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.

```
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.

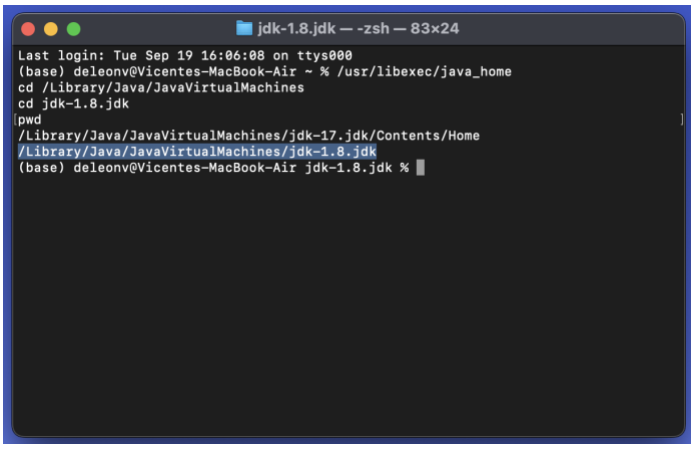
```
B) Enable SSG to Local host in the system setting (please follow tutorial):
    Run:
    cd
    pwd
    ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa (run this only one time in Home Directory to create security key for SSH)
    cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys (exit Home Directory and run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

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

```
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.

```
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
    configure environment variables using: .zprofile

    D.1)FIRST, open terminal and run the following:
    /usr/libexec/java_home
    cd /Library/Java/JavaVirtualMachines
    cd jdk-1.8.jdk
    pwd

    and from output, copy the path- Library/Java/JavaVirtualMachines/jdk-1.8.jdk
```

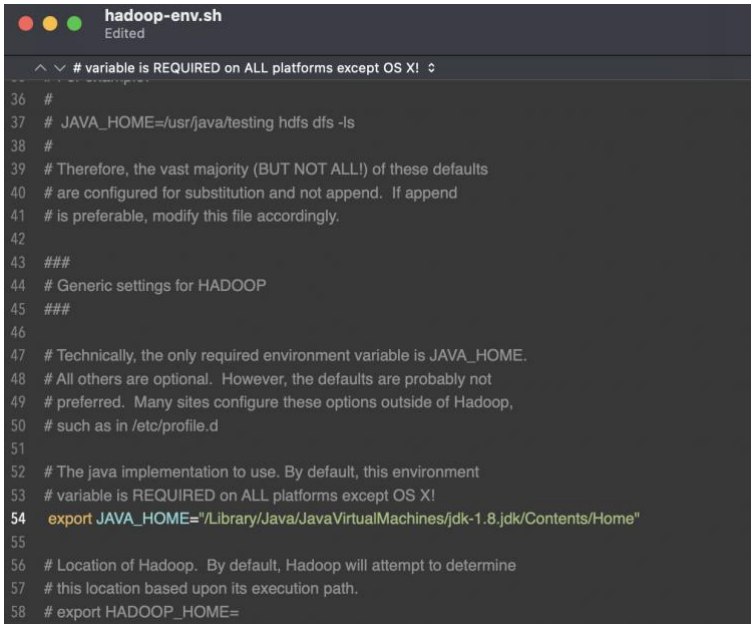


5.

```
D.2)New terminal window run:
type: sudo nano .zprofile
type: export JAVA_HOME="/Library/Java/JavaVirtualMachines/jdk-1.8.jdk/Contents/Home"
type: control + 0 to save changes and press enter
type: control + x to exit

type to update: source .zprofile
Type to check: echo $JAVA_HOME

D.3) New terminal window:
    type: sudo nano .zprofile
    important: my user is deleonv and the haddop version available at the moment was 3.2.3 so I just typed this:

    export HADOOP_HOME=/Users/deleonv/hadoop-3.2.3/  (OR export HADOOP_HOME=~/Desktop/hadoop-3.2.3/)
    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 PATH=$PATH:$HADOOP_HOME/sbin:$HADOOP_HOME/bin
    export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib/native"

    type to update: source .zprofile
```

6.

```
D.4)
    Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: hadoop-env.sh
    Uncomment line 54 that says: # export JAVA_HOME=
    add: export JAVA_HOME="/Library/Java/JavaVirtualMachines/jdk-1.8.jdk/Contents/Home"

D.5)
    Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: core-site.xml
     add the following between <configuration> and </configuration>  :

    <property>
            <name>hadoop.tmp.dir</name>
            <value>/Users/deleonv/hdfs/tmp/</value>
    </property>
    <property>
            <name>fs.default.name</name>
            <value>hdfs://127.0.0.1:9000</value>
    </property>
```



core-site.xml
Edited

```
<?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>hadoop.tmp.dir</name>
      <value>/Users/deleonv/hdfs/tmp/</value>
</property>
<property>
      <name>fs.default.name</name>
      <value>hdfs://127.0.0.1:9000</value>
</property>

</configuration>
```

7.

```
    D.6)
        Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: hdfs-site.xml
        add the following between <configuration> and </configuration>  :



        <property>
        <name>dfs.data.dir</name>
        <value>/Users/deleonv/hdfs/namenode</value>
        </property>
        <property>
        <name>dfs.data.dir</name>
        <value>/Users/deleonv/hdfs/datanode</value>
        </property>
        <property>
        <name>dfs.replication</name>
        <value>1</value>
        </property>

    D.7)
        Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: mapred-site.xml
        add the following between <configuration> and </configuration>  :

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

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.data.dir</name>
<value>/Users/deleonv/hdfs/namenode</value>
</property>
<property>
<name>dfs.data.dir</name>
<value>/Users/deleonv/hdfs/datanode</value>
</property>
<property>
<name>dfs.replication</name>
<value>1</value>
</property>

</configuration>
```

mapred-site.xml
Edited

```
<?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>mapreduce.framework.name</name>
  <value>yarn</value>
</property>

</configuration>
```

8.

```
D.8)
    Go to the hadoop folder within desktop and search for file and open it manually using CotEditor: yarn-site.xml
    add the following between <configuration> and </configuration>  :

    <property>
            <name>yarn.nodemanager.aux-services</name>
            <value>mapreduce_shuffle</value>
        </property>
        <property>
            <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
            <value>org.apache.hadoop.mapred.ShuffleHandler</value>
        </property>
        <property>
            <name>yarn.resourcemanager.hostname</name>
            <value>127.0.0.1</value>
        </property>
        <property>
            <name>yarn.acl.enable</name>
            <value>0</value>
        </property>
        <property>
            <name>yarn.nodemanager.env-whitelist</name>
            <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRE
            D_HOME</value>
        </property>
```

yarn-site.xml 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>
  <property>
    <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
    <value>org.apache.hadoop.mapred.ShuffleHandler</value>
  </property>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>127.0.0.1</value>
  </property>
  <property>
    <name>yarn.acl.enable</name>
    <value>0</value>
  </property>
  <property>
    <name>yarn.nodemanager.env-whitelist</name>
    <value>JAVA_HOME,HADOOP_COMMON_HOME,HADOOP_HDFS_HOME,HADOOP_CONF_DIR,CLASSPATH_PERPEND_DISTCACHE,HADOOP_YARN_HOME,HADOOP_MAPRED_HOME</value>
  </property>

</configuration>
```

9.

```
E) This WORKS!!! NEW for "It looks like the HDFS Namenode formatting process has started successfully.:"
cd ~/Desktop/hadoop-3.2.3/libexec
./hdfs-config.sh

then type: hdfs namenode -format


F) STarting hadoop services:
type: start-all.sh

G) Got to localhost: http://localhost:9870/dfshealth.html#tab-overview

H) Stop and close Hadoop
type: stop-all.sh
```