



实验3：常用Linux操作和Hadoop操作

廖 军

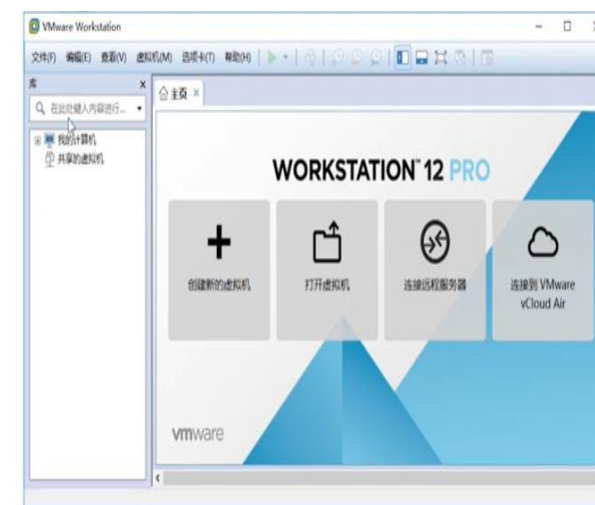
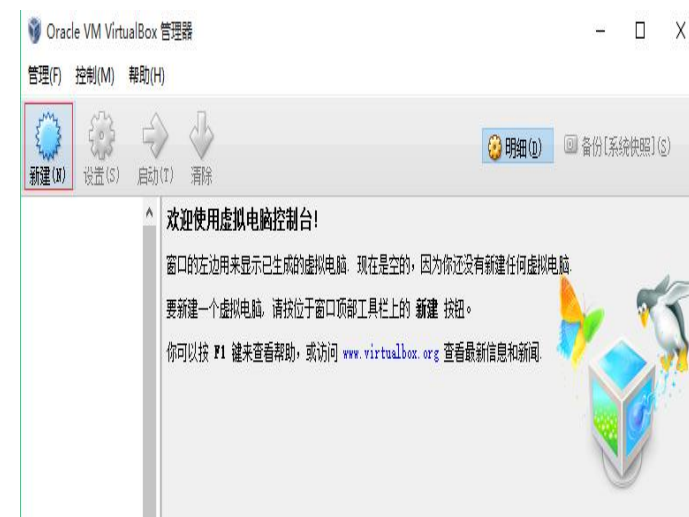
重庆大学大数据与软件学院

E-mail: liaojun@cqu.edu.cn

Linux虚拟机安装和操作

工具准备

- 1. 一、材料和工具
- 1、下载VirtualBox虚拟机软件或vmare虚拟机软件
- 2. 下载Ubuntu LTS 14.04 ISO映像文件或16.04版本



详细安装教程请参考《大数据技术原理与应用 第二章 大数据处理架构Hadoop 学习指南》
使用开源虚拟机软件VirtualBox安装Ubuntu: <http://dbl原因.xmu.edu.cn/blog/337-2/>

Hadoop安装和操作

Hadoop基本安装配置主要包括以下几个步骤：

- 创建Hadoop用户
- SSH登录权限设置
- 安装Java环境
- 单机安装配置
- 伪分布式安装配置

详细安装配置过程请参考厦门大学数据库实验室出品教程

《**Hadoop安装教程**》

hadoop2版本 <http://dblab.xmu.edu.cn/blog/install-hadoop/>

或hadoop 3版本 <http://dblab.xmu.edu.cn/blog/2441-2/>

Hadoop安装和操作

1、创建Hadoop用户：

如果安装 Ubuntu 的时候不是用的 “hadoop” 用户，那么需要增加一个名为 hadoop 的用户

首先按 **ctrl+alt+t** 打开终端窗口，输入如下命令创建新用户：

```
$ sudo useradd -m hadoop -s /bin/bash
```

上面这条命令创建了可以登陆的 hadoop 用户，并使用 /bin/bash 作为 shell

接着使用如下命令设置密码，可简单设置为 hadoop，按提示输入两次密码：

```
$ sudo passwd hadoop
```

可为 hadoop 用户增加管理员权限，方便部署，避免一些对新手来说比较棘手的权限问题：

```
$ sudo adduser hadoop sudo
```

Hadoop安装和操作

2、SSH登录权限设置

SSH是什么？

SSH 为 Secure Shell 的缩写，是建立在应用层和传输层基础上的安全协议。SSH 是目前较可靠，专为远程登录会话和其他网络服务提供安全性的协议。利用 SSH 协议可以有效防止远程管理过程中的信息泄露问题。SSH最初是UNIX系统上的一个程序，后来又迅速扩展到其他操作平台。SSH是由[客户端](#)和[服务端](#)的软件组成，服务端是一个守护进程(daemon)，它在后台运行并响应来自客户端的连接请求，客户端包含ssh程序以及像scp（远程拷贝）、slogin（远程登陆）、sftp（安全文件传输）等其他的应用程序

配置SSH的原因：

Hadoop名称节点（NameNode）需要启动集群中所有机器的Hadoop守护进程，这个过程需要通过SSH登录来实现。Hadoop并没有提供SSH输入密码登录的形式，因此，为了能够顺利登录每台机器，需要将所有机器配置为名称节点可以无密码登录它们

Hadoop安装和操作

3、安装Java环境

- Java环境可选择 Oracle 的 JDK，或是 OpenJDK
- 可以在Ubuntu中直接通过命令安装 OpenJDK 7

```
$ sudo apt-get install openjdk-7-jre openjdk-7-jdk
```

- 还需要配置一下 JAVA_HOME 环境变量
具体请参考网络教程：<http://dblab.xmu.edu.cn/blog/install-hadoop/>
或 <http://dblab.xmu.edu.cn/blog/2441-2/>

Hadoop安装和操作

4、单机安装配置

Hadoop 2 安装文件的下载

Hadoop 2 可以到官网下载，需要下载 **hadoop-2.x.y.tar.gz** 这个格式的文件，这是编译好的，另一个包含 **src** 的则是 Hadoop 源代码，需要进行编译才可使用

- 如果读者是使用虚拟机方式安装Ubuntu系统的用户，请用虚拟机中的Ubuntu自带firefox浏览器访问本指南，再点击下载地址，才能把hadoop文件下载到虚拟机ubuntu中。请不要使用Windows系统下的浏览器下载，文件会被下载到Windows系统中，虚拟机中的Ubuntu无法访问外部Windows系统的文件，造成不必要的麻烦。
- 如果读者是使用双系统方式安装Ubuntu系统的用户，请进去Ubuntu系统，在Ubuntu系统打开firefox浏览器，再点击下载

Hadoop安装和操作

4、单机安装配置

选择将 Hadoop 安装至 /usr/local/ 中

```
$ sudo tar -zxf ~/下载/hadoop-2.6.0.tar.gz -C /usr/local # 解压到/usr/local中  
$ cd /usr/local/  
$ sudo mv ./hadoop-2.6.0/ ./hadoop # 将文件夹名改为hadoop  
$ sudo chown -R hadoop:hadoop ./hadoop # 修改文件权限
```

Hadoop 解压后即可使用。输入如下命令来检查 Hadoop 是否可用，成功则会显示 Hadoop 版本信息：

```
$ cd /usr/local/hadoop  
$ ./bin/hadoop version
```

Hadoop 默认模式为非分布式模式（本地模式），无需进行其他配置即可运行。

Hadoop安装和操作

5、伪分布式安装配置

- Hadoop 可以在单节点上以伪分布式的方式运行，Hadoop 进程以分离的 Java 进程来运行，节点既作为 NameNode 也作为 DataNode，同时，读取的是 HDFS 中的文件
- Hadoop 的配置文件位于 /usr/local/hadoop/etc/hadoop/ 中，伪分布式需要修改2个配置文件 **core-site.xml** 和 **hdfs-site.xml**
- Hadoop的配置文件是 xml 格式，每个配置以声明 property 的 name 和 value 的方式来实现

实验步骤：

- 修改配置文件：core-site.xml，hdfs-site.xml，mapred-site.xml
- 初始化文件系统hadoop namenode -format
- 启动所有进程start-all.sh
- 访问web界面，查看Hadoop信息
- 运行实例

5、伪分布式安装配置

修改配置文件 **core-site.xml**

```
<configuration>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>file:/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>
```

- hadoop.tmp.dirhadoop表示存放数据的目录，即包括NameNode的数据，也包括DataNode的数据。该路径任意指定，只要实际存在该文件夹即可

- name为fs.defaultFS的值，表示hdfs路径的逻辑名称

Hadoop安装和操作

5、伪分布式安装配置

修改配置文件 **hdfs-site.xml**

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>1</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/usr/local/hadoop/tmp/dfs/name</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/usr/local/hadoop/tmp/dfs/data</value>
  </property></configuration>
```

- dfs.replication表示副本的数量，伪分布式要设置为1
- dfs.namenode.name.dir表示本地磁盘目录，是存储fsimage文件的地方
- dfs.datanode.data.dir表示本地磁盘目录，HDFS数据存放block的地方

关于三种Shell命令方式的区别：

1. hadoop fs
2. hadoop dfs
3. hdfs dfs

- hadoop fs适用于任何不同的文件系统，比如本地文件系统和HDFS文件系统
- hadoop dfs只能适用于HDFS文件系统
- hdfs dfs跟hadoop dfs的命令作用一样，也只能适用于HDFS文件系统

详细安装配置过程请参考厦门大学数据库实验室出品教程

《**Hadoop安装教程**》

<http://dblab.xmu.edu.cn/blog/install-hadoop/>
或 <http://dblab.xmu.edu.cn/blog/2441-2/>