# Hadoop：单机安装配置[¶](" \l "Hadoop：单机安装配置)

## 实验介绍[¶](" \l "实验介绍)

#### 1. 实验内容[¶](" \l "1.-实验内容)

* hadoop单机模式安装
* 测试安装

#### 2. 实验知识点[¶](" \l "2.-实验知识点)

* 下载解压/环境变量配置
* Linux/shell
* 测试WordCount程序

#### 3. 实验环境[¶](" \l "3.-实验环境)

* hadoop2.6.0
* jupyter kernel终端

#### 4. 适合人群[¶](" \l "4.-适合人群)

本课程难度为一般，属于初级级别课程，适合具有linux基础的用户。

## 实验要求[¶](" \l "实验要求)

* 单机模式：默认情况下运行为一个单独机器上的独立Java进程，主要用于调试环境

## 实验步骤[¶](" \l "实验步骤)

需要先添加用来运行Hadoop进程的用户组hadoop及用户hadoop。

#### 1. 添加用户及用户组[¶](" \l "1.-添加用户及用户组)

创建用户和用户组hadoop

$ sudo mkdir -p /hadoop

$ sudo groupadd hadoop

$ sudo useradd -g hadoop -G hadoop -d /hadoop hadoop

$ sudo chown -R hadoop:hadoop /hadoop

按照提示输入hadoop用户的密码，例如密码设定为 hadoop。注意输入密码的时候是不显示的。

$ sudo passwd hadoop

Enter new UNIX password: hadoop

Retype new UNIX password: hadoop

passwd: password updated successfully

#### 2. 添加sudo权限[¶](" \l "2.-添加sudo权限)

将hadoop用户添加进sudo用户组

$ sudo usermod -G sudo hadoop

#### 3. 安装及配置依赖的软件包[¶](" \l "3.-安装及配置依赖的软件包)

#### (1) hadoop环境需要预安装openssh-server、java等，这些软件包在实验环境中如果没有，需要手工安装。[¶](" \l "(1)-hadoop环境需要预安装openssh-server、java等，这)

安装并启动openssh-verver：

$ sudo apt-get install openssh-server -y

$ sudo service ssh start

验证环境执行下列指令：

$ ssh -V

OpenSSH\_6.6.1p1 Ubuntu-2ubuntu2.8, OpenSSL 1.0.1f 6 Jan 2014

$ java -version

java version "1.7.0\_121"

OpenJDK Runtime Environment (IcedTea 2.6.8) (7u121-2.6.8-1ubuntu0.14.04.3)

OpenJDK 64-Bit Server VM (build 24.121-b00, mixed mode)

#### (2) 配置ssh免密码登录[¶](" \l "(2)-配置ssh免密码登录)

切换到hadoop用户，需要输入添加hadoop用户时配置的密码。后续步骤都将在hadoop用户的环境中执行。

$ su - hadoop

配置ssh环境免密码登录。

在/hadoop目录下执行

$ ssh-keygen -t rsa #一路回车

$ cat .ssh/id\_rsa.pub >> .ssh/authorized\_keys

$ chmod 600 .ssh/authorized\_keys

验证登录本机是否还需要密码，第一次需要输入yes，以后不需要密码就可以登录。

$ ssh localhost

The authenticity of host 'localhost (::1)' can't be established.

ECDSA key fingerprint is 33:5d:12:e4:d5:59:8b:a3:a3:46:45:fd:16:f7:51:c8.

Are you sure you want to continue connecting (yes/no)? yes

Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.

The programs included with the Debian GNU/Linux system are free software;

the exact distribution terms for each program are described in the

individual files in /usr/share/doc/\*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent

permitted by applicable law.

$

#### 4.下载并安装Hadoop[¶](#4.下载并安装Hadoop)

在hadoop用户登录的环境中进行下列操作：

#### (1) 下载Hadoop 2.6.0[¶](#(1)-下载Hadoop-2.6.0)

这里我们将它的压缩包已经下载放到了根目录下，我们要将其拷贝到/hadoop目录下：

$ cp /home/ds/data/tools/hadoop-2.6.0.tar.gz /hadoop/

#### (2) 解压并安装[¶](" \l "(2)-解压并安装)

$ cd /hadoop

$ tar zxvf hadoop-2.6.0.tar.gz

$ mv hadoop-2.6.0 hadoop

$ chmod 777 hadoop

#### (3) 配置Hadoop[¶](#(3)-配置Hadoop)

配置环境变量的文件我们已经写好，命名为hadoop\_bash\_profile.src，并放到/root目录下，使用以下命令将其拷贝到/hadoop下：

$ sudo cp /root/hadoop\_bash\_profile.src /hadoop/.bash\_profile

此时可能提示需要输入password，输入 hadoop；

[sudo] password for hadoop: hadoop

使用如下命令查看：

$ more /hadoop/.bash\_profile

如果该文件下是否包含如下内容：

#HADOOP START

export JAVA\_HOME=/usr/lib/jvm/java-7-openjdk-amd64

export HADOOP\_HOME=/hadoop/hadoop

export PATH=$PATH:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

export HADOOP\_MAPRED\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_HOME=$HADOOP\_HOME

export HADOOP\_HDFS\_HOME=$HADOOP\_HOME

export YARN\_HOME=$HADOOP\_HOME

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib"

#HADOOP END

配置成功后，激活新加的环境变量

$ bash

$ source ~/.bash\_profile

至此，Hadoop单机模式安装完成，可以通过下述步骤的测试来验证安装是否成功。

## 实验结果[¶](" \l "实验结果)

创建输入的数据，暂时采用/etc/protocols文件作为测试

$ cd /hadoop/hadoop

$ mkdir input

$ cp /etc/protocols ./input

执行Hadoop WordCount应用（词频统计）

$ bin/hadoop jar share/hadoop/mapreduce/sources/hadoop-mapreduce-examples-2.6.0-sources.jar org.apache.hadoop.examples.WordCount input output

查看生成的单词统计数据

$ cat output/\*

## 练习题[¶](" \l "练习题)

请使用hadoop的wordcount对日志文件/var/log/dpkg.log进行词频统计。 将你执行的命令，和输出的结果粘贴到下面的文本框中。

作业练习：

###输入你的作业代码###

###作业代码结束###