

ubuntu 安装 hadoop

选择 jdk1.7\_6 位、Hadoop-2.10.1

先创建一个 java 文件夹

```
sudo mkdir -p /usr/local
```

再将当前目录下的 jdk-8u291-linux-x64.tar.gz 解压到 java 文件夹

```
sudo tar -zxvf ./jdk-8u291-linux-x64.tar.gz -C /usr/local/java
```

配置环境变量

```
sudo vim ~/.bashrc
```

在末尾添加

```
export JAVA_HOME=/usr/local/jdk/jdk1.8.0_291
```

```
export CLASSPATH=.:$JAVA_HOME/lib/*jar
```

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

使环境变量生效

```
source ~/.bashrc
```

检查 JDK 是否安装成功

```
java -version
```

安装 hadoop

```
sudo tar -zxvf ./hadoop-2.10.1.tar.gz -C /usr/local
```

重命名并修改权限

```
cd /usr/local
```

```
sudo mv ./hadoop-2.10.1/ ./hadoop
```

```
sudo chown -R 你的用户名 ./hadoop
```

ssh 配置

更新 apt 包

```
sudo apt-get update
```

安装 ssh server

```
sudo apt-get install openssh-server
```

配置 ssh

```
ssh localhost
```

再输入密码

接下来配置无密码登录,进入.ssh 文件夹

```
cd ~/.ssh/
```

使用 RSA 算法生成秘钥对

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

运行一致对于所有选择直接回车即可

然后将公钥加入到授权

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

再次 ssh localhost 使就不需要密码了

配置 Hadoop 伪分布式

伪分布式需要修改 2 个配置文件 **core-site.xml** 和 **hdfs-site.xml** , 路径在 :

/usr/local/hadoop/etc/hadoop,

打开 core-site.xml, 添加如下配置

```
<configuration>
```

```
    <property>
```

```
        <name>hadoop.tmp.dir</name>
```

```
        <value>/usr/local/hadoop/tmp</value>
```

```
        <description>Abase for other temporary directories.</description>
```

```
    </property>
```

```
    <property>
```

```
        <name>fs.defaultFS</name>
```

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

```
    </property>
```

```
</configuration>
```

打开 hdfs-site.xml, 添加如下配置

```
<configuration>
```

```
    <property>
```

```
        <name>dfs.replication</name>
```

```
        <value>1</value>
```

```
    </property>
```

```

    <property>
        <name>dfs.namenode.name.dir</name>
        <value>/usr/local/hadoop/tmp/dfs/name</value>
    </property>
    <property>
        <name>dfs.datanode.data.dir</name>
        <value>/usr/local/hadoop/tmp/dfs/data</value>
    </property>
</configuration>

```

手动修改 hadoop 环境中的 JAVA\_HOME

cd etc/hadoop

vim hadoop-env.sh

在文件中修改

```
export JAVA_HOME=/usr/local/java/jdk1.8.0_181
```

切换回到 hadoop 主目录/usr/local/Hadoop

进行 namenode 的格式化

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

开启进程

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

验证已开启的进程

```
jps
```

可以看到 Datanode、JPS、namenode、secondaryNameNode 已开启

首先在 HDFS 中创建用户文件夹

在/usr/local/hadoop 目录下：

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

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

将 etc/hadoop 下所有的 xml 问价复制到 input 文件夹

```
./bin/hdfs dfs -put ./etc/hadoop/*.xml /user/hadoop/input
```

可以通过 ls 查看复制后的文件夹情况

```
./bin/hdfs dfs -ls /user/hadoop/input
```

运行一个例子：

```
./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-2.10.1.jar grep  
/user/hadoop/input output 'dfs[a-z]+'
```

查看运行结果

```
./bin/hdfs dfs -cat output/*
```

得到如下结果即可：

```
1 dfsadmin
```