes-MacBook-Air jdk-1.8.jdk %
```

5.

```
29 D.2) New terminal window run:
30 type: sudo nano .zprofile
31 type: export JAVA_HOME="/Library/Java/JavaVirtualMachines/jdk-1.8.jdk/Contents/Home"
32 type: control + 0 to save changes and press enter
33 type: control + x to exit
34
35 type to update: source .zprofile
36 Type to check: echo $JAVA_HOME
37
38 D.3) New terminal window:
39 type: sudo nano .zprofile
40 important: my user is deleonv and the haddop version available at the moment was 3.2.3 so I just typed this:
41
42 export HADOOP_HOME=/Users/deleonv/hadoop-3.2.3/ (OR export HADOOP_HOME=~/.Desktop/hadoop-3.2.3/)
43 export HADOOP_INSTALL=$HADOOP_HOME
44 export HADOOP_MAPRED_HOME=$HADOOP_HOME
45 export HADOOP_COMMON_HOME=$HADOOP_HOME
46 export HADOOP_HDFS_HOME=$HADOOP_HOME
47 export YARN_HOME=$HADOOP_HOME
48 export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
49 export PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
50 export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"
51
52 type to update: source .zprofile
```

6.

```
54 D.4)
55 Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: hadoop-env.sh
56 Uncomment line 54 that says: # export JAVA_HOME=
57 add: export JAVA_HOME="/Library/Java/JavaVirtualMachines/jdk-1.8.jdk/Contents/Home"
58
59 D.5)
60 Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: core-site.xml
61 add the following between <configuration> and </configuration> :
62
63 <property>
64 <name>hadoop.tmp.dir</name>
65 <value>/Users/deleonv/hdfs/tmp/</value>
66 </property>
67 <property>
68 <name>fs.default.name</name>
69 <value>hdfs://127.0.0.1:9000</value>
70 </property>
```

```
hadoop-env.sh
Edited

^ ^ # variable is REQUIRED on ALL platforms except OS X! ^
36 #
37 # JAVA_HOME=/usr/java/testing hdfs dfs -ls
38 #
39 # Therefore, the vast majority (BUT NOT ALL!) of these defaults
40 # are configured for substitution and not append. If append
41 # is preferable, modify this file accordingly.
42
43 ###
44 # Generic settings for HADOOP
45 ###
46
47 # Technically, the only required environment variable is JAVA_HOME.
48 # All others are optional. However, the defaults are probably not
49 # preferred. Many sites configure these options outside of Hadoop,
50 # such as in /etc/profile.d
51
52 # The java implementation to use. By default, this environment
53 # variable is REQUIRED on ALL platforms except OS X!
54 export JAVA_HOME="/Library/Java/JavaVirtualMachines/jdk-1.8.jdk/Contents/Home"
55
56 # Location of Hadoop. By default, Hadoop will attempt to determine
57 # this location based upon its execution path.
58 # export HADOOP_HOME=
```

```
core-site.xml
Edited

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
19 <configuration>
20
21 <property>
22 <name>hadoop.tmp.dir</name>
23 <value>/Users/deleov/hdfs/tmp</value>
24 </property>
25 <property>
26 <name>fs.default.name</name>
27 <value>hdfs://127.0.0.1:9000</value>
28 </property>
29 |
30 </configuration>
31
```

7.

```
72 D.6)
73 Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: hdfs-site.xml
74 add the following between <configuration> and </configuration> :
75
76
77 <property>
78 <name>dfs.data.dir</name>
79 <value>/Users/deleov/hdfs/namenode</value>
80 </property>
81 <property>
82 <name>dfs.data.dir</name>
83 <value>/Users/deleov/hdfs/datanode</value>
84 </property>
85 <property>
86 <name>dfs.replication</name>
87 <value>1</value>
88 </property>
89
90 D.7)
91 Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: mapred-site.xml
92 add the following between <configuration> and </configuration> :
93
94 <property>
95 <name>mapreduce.framework.name</name>
96 <value>yarn</value>
97 </property>
```

```
hdfs-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
19 <configuration>
20
21 <property>
22 <name>dfs.data.dir</name>
23 <value>/Users/deleov/hdfs/namenode</value>
24 </property>
25 <property>
26 <name>dfs.data.dir</name>
27 <value>/Users/deleov/hdfs/datanode</value>
28 </property>
29 <property>
30 <name>dfs.replication</name>
31 <value>1</value>
32 </property>
33
34 </configuration>
35
```

```
mapred-site.xml
Edited

1 <?xml version="1.0"?>
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
19 <configuration>
20
21 <property>
22 <name>mapreduce.framework.name</name>
23 <value>yarn</value>
24 </property>
25
26 </configuration>
27
```

8.

```
99 D.B)
100 Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: yarn-site.xml
101 add the following between <configuration> and </configuration> :
102
103 <property>
104 <name>yarn.nodemanager.aux-services</name>
105 <value>mapreduce_shuffle</value>
106 </property>
107 <property>
108 <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
109 <value>org.apache.hadoop.mapred.ShuffleHandler</value>
110 </property>
111 <property>
112 <name>yarn.resourcemanager.hostname</name>
113 <value>127.0.0.1</value>
114 </property>
115 <property>
116 <name>yarn.acl.enable</name>
117 <value>0</value>
118 </property>
119 <property>
120 <name>yarn.nodemanager.env-whitelist</name>
121 <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
122 </property>
123
```

```
yarn-site.xml
XML

1 <?xml version="1.0"?>
2 <!--
3 Licensed under the Apache License, Version 2.0 (the "License");
4 you may not use this file except in compliance with the License.
5 You may obtain a copy of the License at
6
7 http://www.apache.org/licenses/LICENSE-2.0
8
9 Unless required by applicable law or agreed to in writing, software
10 distributed under the License is distributed on an "AS IS" BASIS,
11 WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
12 See the License for the specific language governing permissions and
13 limitations under the License. See accompanying LICENSE file.
14 -->
15
16 <configuration>
17
18 <property>
19 <name>yarn.nodemanager.aux-services</name>
20 <value>mapreduce_shuffle</value>
21 </property>
22 <property>
23 <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
24 <value>org.apache.hadoop.mapred.ShuffleHandler</value>
25 </property>
26 <property>
27 <name>yarn.resourcemanager.hostname</name>
28 <value>127.0.0.1</value>
29 </property>
30 <property>
31 <name>yarn.acl.enable</name>
32 <value>0</value>
33 </property>
34 <property>
35 <name>yarn.nodemanager.env-whitelist</name>
36 <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
37 </property>
38
39 </configuration>
40
```

9.

```
126 E) This WORKS!!! NEW for "It looks like the HDFS Namenode formatting process has started successfully.:"
127 cd ~/Desktop/hadoop-3.2.3/libexec
128 ./hdfs-config.sh
129
130 then type: hdfs namenode -format
131
132
133 F) Starting hadoop services:
134 type: start-all.sh
135
136 G) Got to localhost: http://localhost:9870/dfshealth.html#tab-overview
137
138 H) Stop and close Hadoop
139 type: stop-all.sh
```