**Steps to Install Hadoop on MacOS machine-**

**Prerequisites to install Hadoop.**

1. **Install Java and check if Java is installed on your local machine.**

If you have Java already installed, you can verify by running the command “java -version” on the terminal.

If it is not installed, you can install Java from the official Java website.

1. **Install Homebrew (Optional)**

Run the below command in the terminal.

/bin/bash -c "$(curl -fsSL <https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh>)"

**Hadoop Installation**

1. **Download Hadoop from the official release.**

Download the Hadoop tar file from the official website <https://hadoop.apache.org/release.html>. I have downloaded the Hadoop 3.3.0 version. Here are the snapshots of the versions I downloaded.

Webpage link - <https://hadoop.apache.org/release/3.3.0.html>

A screenshot of a computer

Description automatically generated

1. **Extract and Move Hadoop Files**

Once you download the Hadoop you can then extract the downloaded Hadoop tar file to the Hadoop folder in your local machine. To achieve this, you can run the below command in the terminal.

The commands are to be run in the folder where you have downloaded the tar file.

tar -xzvf hadoop-3.3.0.tar.gz

mv hadoop-3.3.0 ~/hadoop

1. **Configure the environment variables.**

Edit the .bash\_profile.sh or .zshrc file in your home directory. These are hidden files and should be able to be listed using the ls -a command.

export HADOOP\_HOME=~/hadoop

export PATH=$PATH:$HADOOP\_HOME/bin

export PATH=$PATH:$HADOOP\_HOME/sbin

1. **Generate the SSH key and set up the SSH environment.**

Run the below command in the terminal home directory

ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa

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

1. **Configure Hadoop**

Navigate to the Hadoop configuration directory and edit **hadoop-env.sh**:

cd ~/hadoop/etc/hadoop

nano hadoop-env.sh

Add or modify the following line to set the Java home:

export JAVA\_HOME=/path/to/your/java/home

1. **Configure Core-Site.xml**

Edit **core-site.xml**: Add the following configuration

**<configuration>**

**<property>**

**<name>fs.defaultFS</name>**

**<value>hdfs://localhost:9000</value>**

**</property>**

**</configuration>**

1. **Configure HDFS-Site.xml**

Edit **hdfs-site.xml**: Add the following configuration

<configuration>

<property>

<name>dfs.replication</name>

<value>1</value>

</property>

</configuration>

1. **Format HDFS:**

Format the Hadoop file system by running the below command –

hdfs namenode -format

1. **Start Hadoop services**

Run the below command on the terminal –

start-dfs.sh

Once you run the below command you may see the name node and data node has been started.

Should you see any issues like port refused connection, it is because of the firewall restricting the ssh protocol and remote login. To enable it- check if you have the ssh installed, if not install the ssh.

Once you install the ssh change the system settings to enable the remote login.

Run the below command –

sudo systemsetup -setremotelogin on

Once you run the above command you can see the error as you need full disk access to enable this. To do that you have to go to system preferences 🡪 privacy 🡪 Full Disk Access🡪 options and then select the terminal application by clicking on the plus sign.

Once done then restart the terminal and run the above command. Once it’s done you can run the start-dfs.sh to check whether the Hadoop has started or not.

Once it is started you can visit the localhost:9870 to check the namenode server details.

To start the YARN run the start-yarn.sh script to start the yarn.

That is it for Hadoop installation and running on Mac.