

Hadoop 的配置有些麻烦，目前没有一键配置的功能，虽然当时我在安装过程中也参考了有关教程，但还是遇到了很多坑，一些老版本的安装过程已不适用于 `hadoop2.x`，下面就介绍一下具体步骤。

安装 Java

因为之后使用 Hadoop 需要运行 jar 包，所以 Java 环境是必须的，这里不作赘述，相信大多数学习 Hadoop 的小伙伴计算机上都早已经搭好 Java 环境。

下载 Hadoop 源码

点击 [这里](#)，有所有的历史版本。本教程以 `hadoop-2.8.2` 为例，下载 `hadoop-2.8.2.tar.gz` 即可。下载完成后，解压到本地合适的目录下。

添加 Hadoop 环境变量

在 `.bash_profile` 文件中配置 Hadoop 的环境变量，使用 `vim` 打开该文件，在终端运行如下命令打开文件：

```
vim ~/.bash_profile
```

进入编辑模式后添加以下两行代码（修改成自己机子上的路径）：

```
export HADOOP_HOME=/Users/Cyan/coding/hadoop/hadoop-2.8.2
export PATH=$PATH:$HADOOP_HOME/bin
```

退出编辑模式，使用 `:wq` 保存修改，然后运行 `source` 命令使文件中的修改立即生效：

```
#source ~/.bash_profile
```

在终端执行命令：

```
hadoop version
```

结果如下，说明 hadoop 路径配置好了：

```
localhost:sbin Cyan$ hadoop version
Hadoop 2.8.2
Subversion https://git-wip-us.apache.org/repos/asf
f80c41f798373828fb
Compiled by jdu on 2017-10-19T20:39Z
Compiled with protoc 2.5.0
From source with checksum dce55e5afe30c210816b39b6
This command was run using /Users/Cyan/coding/hado
adoop-common-2.8.2.jar
```

添加好了环境变量，下面就是修改一些相关配置文件。

修改 Hadoop 的配置文件

需要修改的 Hadoop 配置文件都在目录 `etc/hadoop` 下，包括：

- `hadoop-env.sh`
- `core-site.xml`
- `hdfs-site.xml`
- `mapred-site.xml`
- `yarn-site.xml`

下面我们逐步进行配置：

1. 修改 `hadoop-env.sh` 文件

直接设置 `JAVA_HOME` 的路径，不要用 `$JAVA_HOME` 代替，因为 Hadoop 对系统变量的支持不是很好。修改下面两个路径（用你机子上的 `JAVA` 路径和 Hadoop 目录的路径代替）：

```
export
JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk1.8.0_151.jdk/Contents/Home
export HADOOP_CONF_DIR=/Users/Cyan/coding/hadoop/hadoop-2.8.2/etc/hadoop
```

2. 修改 **core-site.xml** 文件

设置 Hadoop 的临时目录和文件系统，localhost:9000 表示本地主机。如果使用远程主机，要用相应的 IP 地址来代替，填写远程主机的域名，则需要到 /etc/hosts 文件中做 DNS 映射。在 **core-site.xml** 文件里作如下配置：

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

  <!--用来指定hadoop运行时产生文件的存放目录 自己创建-->
  <property>
    <name>hadoop.tmp.dir</name>
    <value>/Users/Cyan/coding/hadoop/hadoop-2.8.2/tmp</value>
  </property>
</configuration>
```

3. 修改 **hdfs-site.xml** 文件

hdfs-site.xml 的配置修改如下：

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <!--不是root用户也可以写文件到hdfs-->
  <property>
    <name>dfs.permissions</name>
    <value>>false</value>    <!--关闭防火墙-->
  </property>
</configuration>
```

4. 修改 **mapred-site.xml** 文件

由于根目录下 **etc/hadoop** 中没有 **mapred-site.xml** 文件，所以需要创建该文件。但是目录中提供了 **mapred-site.xml.template** 模版文件。我们将其重命名为 **mapred-site.xml**，然后将 yarn 设置成数据处理框架：

```
<configuration>
  <property>
    <!--指定mapreduce运行在yarn上-->
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
```

```
</configuration>
```

5. 修改 yarn-site.xml 文件

配置数据的处理框架 yarn:

```
<configuration>
  <!-- Site specific YARN configuration properties -->
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
  <property>
    <name>yarn.resourcemanager.address</name>
    <value>localhost:9000</value>
  </property>
</configuration>
```

至此需要修改和配置的文件都已经没有问题了。

启动 Hadoop

1. 启动 NameNode

在终端运行命令:

```
hadoop namenode -format
```

结果如下就是成功了:

```
18/01/19 19:49:19 INFO util.GSet: Computing capacity
18/01/19 19:49:19 INFO util.GSet: VM type =
18/01/19 19:49:19 INFO util.GSet: 0.02999999932944
18/01/19 19:49:19 INFO util.GSet: capacity =
Re-format filesystem in Storage Directory /Users/C
r N) Y
18/01/19 19:49:23 INFO namenode.FSImage: Allocated
63686
18/01/19 19:49:23 INFO common.Storage: Storage dir
dfs/name has been successfully formatted.
18/01/19 19:49:23 INFO namenode.FSImageFormatProto
adoop-2.8.2/tmp/dfs/name/current/fsimage.ckpt_0000
18/01/19 19:49:23 INFO namenode.FSImageFormatProto
.8.2/tmp/dfs/name/current/fsimage.ckpt_000000000000
18/01/19 19:49:23 INFO namenode.NNStorageRetention
18/01/19 19:49:23 INFO util.ExitUtil: Exiting with
18/01/19 19:49:23 INFO namenode.NameNode: SHUTDOWN
/*****
SHUTDOWN_MSG: Shutting down NameNode at localhost/
*****/
```

2. 启动 HDFS

在终端首先进入 /sbin 目录：

```
cd /Users/Cyan/coding/hadoop/hadoop-2.8.2/sbin
```

然后启动 HDFS：

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

如果成功了，过程中需要输三次密码。

如果报错 “connection refused”，则需要在计算机系统设置中打开远程登录许可。

点击 **Sharing**（共享）：




勾选 **Remote Login**（远程登录），然后添加当前用户：




Computer Name: Cyan的MacBook Pro

Computers on your local network can
CyandeMacBook-Pro.local

On	Service
<input type="checkbox"/>	Screen Sharing
<input type="checkbox"/>	File Sharing
<input type="checkbox"/>	Printer Sharing
<input checked="" type="checkbox"/>	Remote Login
<input type="checkbox"/>	Remote Management
<input type="checkbox"/>	Remote Apple Events
<input type="checkbox"/>	Internet Sharing
<input type="checkbox"/>	Bluetooth Sharing

 Remote Login: C

To log in to this compu

Allow access for: 

这样就会解决 **connection** 的问题，如果还有其他错误，请检查路径添加的版本和本机的版本是否匹配。

3. 启动 yarn

在终端首先进入 **/sbin** 目录：

```
cd /Users/Cyan/coding/hadoop/hadoop-2.8.2/sbin
```

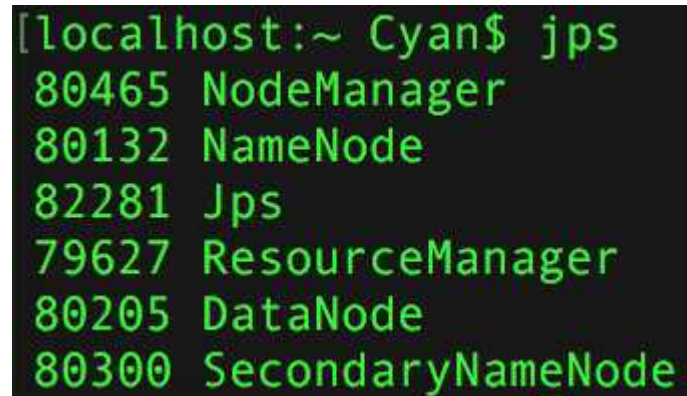
然后启动 yarn：

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

在终端执行：

```
jps
```

结果如下，证明 Hadoop 可以成功启动：

A terminal window with a black background and green text. The prompt is [localhost:~ Cyan\$. The command jps has been executed, and the output lists six processes with their respective PIDs: NodeManager (80465), NameNode (80132), Jps (82281), ResourceManager (79627), DataNode (80205), and SecondaryNameNode (80300).

```
[localhost:~ Cyan$ jps
80465 NodeManager
80132 NameNode
82281 Jps
79627 ResourceManager
80205 DataNode
80300 SecondaryNameNode
```

我们也可以在浏览器中打开 localhost:50070/ 来查看 Hadoop 的启动情况：

Overview

localhost:9

Started:	Fr
Version:	2.1
Compiled:	Th
Cluster ID:	Cl
Block Pool ID:	BP

Hadoop 的安装启动就完成啦！接下来就可以通过一些 **shell** 命令来操作 Hadoop 下的文件了，例如：

<code>hadoop fs -ls /</code>	查看根目录下的文件及文件夹
<code>hadoop fs -mkdir /test</code>	在根目录下创建一个文件夹 <code>testdata</code>
<code>hadoop fs -rm /.../...</code>	移除某个文件
<code>hadoop fs -rmr /...</code>	移除某个空的文件夹