

Step 1: Setup Prerequisites

1a) Install Homebrew for Easy Installation of Hadoop

Mac Users: (Homepage: <http://brew.sh/>)

```
/usr/bin/ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install/master/install)"
```

Linux: (Homepage: <http://linuxbrew.sh/>)

```
sudo yum install ruby
```

```
ruby -e "$(curl -fsSL https://raw.githubusercontent.com/Linuxbrew/install/master/install)"
```

```
PATH="$HOME/.linuxbrew/bin:$PATH"
```

```
echo 'export PATH="$HOME/.linuxbrew/bin:$PATH"' >> ~/.bash_profile
```

1b) Install wget for easy download of .tar.gz files

```
brew install wget
```

Step 2: User Homebrew to Install Hadoop (Mac or Linux) (204mb)

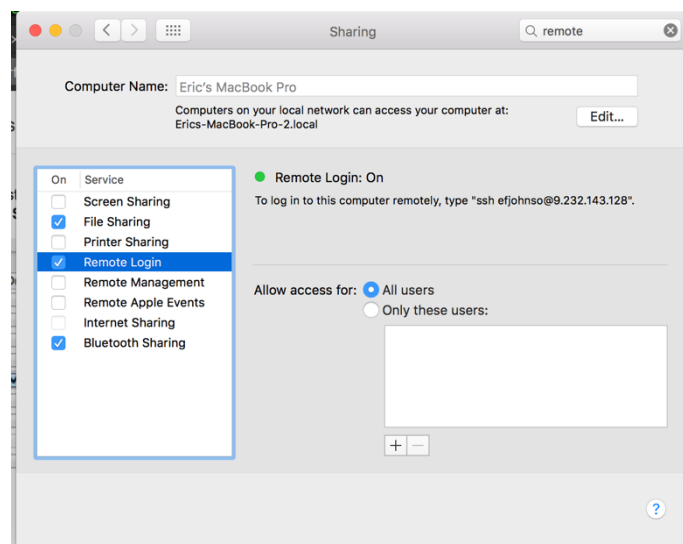
Install Java 1.8: https://www.java.com/en/download/mac_download.jsp

Step 3: Use Homebrew to Install Hadoop (Mac or Linux) (204mb)

1a) Install Hadoop

```
brew install hadoop
```

1b) (Mac Only) - enable "remote sharing" : System Preference > Sharing > Remote Login



Step 4: Configure Hadoop

4a) Edit **hadoop-env.sh**

```
sudo vi /usr/local/Cellar/hadoop/2.7.2/libexec/etc/hadoop/hadoop-env.sh
```

Type: /export HADOOP_OPTS to jump to the line you need

Type: ctrl + c <THEN> SHIFT + I

DELETE the line you are on and replace with:

```
export HADOOP_OPTS="$HADOOP_OPTS -Djava.net.preferIPv4Stack=true -Djava.security.krb5.realm= -  
Djava.security.krb5.kdc="
```

Type: ctrl + c <THEN> SHIFT + : <THEN> wq! ENTER

4b) Edit **hdfs-site.xml**

```
sudo vi /usr/local/Cellar/hadoop/2.7.2/libexec/etc/hadoop/hdfs-site.xml
```

Type: ctrl + c <THEN> SHIFT + I

Between the <configuration></configuration> tags copy and paste the following text:

```
<property>  
<name>dfs.replication</name>  
<value>1</value>  
</property>
```

Type: ctrl + c <THEN> SHIFT + : <THEN> wq! ENTER

4c) Edit **core-site.xml**

```
sudo vi /usr/local/Cellar/hadoop/2.7.2/libexec/etc/hadoop/core-site.xml
```

Type: ctrl + c <THEN> SHIFT + I

Between the <configuration></configuration> tags copy and paste the following text:

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

Type: ctrl + c <THEN> SHIFT + : <THEN> wq! ENTER

4d) Edit **mapred-site.xml**

```
sudo vi /usr/local/Cellar/hadoop/2.7.2/libexec/etc/hadoop/mapred-site.xml
```

Type: ctrl + c <THEN> SHIFT + I

Between the <configuration></configuration> tags copy and paste the following text:

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

Type: ctrl + c <THEN> SHIFT + : <THEN> wq! ENTER

4e) Edit **yarn-site.xml**

```
sudo vi /usr/local/Cellar/hadoop/2.7.2/libexec/etc/hadoop/yarn-site.xml
```

Type: ctrl + c <THEN> SHIFT + I

Between the <configuration></configuration> tags copy and paste the following text:

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

Big Data Analytics

Hadoop Installation Guide

Eric Johnson / Ching-Yung Lin
efj2106@columbia.edu

Type: ctrl + c <THEN> SHIFT + : <THEN> wq! ENTER

4f) Add easy launch words to your "profile"

```
sudo vi ~/.profil
```

Type: ctrl + c <THEN> SHIFT + I

In a blank section copy and paste the two following lines:

```
alias hstart="/usr/local/Cellar/hadoop/2.7.2/sbin/start-dfs.sh;/usr/local/Cellar/hadoop/2.7.2/sbin/start-yarn.sh"
alias hstop="/usr/local/Cellar/hadoop/2.7.2/sbin/stop-yarn.sh;/usr/local/Cellar/hadoop/2.7.2/sbin/stop-dfs.sh"
```

Type: ctrl + c <THEN> SHIFT + : <THEN> wq! ENTER

Step 5: Configure SSH Key for localhost

```
ssh-keygen -t rsa
```

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

Test you can access localhost by typing:

```
ssh localhost
```

```
Eric's-MacBook-Pro-2:2.7.2 efjohnso$ ssh localhost
Last login: Thu Sep 15 15:02:46 2016 from ::1
```

Step 6: Install java-8

Download and install:

https://www.java.com/en/download/mac_download.jsp

Step 7: Setup local variables

```
sudo vi $HOME/.bashrc
```

Type: ctrl + c <THEN> SHIFT + I

COPY AND PASTE THE FOLLOWING:

```
# Set Hadoop-related environment variables
export HADOOP_HOME=/usr/local/Cellar/hadoop/2.7.2

# Set JAVA_HOME (we will also configure JAVA_HOME directly for Hadoop later on)
export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk1.8.0_101.jdk/Contents/Home

# Some convenient aliases and functions for running Hadoop-related commands
unalias fs &> /dev/null
alias fs="hadoop fs"
unalias hls &> /dev/null
alias hls="fs -ls"

# If you have LZOP compression enabled in your Hadoop cluster and
# compress job outputs with LZOP (not covered in this tutorial):
# Conveniently inspect an LZOP compressed file from the command
# line; run via:
#
# $ lzohead /hdfs/path/to/lzop/compressed/file.lzo
#
# Requires installed 'lzop' command.
#
lzohead () {
    hadoop fs -cat $1 | lzop -dc | head -1000 | less
}
```

```
# Add Hadoop bin/ directory to PATH
```

Type: ctrl + c <THEN> SHIFT + : <THEN> wq! ENTER

Then type:

```
export PATH=$PATH:$HADOOP_HOME/bin
```

```
export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk1.8.0_101.jdk/Contents/Home
```

Step 8: Start Hadoop (runs start-dfs.sh and start-yarn.sh)

To Start: hstart

Then if you want to later Stop: hstop

*Check what nodes are running by typing: jps

```
Eric-MacBook-Pro-2:~ efjohnso$ jps
6018  ResourceManager
77524
96580  DataNode
96695  SecondaryNameNode
10343  Jps
6334  NameNode
```

Step 9: Format HDFS and make new directory

```
cd /usr/local/Cellar/hadoop/2.7.2/
```

```
./sbin/start-dfs.sh
```

```
./bin/hdfs namenode -format
```

```
./bin/hdfs dfs -mkdir /user
```

```
./bin/hdfs dfs -mkdir /user/(your computer name - for me it was efjohnso)
```

```
./sbin/start-yarn.sh
```

Check: <http://localhost:8080/>

Big Data Analytics Hadoop Installation Guide

Eric Johnson / Ching-Yung Lin
efj2106@columbia.edu

Other Access Hadoop FileSystem at:

Resource Manager: <http://localhost:50070>

JobTracker: <http://localhost:8088>

Specific Node: Information: <http://localhost:8042>

Step 10: Download Hadoop Examples jar and test

```
cd /usr/local/Cellar/hadoop/2.7.2/
```

```
wget https://www.dropbox.com/s/cyuah7lc31g0x3h/hadoop-mapreduce-examples-2.6.0.jar
```

Example 1: Approximate the value of Pi

Execute: `./bin/hadoop jar hadoop-mapreduce-examples-2.6.0.jar pi 10 100`

Example 2: Solve a Sudoku Puzzle

Create the following file:

```
sudo vi puzz1e1.dta
```

Type: `ctrl + c` <THEN> `SHIFT + I` (COPY AND PASTE FOLLOWING)

```
8 5 ? 3 9 ? ? ? ?
? ? 2 ? ? ? ? ? ?
? ? 6 ? 1 ? ? ? 2
? ? 4 ? ? 3 ? 5 9
? ? 8 9 ? 1 4 ? ?
3 2 ? 4 ? ? 8 ? ?
9 ? ? ? 8 ? 5 ? ?
? ? ? ? ? ? 2 ? ?
? ? ? ? 4 5 ? 7 8
```

Type: `ctrl + c` <THEN> `SHIFT + :` <THEN> `wq!` ENTER

Execute: `./bin/hadoop jar hadoop-mapreduce-examples-2.6.0.jar sudoku puzz1e1.dta`

Example 3: Count the words in Hillary Clinton's Speech

```
sudo vi hillary.txt
```

Type: `ctrl + c` <THEN> `SHIFT + I` (COPY AND PASTE FOLLOWING)

First, it's wonderful to be back in Michigan. You can really feel the energy and dynamism that is driving this state's comeback. And in Detroit, we've got new businesses that are opening. Neighborhoods like Midtown and Eastern Market are coming back. The auto industry just had its best year ever. Over in Ann Arbor, high tech firms are thriving. The next generation of engineers are getting trained up in Houghton. And here at Futuramic – so well named – you are on the front lines of what I believe will be a true manufacturing renaissance in America. And I just was given a short, but exciting, tour by Mark Jurcak and John Couch, who were telling me about how this company was started as, and for most of its early history, was an auto supply company. And then in 2000 as the market began to change and some of the auto companies began to realign, they were faced with a choice. We all face choices in life, don't we? And this company could have just said, 'Hey, you know, our business is not going to be what it was, we've got to just fold up, let's just kind of quit.' But that's not what happened here, and what happened here is what can happen across America. You are in now, what is largely an aerospace company. And because of the work force and the work ethic and the commitment of Futuramic, you are seeing the future unfold. So I got to see what's happening here to help build the sled rocket that is going to go from Macomb to Mars. I saw the two halves of an F-35 nose come waiting to be put together. I talked with some of the workers about the absolute perfection that is required to do this work. And what I believe with all my heart, is that what's happening here can happen in so many places if we put our minds to it. If we support advanced manufacturing. If we are the kind of country that once again understands how important it is the build things. We are builders and we need to get back to building! So we're making progress, none of us can be satisfied until the economic revitalization we're seeing in some parts of Michigan reaches every community. But it is inspiring to see this combination of old-fashioned hard work and cutting-edge innovation. And I know my opponent in this election was here in Michigan about a week ago, and it was like he was in a different place. When he visited Detroit on Monday, he talked only of failure, poverty, and crime. He's missing so much about what makes Michigan great. And the same is true when it comes to our country. He describes America as an embarrassment. He said – and I quote – 'We're becoming a third-world country.' Look around you, my friends. Go visit with the workers building rockets. That doesn't happen in third world countries. Now we have a lot of urgent and important work to do – and that's what I'm going to talk about today – because all the people that I've met in this campaign really prove how wrong this negative, pessimistic view is. America's best days are still ahead of us if we make up our minds to actually go out and make that happen. Just consider our assets: We have the most dynamic, productive workforce in the world, bar none. We have the most innovative businesses. The top colleges, universities, community colleges, training programs in the world. And the best science and technology. We have enormous capacity for clean energy production. We are resilient, determined, hard-working. There is nothing America can't do – if we do it together. And I know this because this is how I was raised. And I don't think Mr. Trump understands any of it. He hasn't offered any credible solutions for the very real economic challenges we face. Those challenges emerged long before the Great Recession, and they have persisted through our recovery. There is too much inequality, too little upward mobility. It is just too hard to get ahead today. But there are common-sense things that your government could do that would give Americans more opportunities to succeed. Why don't we do it? Because powerful special interests and the tendency to put ideology ahead of political progress have led to gridlock in Congress. How can you not be frustrated, and even angry, when you see nothing getting done? And a lot of people feel no one is on their side and no one has their back and that is not how it's supposed to be in America. If I am an fortunate enough to be your President, I will have your back every single day that I serve. My mission in the White House will be to make our economy work for everyone, not just those at the top. This is personal for me. I am the product of the American middle class, I was born in Chicago, I was raised in a suburb. But my grandfather worked at the Scranton lace mill in Scranton, Pennsylvania, for 50 years. And because he worked hard, my Dad was able to go to college, and eventually start his own small business – and then send me out into the world to follow my dreams. No matter how far those dreams have taken me, I have always remembered, I'm the daughter of a small-business owner and the granddaughter of a factory worker – and proud of both. So here's what I want. I want every American family to be able to tell the same story. If you work hard, you do your part, you should be able to give your children all the opportunities they deserve. That is the basic bargain of America. Now whether we will be able to renew that bargain on even better terms for the 21st century depends in large measure, on the outcome of this election. So here are four questions that I hope the American people will ask of both candidates – and that the answers should make your choice in November crystal clear: First, which candidate has a real plan to create good-paying jobs? Second, who will restore fairness to our economy and ensure that those at the top pay their fair share of taxes? Third, who will really go to bat for working families? And fourth, who can bring people together to deliver results that will make a difference in your lives? Now I hope that after giving a fair hearing to both sides, you will join the millions of people across our country supporting this campaign – not just Democrats, but a growing number of Republicans and Independents as well. Now when it comes to creating jobs, I would argue it's not even close. Even conservative experts say Trump's agenda will pull our economy back into recession. And according to an independent analysis by a former economic advisor to Senator John McCain, if you add up all of Trump's ideas – from cutting taxes for the wealthy and corporations, to starting a trade war with China, to deporting millions of hard-working immigrants – the result would be a loss of 3.4 million jobs. By contrast, the same analyst found that with our plans, the economy would create more than 19 million new jobs. So let me tell you how we would do that. I believe every American willing to work hard should be able to find a job that provides dignity, pride and decent pay that can support a family. So starting on Day One, we will work with both parties to pass the biggest investment in new, good-paying jobs since World War II. We will put Americans to work building and modernizing our roads, our bridges, our tunnels, our railways, our ports, our airports. We are way overdue for this, my friends. We are living off the investments that were made by our parents' and grandparents' generations. We will also help cities like Detroit and Flint connect underserved neighborhoods to opportunity, expanding affordable housing, and we will repair

Big Data Analytics

Hadoop Installation Guide

Eric Johnson / Ching-Yung Lin
efj2106@columbia.edu

[illegible]

Type: **ctrl + c** <THEN> **SHIFT + :** <THEN> **wq!** ENTER

Add file to HDFS directory then execute:

```
Execute: ./bin/hadoop jar hadoop-mapreduce-examples-2.6.0.jar wordcount hillary.txt output.txt
```