

1. 安装JDK

1.1 JDK安装步骤

1. 下载JDK安装包（下载Linux系统的 .tar.gz 的安装包）

<https://www.oracle.com/java/technologies/javase/javase-jdk8-downloads.html>

2. 更新Ubuntu源

```
sudo apt-get update
```

3. 将JDK压缩包解压到Ubuntu系统中 /usr/local/ 中

```
sudo tar -zxvf jdk-8u251-linux-x64.tar.gz -C /usr/local/
```

4. 将解压的文件夹重命名为 jdk8

```
cd /usr/local/
```

```
sudo mv jdk1.8.0_251/ jdk8
```

5. 添加到环境变量

```
cd /home/tarena/
```

```
sudo gedit .bashrc
```

在文件末尾添加如下内容:

```
1 export JAVA_HOME=/usr/local/jdk8
2 export JRE_HOME=$JAVA_HOME/jre
3 export CLASSPATH=.:$JAVA_HOME/lib:$JRE_HOME/lib
4 export PATH=.:$JAVA_HOME/bin:$PATH
```

```
source .bashrc
```

6. 验证是否安装成功

```
java -version
```

出现java的版本则证明安装并添加到环境变量成功 java version "1.8.0_251"

2. 安装Hadoop并配置伪分布式

2.1 Hadoop安装配置步骤

1. 安装SSH

```
sudo apt-get install ssh
```

2. 配置免登录认证,避免使用Hadoop时的权限问题

```
ssh-keygen -t rsa （输入此条命令后一路回车）
```

```
cd ~/.ssh
```

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

```
ssh localhost （发现并未让输入密码即可连接）
```

```
exit （退出远程连接状态）
```

3. 下载Hadoop 2.10（374M）

<https://archive.apache.org/dist/hadoop/common/hadoop-2.10.0/hadoop-2.10.0.tar.gz>

4. 解压到 /usr/local 目录中,并将文件夹重命名为 hadoop , 最后设置权限

```
sudo tar -zxvf hadoop-2.10.0.tar.gz -C /usr/local/  
cd /usr/local  
sudo mv hadoop-2.10.0/ hadoop2.10  
sudo chown -R tarena hadoop2.10/
```

5. 验证Hadoop

```
cd /usr/local/hadoop2.10/bin  
./hadoop version （此处出现hadoop的版本）
```

6. 设置JAVE_HOME环境变量

```
sudo gedit /usr/local/hadoop2.10/etc/hadoop/hadoop-env.sh  
把原来的export JAVA_HOME=${JAVA_HOME}改为  
export JAVA_HOME=/usr/local/jdk8
```

7. 设置Hadoop环境变量

```
sudo gedit /home/tarena/.bashrc  
在末尾追加
```

```
1 export HADOOP_HOME=/usr/local/hadoop2.10  
2 export CLASSPATH=.:{JAVA_HOME}/lib:${HADOOP_HOME}/sbin:$PATH  
3 export PATH=.:${HADOOP_HOME}/bin:${HADOOP_HOME}/sbin:$PATH
```

```
source /home/tarena/.bashrc
```

8. 伪分布式配置, 修改2个配置文件 (core-site.xml 和 hdfs-site.xml)

9. 修改core-site.xml

```
sudo gedit /usr/local/hadoop2.10/etc/hadoop/core-site.xml
```

添加如下内容

```
1 <configuration>  
2   <property>  
3     <name>hadoop.tmp.dir</name>  
4     <value>file:/usr/local/hadoop2.10/tmp</value>  
5   </property>  
6   <property>  
7     <name>fs.defaultFS</name>  
8     <value>hdfs://localhost:9000</value>  
9   </property>  
10 </configuration>
```

10. 修改hdfs-site.xml

```
sudo gedit /usr/local/hadoop2.10/etc/hadoop/hdfs-site.xml
```

添加如下内容

```
1 <configuration>  
2   <property>  
3     <name>dfs.replication</name>  
4     <value>1</value>
```

```

5     </property>
6     <property>
7         <name>dfs.namenode.name.dir</name>
8         <value>file:/usr/local/hadoop2.10/tmp/dfs/name</value>
9     </property>
10    <property>
11        <name>dfs.datanode.data.dir</name>
12        <value>file:/usr/local/hadoop2.10/tmp/dfs/data</value>
13    </property>
14 </configuration>

```

11. 配置YARN - 1

```

cd /usr/local/hadoop2.10/etc/hadoop
cp mapred-site.xml.template mapred-site.xml
sudo gedit mapred-site.xml

```

添加如下配置

```

1 <property>
2     <name>mapreduce.framework.name</name>
3     <value>yarn</value>
4 </property>

```

12. 配置YARN - 2

```

sudo gedit yarn-site.xml

```

添加如下配置：

```

1 <property>
2     <name>yarn.nodemanager.aux-services</name>
3     <value>mapreduce_shuffle</value>
4 </property>

```

13. 执行NameNode格式化

```

cd /usr/local/hadoop2.10/bin
./hdfs namenode -format

```

出现 Storage directory /usr/local/hadoop2.10/tmp/dfs/name has been successfully formatted 则表示格式化成功

14. 启动Hadoop所有组件

```

cd /usr/local/hadoop2.10/sbin
./start-all.sh

```

启动时可能会出现警告，直接忽略即可，不影响正常使用

15. 启动成功后，可访问Web页面查看 NameNode 和 Datanode 信息，还可以在线查看 HDFS 中的文件

<http://localhost:50070>

16. 查看Hadoop相关组件进程

```

jps

```

会发现如下进程

```
1 | NameNode --- 50070
2 | DataNode --- 50075
3 | SecondaryNameNode --- 50090
4 | ResourceManager --- 8088
5 | NodeManager
```

17. 测试 - 将本地文件上传至hdfs

hadoop fs -put 一个本地的任意文件 /

hadoop fs -ls /

也可以在浏览器中Utilities->Browse the file system查看

3. Hive安装

3.1 详细安装步骤

1. 下载hive安装包 (2.3.7版本)

<http://us.mirrors.quenda.co/apache/hive/>

2. 解压到 /usr/local/ 目录下

```
sudo tar -zxvf apache-hive-2.3.7-bin.tar.gz -C /usr/local
```

3. 给文件夹重命名

```
sudo mv /usr/local/apache-hive-2.3.7-bin /usr/local/hive2.3.7
```

4. 设置环境变量

```
sudo gedit /home/tarena/.bashrc
```

在末尾添加如下内容

```
1 | export HIVE_HOME=/usr/local/hive2.3.7
2 | export PATH=.:${HIVE_HOME}/bin:$PATH
```

5. 刷新环境变量

```
source /home/tarena/.bashrc
```

6. 下载并添加连接MySQL数据库的jar包 (8.0.19 Ubuntu Linux Ubuntu Linux 18.04)

下载链接: <https://downloads.mysql.com/archives/c-j/>

解压后找到 mysql-connector-java-8.0.19.jar

将其拷贝到 /usr/local/hive2.3.7/lib

```
sudo cp -p mysql-connector-java-8.0.19.jar /usr/local/hive2.3.7/lib/
```

7. 创建hive-site.xml配置文件

```
sudo touch /usr/local/hive2.3.7/conf/hive-site.xml
```

```
sudo gedit /usr/local/hive2.3.7/conf/hive-site.xml
```

并添加如下内容

```
1 | <configuration>
2 |     <property>
3 |         <name>javax.jdo.option.ConnectionURL</name>
4 |         <value>jdbc:mysql://localhost:3306/hive?
createDatabaseIfNotExist=true</value>
```

```

5      <description>JDBC connect string for a JDBC
metastore</description>
6      </property>
7      <property>
8          <name>javax.jdo.option.ConnectionDriverName</name>
9          <value>com.mysql.cj.jdbc.Driver</value>
10         <description>Driver class name for a JDBC
metastore</description>
11         </property>
12         <property>
13             <name>javax.jdo.option.ConnectionUserName</name>
14             <value>root</value>
15             <description>username to use against metastore
database</description>
16         </property>
17         <property>
18             <name>javax.jdo.option.ConnectionPassword</name>
19             <value>123456</value>
20             <description>password to use against metastore
database</description>
21         </property>
22 </configuration>

```

8. 在hive配置文件中添加hadoop路径

```

cd /usr/local/hive2.3.7/conf
sudo cp -p hive-env.sh.template hive-env.sh
sudo gedit /usr/local/hive2.3.7/conf/hive-env.sh
添加如下内容:

```

```

1  HADOOP_HOME=/usr/local/hadoop2.10
2  export HIVE_CONF_DIR=/usr/local/hive2.3.7/conf

```

9. hive元数据初始化

```
schematool -dbType mysql -initSchema
```

10. 测试hive

```
hive
```

```
hive>show databases;
```

如果能够正常显示内容，则hive安装并配置完毕