**电 子 科 技 大 学**

**实** **验**

**报** **告**

**学生姓名：** 赵则橦

**学号：** 2017221305013

**指导教师：** 黄俊

**实验地点：** 信软楼

**实验时间：2019/11**

**一、 实验名称：**构建Hadoop单机环境模式

**二、** **实验学时：**2学时

**三、 实验目的：**

1. 掌握 Hadoop 的环境搭建

2. 掌握 Hadoop、Linux 的基本命令。

**四、 实验原理**

首先需要创建用户，然后授予用户权利，在 linux 系统上，如果用户没有读写运行的权利将会有许多事务收到限制。然后安装 ssh，安装 ssh 的目的在能在后期做多设备 hadoop 实验时候能够登录到别人的设备上，并且无需密码。运行

* wordCount 代码在 hadoop 的 share 包里，能够检索文本里面单词出现的次数，但这次实验只在单机上运行，所以无论输入和输出都在单机上进行。

**五、 实验内容**

首先安装 SSH ， SSH 为 Secure Shell 的缩写，由 IETF 的网络小组（Network Working Group）所制定；SSH 为建立在应用层基础上的安全协议。安装 SSH 的目的在于为了方便配置集群,能在一台机器上登陆到集群中其他机器上。

然后在 ubuntu 上安装 hadoop，Hadoop 是一个由 Apache 基金会所开发的分布式系统基础架构。用户可以在不了解分布式底层细节的情况下，开发分布式程序。

最后运行在 hadoop 代码包上的 wordCount 程序，并将 input 里面的 readme.txt 作为输入，将 output 作为输出，利用命令将结果输出到屏幕上。

**六、 实验设备及环境**

Ubuntu jdk 1.8 hadoop 2.7.X

**七、 实验步骤**

1. 在 ubuntu 中创建 hadoop 用户。
2. 安装 ssh。

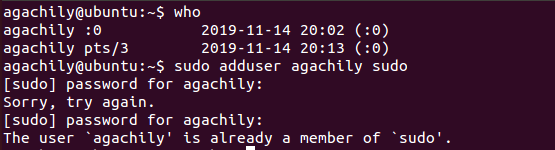
1. 安装 java。
2. 安装 hadoop。
3. 测试 wordcount。

**八、 实验结果与分析（含重要数据结果分析或核心代码流程分析）**

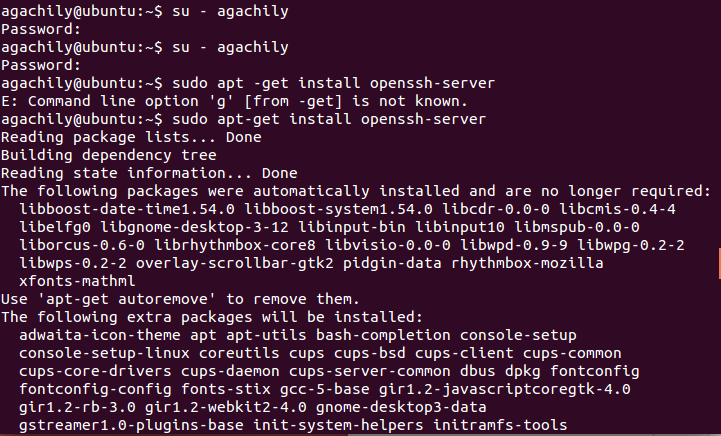
1. 在 ubuntu 中创建 hadoop 用户:由于实验在自己电脑上完成，此步骤已经在

此之前完成，并且为用户增加管理员权限

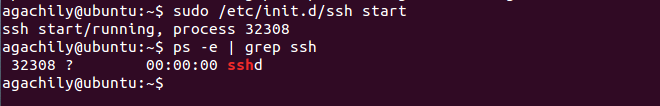
系统中已有用户agachily



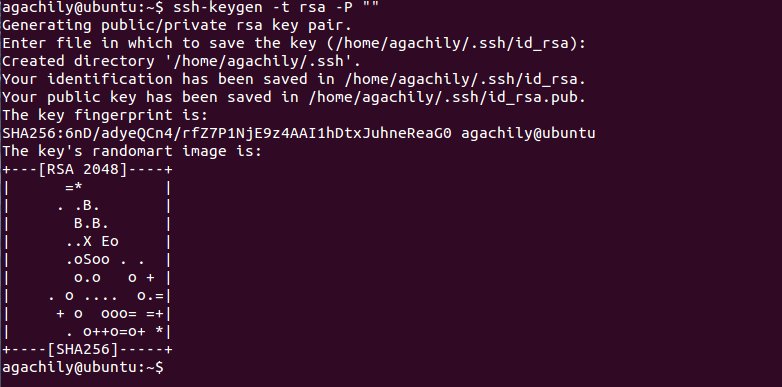
1. 安装 ssh



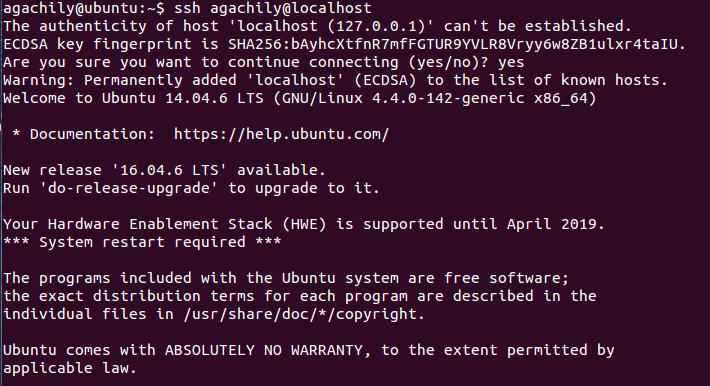
安装好ssh服务后，打开这个服务，查看ssh服务是否正确启动：



设置免密码登录，生成私钥和公钥，在／home／hadoop/.ssh里面生成两个文件 id\_rsa、id\_rsa.pub，前者是私钥，后者是公钥

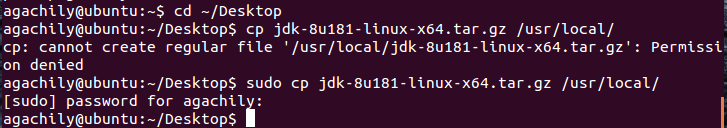


将公钥追加到authorized\_keys里面，它保存所有允许以当前用户身份登录到ssh客户端用户的公钥内容，然后登陆ssh，然后退出

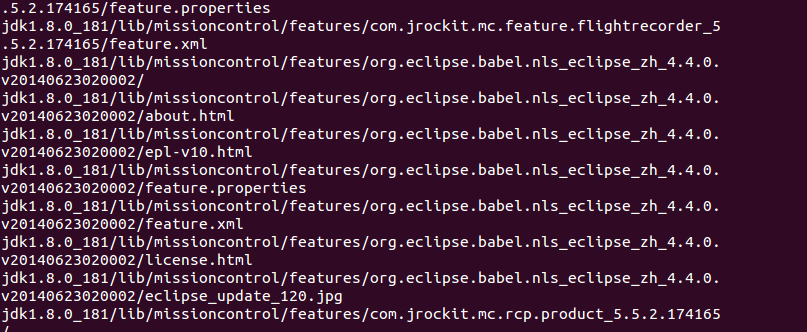


1. 安装java环境

先是把压缩包复制到欲安装目录下：



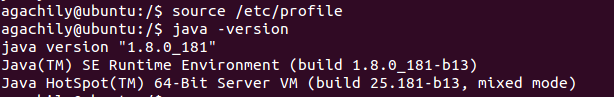
进入上述目录解压：



1. 配置路径安装

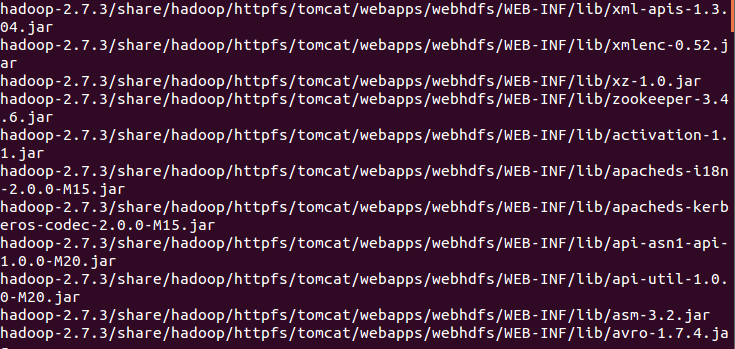
打开配置文件：  
sudo gedit /etc/profile  
添加以下命令语句：  
export JAVA\_HOME=/usr/local/jdk1.8.0\_181  
export PATH=$JAVA\_HOME/bin:$PATH  
export CLASS\_PATH=.:$JAVA\_HOME/lib/dt.jar:$JAVA\_HOME/lib/tools.jar

应用配置并测试：



1. 安装hadoop

解压并copy到/usr/local



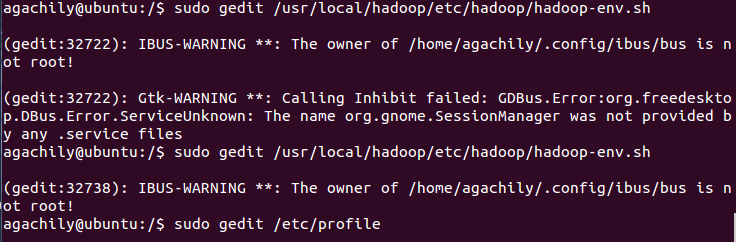
切换到到/usr/local/，对hadoop-2.7.3重命名，并赋予权限

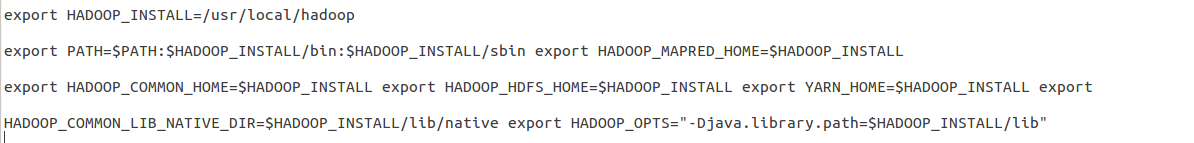


修改目录主人



打开配置文件并进行修改操作



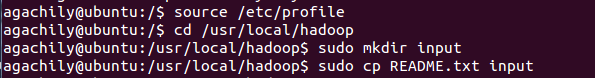


应用配置：

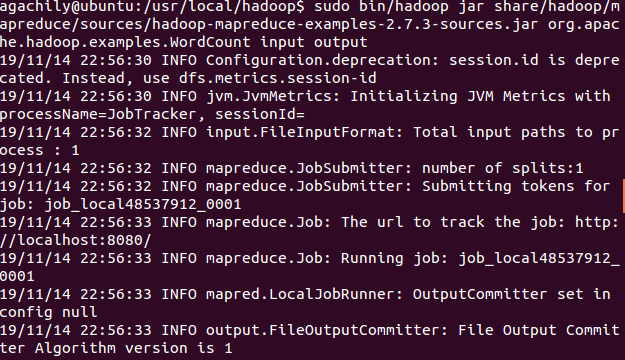


1. 测试

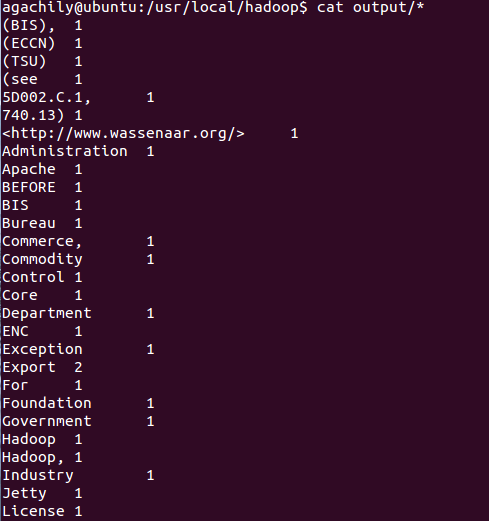
进入hadoop 目录，创建目录和测试用例



运行测试样例：



执行 cat output/\*，查看字符统计结果



1. **总结及心得体会：**

通过这次实验，我成功构建了Hadoop单机环境模式，在本地上安装了ssh，配置了java环境和ssh。我首先对ssh进行了一定的了解，ssh为建立在应用层基础上的安全协议，安装ssh可以方便配置集群，能在一台机器上登陆到集群中其他机器上。至于hadoop则是一个分布式系统基础架构，方便用户在不了解分布式底层细节情况下开发分布式程序。在hadoop代码包中有一个WordCount程序，将input中的readme进行输入，将output作为输出。在配置java环境中，由于实验指导书中的配置文件有误，导致我很久无法配置成功，希望这个问题能够得到纠正，但在配置java环境时，我也对配置文件中的三个部分具体是做什么的有了更清楚的认识。

在配置hadoop时，我是安装到了/usr/local目录下，在后面我由于有几次没有注意操作目录而没有成功运行，这也让我长了个记性。这次实验总的来说是简单易行的，从中我熟悉了linux的操作环境，也对hadoop和java的配置文件进行了详细的阅读，知道了为啥还需要这些文件。我想的是，不像在windows系统中，在linux系统下只有一个解压的过程，所以通过配置文件是和系统进行沟通，使得文件可以成功运行。这也让我认识到了linux系统和windows系统中的一些区别。

1. **对本实验过程及方法、手段的改进建议：**

希望能够提供更详细明确的实验指导书