

给力星 (<http://www.powerxing.com/>)

追逐内心的平和

首页 (<http://www.powerxing.com/>)

■ 笔记 (<http://www.powerxing.com/notes/>)

搜藏 (<http://www.powerxing.com/articles/>)

代码 (<http://www.powerxing.com/snippets/>)

音乐 (<http://www.powerxing.com/music/>)

关于 (<http://www.powerxing.com/about/>)

搜索

Hadoop安装教程_单机/伪分布式配置 _Hadoop2.6.0/Ubuntu14.04

📅 2014-08-09 (updated: 2016-01-06) 👁 67618



厦门大学林子雨编著《大数据技术原理与应用》
人民邮电出版社出版发行 21世纪高等教育计算机规划教材
中国高校第一本系统介绍大数据知识专业教材
京东、当当网等各大网店畅销书籍
最完备的大数据课程公共服务体系

➡ 点击进入
教材官网

(<http://dblab.xmu.edu.cn/post/bigdata/>)

当开始着手实践 Hadoop 时，安装 Hadoop 往往会成为新手的一道门槛。尽管安装其实很简单，书上有写到，官方网站也有 Hadoop 安装配置教程，但由于对 Linux 环境不熟悉，书上跟官网上简略的安装步骤新手往往 Hold 不住。加上网上不少教程也甚是坑，导致新手折腾老几天愣是没装好，很是打击学习热情。

本教程由厦门大学数据库实验室 (<http://dblab.xmu.edu.cn>)出品，转载请注明。本教程适合于原生 Hadoop 2，包括 Hadoop 2.6.0, Hadoop 2.7.1 等版本，主要参考了官方安装教程 (<http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html>)，步骤详细，辅以适当说明，相信按照步骤来，都能顺利安装并运行**Hadoop**。另外有Hadoop安装配置简略版 (<http://www.powerxing.com/install-hadoop-simplify/>)方便有基础的读者快速完成安装。此外，希望读者们能多去了解一些 Linux 的知识，以后出现问题时才能自行解决。

环境

本教程使用 **Ubuntu 14.04 64位** 作为系统环境（Ubuntu 12.04 也行，32位、64位均可），请自行安装系统（可参考使用VirtualBox安装Ubuntu (<http://dblab.xmu.edu.cn/blog/337-2/>)）。

如果用的是 CentOS/RedHat 系统，请查看相应的CentOS安装Hadoop教程_单机伪分布式配置 (<http://www.powerxing.com/install-hadoop-in-centos/>)。

本教程基于原生 Hadoop 2，在 **Hadoop 2.6.0 (stable)** 版本下验证通过，可适合任何 Hadoop 2.x.y 版本，例如 Hadoop 2.4.1。

Hadoop版本

Hadoop 有两个主要版本，Hadoop 1.x.y 和 Hadoop 2.x.y 系列，比较老的教材上用的可能是 0.20 这样的版本。Hadoop 2.x 版本在不断更新，本教程均可适用。如果需安装 0.20, 1.2.1这样的版本，本教程也可以作为参考，主要差别在于配置项，配置请参考官网教程或其他教程。

新版是兼容旧版的，书上旧版本的代码应该能够正常运行（我自己没验证，欢迎验证反馈）。

装好了 Ubuntu 系统之后，在安装 Hadoop 前还需要做一些必备工作。

创建hadoop用户

如果你安装 CentOS 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
```

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

Ubuntu终端复制粘贴快捷键

在Ubuntu终端窗口中，复制粘贴的快捷键需要加上 shift，即粘贴是 ctrl+shift+v。

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
```

最后注销当前用户（点击屏幕右上角的齿轮，选择注销），在登陆界面使用刚创建的 hadoop 用户进行登陆。

更新apt

用 hadoop 用户登录后，我们先更新一下 apt，后续我们使用 apt 安装软件，如果没更新可能有一些软件安装不了。按 **ctrl+alt+t** 打开终端窗口，执行如下命令：

```
$ sudo apt-get update
```

若出现如下“Hash校验和不符合”的提示，可通过更改软件源来解决。若没有该问题，则不需要更改。

```
W: 无法下载 bzip2:/var/lib/apt/lists/partial/cn.archive.ubuntu.com_ubuntu_dists_
trusty-updates_universe_i18n_Translation-en Hash 校验和不符合
W: 无法下载 bzip2:/var/lib/apt/lists/partial/cn.archive.ubuntu.com_ubuntu_dists_
trusty-backports_universe_binary-amd64_Packages Hash 校验和不符合
E: Some index files failed to download. They have been ignored, or old ones used
instead.
hadoop@star-lab:~$
```

Ubuntu更新软件源时遇到Hash校验和不符合的问题

[点击查看：如何更改软件源](#)

后续需要更改一些配置文件，我比较喜欢用的是 vim（vi增强版，基本用法相同），建议安装一下（如果你实在还不会用 vi/vim 的，请将后面用到 vim 的地方改为 gedit，这样可以使用文本编辑器进行修改，并且每次文件更改完成后请关闭整个 gedit 程序，否则会占用终端）：

```
$ sudo apt-get install vim
```

安装软件时若需要确认，在提示处输入 y 即可。

```
hadoop@DBLab-XMU:~$ sudo apt-get install vim
正在读取软件包列表... 完成
正在分析软件包的依赖关系树
正在读取状态信息... 完成
将会安装下列额外的软件包：
  vim-runtime
建议安装的软件包：
  ctags vim-doc vim-scripts
下列【新】软件包将被安装：
  vim vim-runtime
升级了 0 个软件包，新安装了 2 个软件包，要卸载 0 个软件包，有 366 个软件包未被升
级。
需要下载 0 B/5,844 kB 的软件包。
解压缩后会消耗掉 28.0 MB 的额外空间。
您希望继续执行吗？ [Y/n]
```

通过命令行安装软件

安装SSH、配置SSH无密码登陆

集群、单节点模式都需要用到 SSH 登陆（类似于远程登陆，你可以登录某台 Linux 主机，并且在上面运行命令），Ubuntu 默认已安装了 SSH client，此外还需要安装 SSH server：

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

安装后，可以使用如下命令登陆本机：

```
$ ssh localhost
```

此时会有如下提示(SSH首次登陆提示)，输入 **yes** 。然后按提示输入密码 **hadoop**，这样就登陆到本机了。

```
hadoop@DBLab-XMU:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is a9:28:e0:4e:89:40:a4:cd:75:8f:0b:8b:57:79:67:86.
Are you sure you want to continue connecting (yes/no)? yes
```

SSH首次登陆提示

但这样登陆是需要每次输入密码的，我们需要配置成SSH无密码登陆比较方便。

首先退出刚才的 **ssh**，就回到了我们原先的终端窗口，然后利用 **ssh-keygen** 生成密钥，并将密钥加入到授权中：

```
$ exit                                # 退出刚才的 ssh localhost
$ cd ~/.ssh/                          # 若没有该目录，请先执行一次ssh loc
alhost
$ ssh-keygen -t rsa                   # 会有提示，都按回车就可以
$ cat ./id_rsa.pub >> ./authorized_keys # 加入授权
```

~的含义

在 Linux 系统中，~ 代表的是用户的主文件夹，即“/home/用户名”这个目录，如你的用户名为 **hadoop**，则 ~ 就代表“/home/hadoop/”。此外，命令中的 # 后面的文字是注释。

此时再用 **ssh localhost** 命令，无需输入密码就可以直接登陆了，如下图所示。

```
hadoop@DBLab-XMU: ~/.ssh$ ssh localhost
Welcome to Ubuntu 14.04.2 LTS (GNU/Linux 3.13.0-49-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

40 packages can be updated.
40 updates are security updates.

Last login: Thu Apr 30 21:20:50 2015 from localhost
hadoop@DBLab-XMU: ~$
```

@给力星
powerxing.com

SSH无密码登录

安装Java环境

Java环境可选择 Oracle 的 JDK，或是 OpenJDK，

按<http://wiki.apache.org/hadoop/HadoopJavaVersions>

(<http://wiki.apache.org/hadoop/HadoopJavaVersions>)中说的，新版本在 OpenJDK 1.7 下是没问题的。为图方便，这边直接通过命令安装 OpenJDK 7。

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

通过上述命令安装 OpenJDK，默认安装位置为 `/usr/lib/jvm/java-7-openjdk-amd64`（32位系统则是 `/usr/lib/jvm/java-7-openjdk-i386`，该路径可以通过执行 `dpkg -L openjdk-7-jdk | grep '/bin/javac'` 命令确定，执行后会输出一个路径，除去路径末尾的 `"/bin/javac"`，剩下的就是正确的路径了）。OpenJDK 安装后就可以直接使用 `java`、`javac` 等命令了。

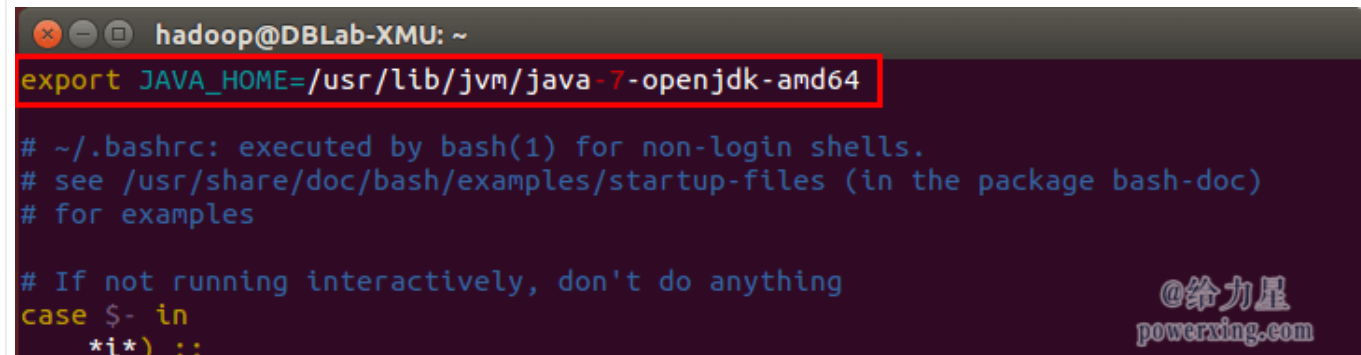
接着需要配置一下 `JAVA_HOME` 环境变量，为方便，我们在 `~/.bashrc` 中进行设置（扩展阅读: 设置Linux环境变量的方法和区别 (<http://www.powerxing.com/linux-environment-variable/>)）：

```
$ vim ~/.bashrc
```

在文件最前面添加如下单独一行（注意 = 号前后不能有空格），并保存：

```
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
```

如下图所示（该文件原本可能不存在，内容为空，这不影响）：



```
hadoop@DBLab-XMU: ~  
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64  
  
# ~/.bashrc: executed by bash(1) for non-login shells.  
# see /usr/share/doc/bash/examples/startup-files (in the package bash-doc)  
# for examples  
  
# If not running interactively, don't do anything  
case $- in  
*(*) ::
```

配置JAVA_HOME变量

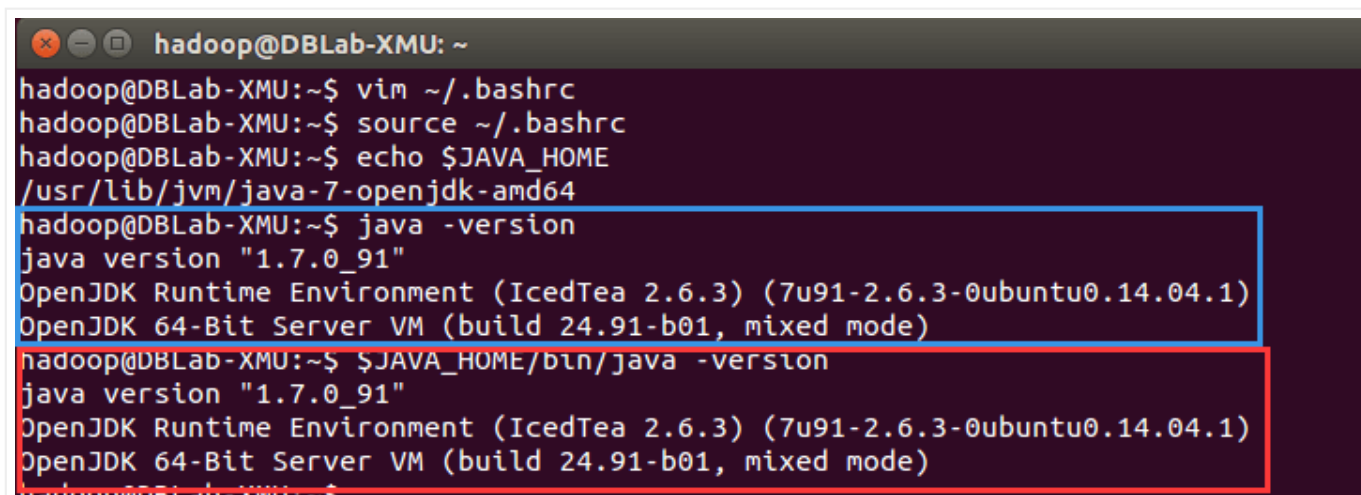
接着还需要让该环境变量生效，执行如下代码：

```
$ source ~/.bashrc    # 使变量设置生效
```

设置好后我们来检验一下是否设置正确：

```
$ echo $JAVA_HOME      # 检验变量值  
$ java -version  
$ $JAVA_HOME/bin/java -version  # 与直接执行 java -version 一样
```

如果设置正确的话，`$JAVA_HOME/bin/java -version` 会输出 java 的版本信息，且和 `java -version` 的输出结果一样，如下图所示：



```
hadoop@DBLab-XMU: ~  
hadoop@DBLab-XMU:~$ vim ~/.bashrc  
hadoop@DBLab-XMU:~$ source ~/.bashrc  
hadoop@DBLab-XMU:~$ echo $JAVA_HOME  
/usr/lib/jvm/java-7-openjdk-amd64  
hadoop@DBLab-XMU:~$ java -version  
java version "1.7.0_91"  
OpenJDK Runtime Environment (IcedTea 2.6.3) (7u91-2.6.3-0ubuntu0.14.04.1)  
OpenJDK 64-Bit Server VM (build 24.91-b01, mixed mode)  
hadoop@DBLab-XMU:~$ $JAVA_HOME/bin/java -version  
java version "1.7.0_91"  
OpenJDK Runtime Environment (IcedTea 2.6.3) (7u91-2.6.3-0ubuntu0.14.04.1)  
OpenJDK 64-Bit Server VM (build 24.91-b01, mixed mode)
```

成功配置JAVA_HOME变量

这样，Hadoop 所需的 Java 运行环境就安装好了。

安装 Hadoop 2

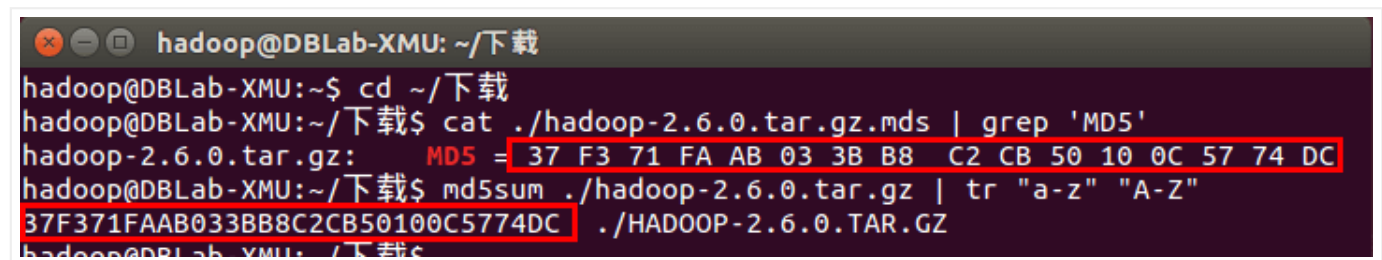
Hadoop 2 可以通过 <http://mirror.bit.edu.cn/apache/hadoop/common/> (<http://mirror.bit.edu.cn/apache/hadoop/common/>) 或者 <http://mirrors.cnnic.cn/apache/hadoop/common/> (<http://mirrors.cnnic.cn/apache/hadoop/common/>) 下载, 本教程选择的是 2.6.0 版本, 下载时请下载 **hadoop-2.x.y.tar.gz** 这个格式的文件, 这是编译好的, 另一个包含 **src** 的则是 Hadoop 源代码, 需要进行编译才可使用。

下载时强烈建议也下载 **hadoop-2.x.y.tar.gz.mds** 这个文件, 该文件包含了检验值可用于检查 **hadoop-2.x.y.tar.gz** 的完整性, 否则若文件发生了损坏或下载不完整, Hadoop 将无法正常运行。

本文涉及的文件均通过浏览器下载, 默认保存在“下载”目录中(若不是请自行更改 **tar** 命令的相应目录)。另外, 如果你用的不是 2.6.0 版本, 则将所有命令中出现的 2.6.0 更改为你所使用的版本。

```
$ cat ~/下载/hadoop-2.6.0.tar.gz.mds | grep 'MD5' # 列出md5检验值
$ # head -n 6 ~/下载/hadoop-2.7.1.tar.gz.mds # 2.7.1版本格式变了,
  可以用这种方式输出
$ md5sum ~/下载/hadoop-2.6.0.tar.gz | tr "a-z" "A-Z" # 计算md5
  值, 并转化为大写, 方便比较
```

若文件不完整则这两个值一般差别很大, 可以简单对比下前几个字符跟后几个字符是否相等即可, 如下图所示, 如果两个值不一样, 请务必重新下载。



```
hadoop@DBLab-XMU: ~/下载
hadoop@DBLab-XMU:~$ cd ~/下载
hadoop@DBLab-XMU:~/下载$ cat ./hadoop-2.6.0.tar.gz.mds | grep 'MD5'
hadoop-2.6.0.tar.gz:      MD5 = 37 F3 71 FA AB 03 3B B8 C2 CB 50 10 0C 57 74 DC
hadoop@DBLab-XMU:~/下载$ md5sum ./hadoop-2.6.0.tar.gz | tr "a-z" "A-Z"
37F371FAAB033BB8C2CB50100C5774DC ./HADOOP-2.6.0.TAR.GZ
hadoop@DBLab-XMU:~/下载$
```

检验文件完整性

我们选择将 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had
  oop
$ 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
```

相对路径与绝对路径的区别

请务必注意命令中的相对路径与绝对路径，本文后续出现的

`./bin/...`，`./etc/...` 等包含 `./` 的路径，均为相对路径，以 `/usr/local/hadoop` 为当前目录。例如在 `/usr/local/hadoop` 目录中执行 `./bin/hadoop version` 等同于执行 `/usr/local/hadoop/bin/hadoop version`。可以将相对路径改成绝对路径来执行，但如果你是在主文件夹 `~` 中执行 `./bin/hadoop version`，执行的会是 `/home/hadoop/bin/hadoop version`，就不是我们所想要的了。

Hadoop单机配置(非分布式)

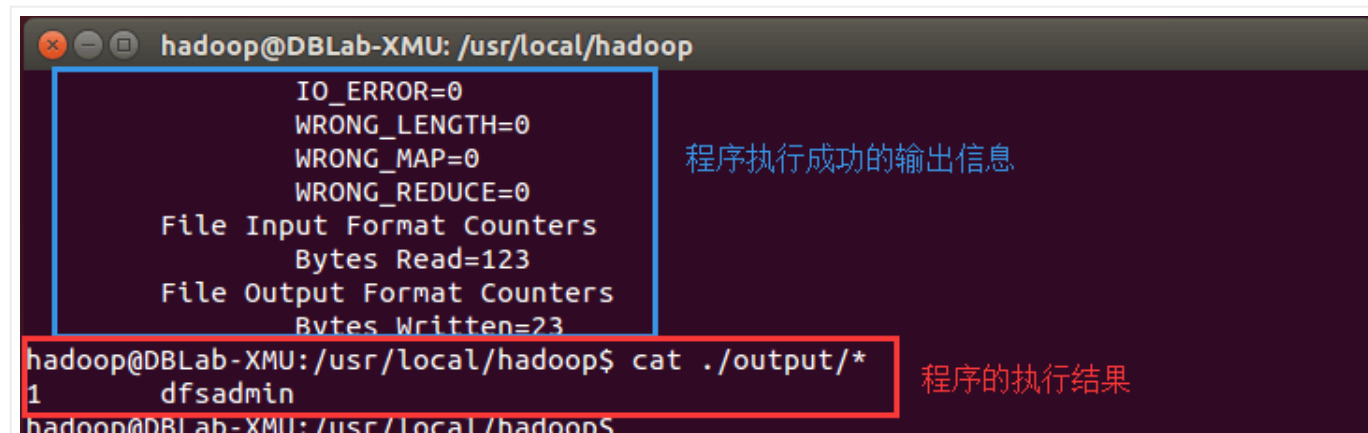
Hadoop 默认模式为非分布式模式，无需进行其他配置即可运行。非分布式即单 Java 进程，方便进行调试。

现在我们可以执行例子来感受下 Hadoop 的运行。Hadoop 附带了丰富的例子（运行 `./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar` 可以看到所有例子），包括 `wordcount`、`terasort`、`join`、`grep` 等。

在此我们选择运行 `grep` 例子，我们将 `input` 文件夹中的所有文件作为输入，筛选当中符合正则表达式 `dfs[a-z.]` 的单词并统计出现的次数，最后输出结果到 `output` 文件夹中。

```
$ cd /usr/local/hadoop
$ mkdir ./input
$ cp ./etc/hadoop/*.xml ./input # 将配置文件作为输入文件
$ ./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-mapreduce-examples-*.jar grep ./input ./output 'dfs[a-z.]+'
$ cat ./output/* # 查看运行结果
```

执行成功后如下所示，输出了作业的相关信息，输出的结果是符合正则的单词 **dfsadmin** 出现了1次



```
hadoop@DBLab-XMU: /usr/local/hadoop
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=123
File Output Format Counters
  Bytes Written=23
hadoop@DBLab-XMU: /usr/local/hadoop$ cat ./output/*
1      dfsadmin
hadoop@DBLab-XMU: /usr/local/hadoop$
```

Hadoop单机模式运行grep的输出结果

注意，Hadoop 默认不会覆盖结果文件，因此再次运行上面实例会提示出错，需要先将 `./output` 删除。

```
$ rm -r ./output
```

Hadoop伪分布式配置

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。

修改配置文件 **core-site.xml** (通过 gedit 编辑会比较方便: `gedit ./etc/hadoop/core-site.xml`)，将当中的

```
<configuration>
</configuration>
```

修改为下面配置：

```
<configuration>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>file:/usr/local/hadoop/tmp</value>
    <description>Abase for other temporary directories.</de
scription>
  </property>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
  </property>
</configuration>
```

同样的，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>
```

配置完成后，执行 NameNode 的格式化：

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

成功的话，会看到 “successfully formatted” 和 “Exitting with status 0” 的提示，若为 “Exitting with status 1” 则是出错。

```
hadoop@DBLab-XMU: /usr/local/hadoop
15/12/17 18:35:26 INFO namenode.FSImage: Allocated new BlockPoolId: BP-965227428-127.0.1.1-1450348526600
15/12/17 18:35:26 INFO common.Storage: Storage directory /usr/local/hadoop/tmp/dfs/name has been successfully formatted.
15/12/17 18:35:27 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
15/12/17 18:35:27 INFO util.ExitUtil: Exiting with status 0
15/12/17 18:35:27 INFO namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at DBLab-XMU/127.0.1.1
*****/
hadoop@DBLab-XMU: /usr/local/hadoop$
```

执行namenode格式化

注意

在这一步时若提示 **Error: JAVA_HOME is not set and could not be found.** 的错误，则需要在文件 `./etc/hadoop/hadoop-env.sh` 中设置 `JAVA_HOME` 变量，即在该文件中找到：

```
export JAVA_HOME=${JAVA_HOME}
```

将这一行改为JAVA安装位置：

```
export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64
```

再重新尝试格式化即可。

接着开启 NameNode 和 DataNode 守护进程。

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

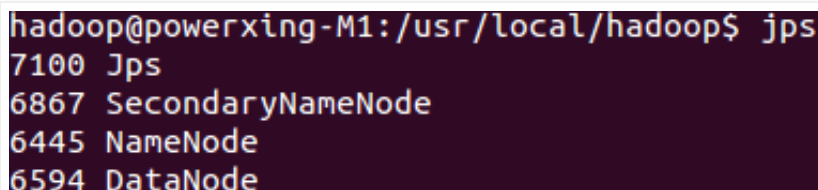
若出现如下SSH提示，输入yes即可。

```
hadoop@DBLab-XMU: /usr/local/hadoop$ sbin/start-dfs.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hadoop-na
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hadoop-da
Starting secondary namenodes [0.0.0.0]
The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.
ECDSA key fingerprint is a9:28:e0:4e:89:40:a4:cd:75:8f:0b:8b:57:79:67:86.
Are you sure you want to continue connecting (yes/no)? yes
```

启动Hadoop时的SSH提示

启动时可能会出现如下 WARN 提示：WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable WARN 提示可以忽略，并不会影响正常使用。

启动完成后，可以通过命令 `jps` 来判断是否成功启动，若成功启动则会列出如下进程：“NameNode”、“DataNode”和“SecondaryNameNode”（如果 SecondaryNameNode 没有启动，请运行 `sbin/stop-dfs.sh` 关闭进程，然后再次尝试启动尝试）。如果没有 NameNode 或 DataNode，那就是配置不成功，请仔细检查之前步骤，或通过查看启动日志排查原因。



```
hadoop@powerxing-M1:/usr/local/hadoop$ jps
7100 Jps
6867 SecondaryNameNode
6445 NameNode
6594 DataNode
```

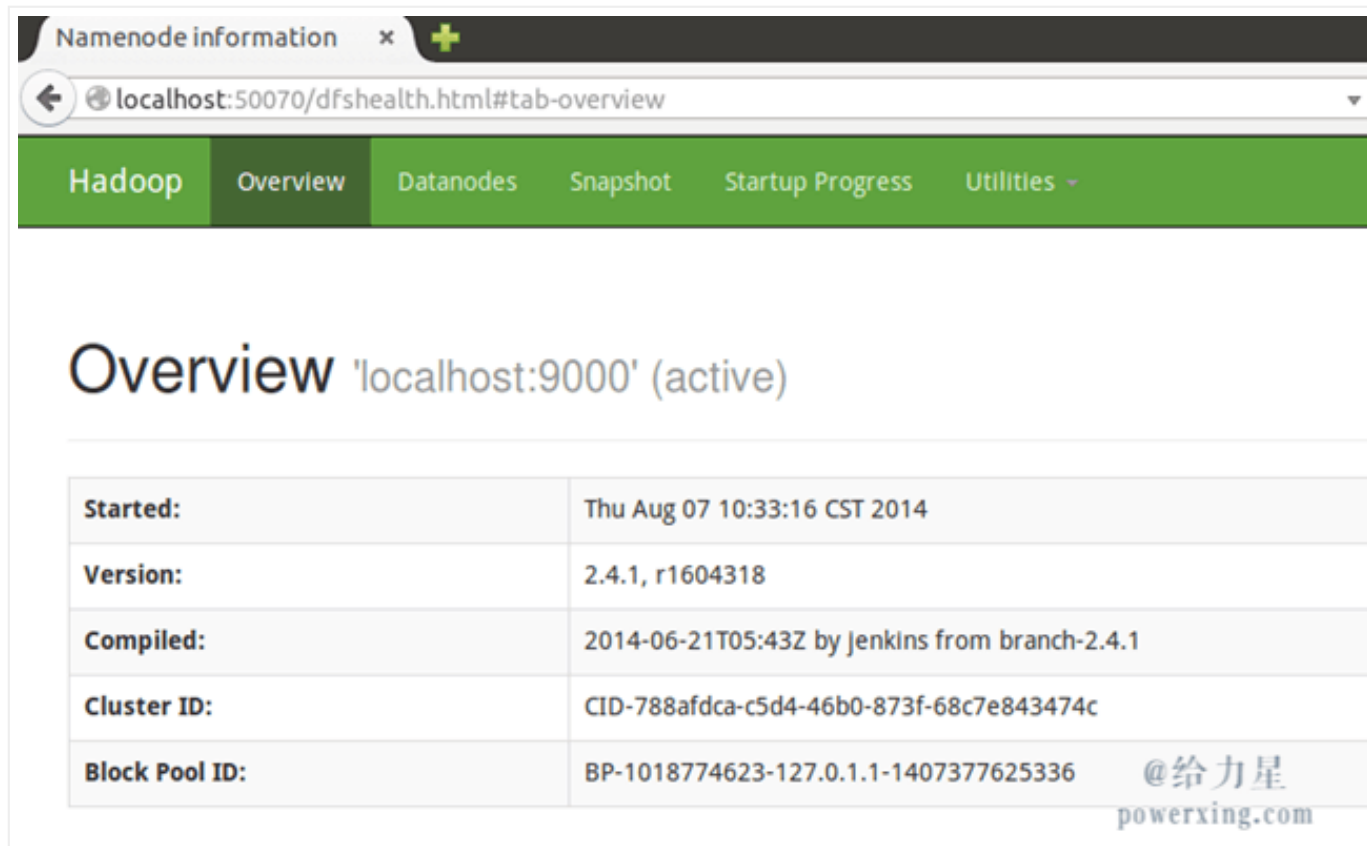
通过jps查看启动的Hadoop进程

通过查看启动日志分析启动失败原因

有时 Hadoop 无法正确启动，如 NameNode 进程没有顺利启动，这时可以查看启动日志来排查原因，注意几点：

- 启动时会提示形如 “DBLab-XMU: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hadoop-namenode-DBLab-XMU.out”，其中 DBLab-XMU 对应你的机器名，但其实启动日志信息是记录在 /usr/local/hadoop/logs/hadoop-hadoop-namenode-DBLab-XMU.log 中，所以应该查看这个后缀为 **.log** 的文件；
- 每一次的启动日志都是追加在日志文件之后，所以得拉到最后面看，看下记录的时间就知道了。
- 一般出错的提示在最后面，通常是写着 Fatal、Error 或者 Java Exception 的地方。
- 可以在网上搜索一下出错信息，看能否找到一些相关的解决方法。

成功启动后，可以访问 Web 界面 <http://localhost:50070> (<http://localhost:50070>) 查看 NameNode 和 Datanode 信息，还可以在线查看 HDFS 中的文件。



Hadoop的Web界面

运行Hadoop伪分布式实例

上面的单机模式，grep 例子读取的是本地数据，伪分布式读取的则是 HDFS 上的数据。要使用 HDFS，首先需要在 HDFS 中创建用户目录：

```
$ ./bin/hdfs dfs -mkdir -p /user/hadoop
```

接着将 `./etc/hadoop` 中的 xml 文件作为输入文件复制到分布式文件系统中，即将 `/usr/local/hadoop/etc/hadoop` 复制到分布式文件系统 `/user/hadoop/input` 中。我们使用的是 `hadoop` 用户，并且已创建相应的用户目录 `/user/hadoop`，因此在命令中就可以使用相对路径如 `input`，其对应的绝对路径就是 `/user/hadoop/input`：

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

复制完成后，可以通过如下命令查看文件列表：

```
$ ./bin/hdfs dfs -ls input
```

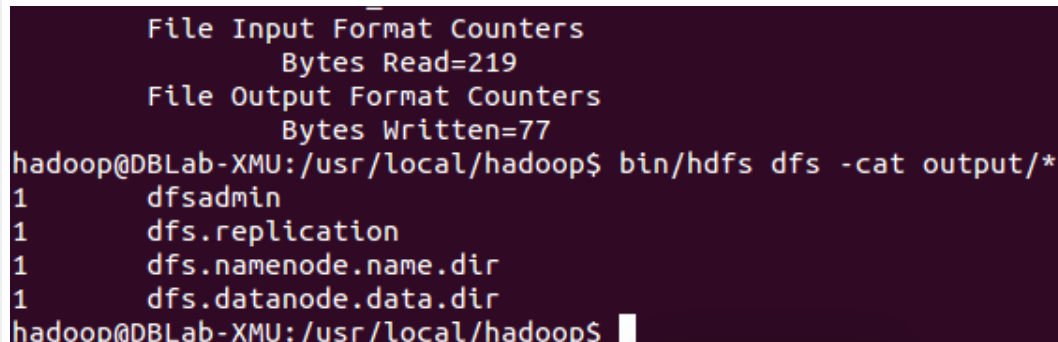

伪分布式运行 MapReduce 作业的方式跟单机模式相同，区别在于伪分布式读取的是HDFS中的文件（可以将单机步骤中创建的本地 input 文件夹，输出结果 output 文件夹都删掉来验证这一点）。

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

查看运行结果的命令（查看的是位于 HDFS 中的输出结果）：

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

结果如下，注意到刚才我们已经更改了配置文件，所以运行结果不同。



```
File Input Format Counters
  Bytes Read=219
File Output Format Counters
  Bytes Written=77
hadoop@DBLab-XMU:/usr/local/hadoop$ bin/hdfs dfs -cat output/*
1      dfsadmin
1      dfs.replication
1      dfs.namenode.name.dir
1      dfs.datanode.data.dir
hadoop@DBLab-XMU:/usr/local/hadoop$
```

The screenshot shows a terminal window with a dark background. It displays the output of the 'hdfs dfs -cat output/*' command, which lists four files: dfsadmin, dfs.replication, dfs.namenode.name.dir, and dfs.datanode.data.dir, each preceded by a '1'. Above the command, it shows 'File Input Format Counters' with 'Bytes Read=219' and 'File Output Format Counters' with 'Bytes Written=77'. A watermark '@给力星 powerxing.com' is visible in the bottom right corner of the terminal output.

Hadoop伪分布式运行grep结果

我们也可以将运行结果取回到本地：

```
$ rm -r ./output      # 先删除本地的 output 文件夹（如果存在）
$ ./bin/hdfs dfs -get output ./output      # 将 HDFS 上的 output
文件夹拷贝到本机
$ cat ./output/*
```

Hadoop 运行程序时，输出目录不能存在，否则会提示错误

“org.apache.hadoop.mapred.FileAlreadyExistsException: Output directory

hdfs://localhost:9000/user/hadoop/output already exists”，因此若要再次执行，需要执行如下命令删除 output 文件夹：

```
$ ./bin/hdfs dfs -rm -r output      # 删除 output 文件夹
```

运行程序时，输出目录不能存在

运行 Hadoop 程序时，为了防止覆盖结果，程序指定的输出目录（如 output）不能存在，否则会提示错误，因此运行前需要先删除输出目录。在实际开发应用程序时，可考虑在程序中加上如下代码，能在每次运行时自动删除输出目录，避免繁琐的命令行操作：

```
1. Configuration conf = new Configuration();
2. Job job = new Job(conf);
3.
4. /* 删除输出目录 */
5. Path outputPath = new Path(args[1]);
6. outputPath.getFileSystem(conf).delete(outputPath, true);
```

若要关闭 Hadoop，则运行

```
$ ./sbin/stop-dfs.sh
```

注意

下次启动 hadoop 时，无需进行 NameNode 的初始化，只需要运行
./sbin/start-dfs.sh 就可以！

启动YARN

（伪分布式不启动 YARN 也可以，一般不会影响程序执行）

有的读者可能会疑惑，怎么启动 Hadoop 后，见不到书上所说的 JobTracker 和 TaskTracker，这是因为新版的 Hadoop 使用了新的 MapReduce 框架（MapReduce V2，也称为 YARN，Yet Another Resource Negotiator）。

YARN 是从 MapReduce 中分离出来的，负责资源管理与任务调度。YARN 运行于 MapReduce 之上，提供了高可用性、高扩展性，YARN 的更多介绍在此不展开，有兴趣的可查阅相关资料。

上述通过 `./sbin/start-dfs.sh` 启动 Hadoop，仅仅是启动了 MapReduce 环境，我们可以启动 YARN，让 YARN 来负责资源管理与任务调度。

首先修改配置文件 **mapred-site.xml**，这边需要先进行重命名：

```
$ mv ./etc/hadoop/mapred-site.xml.template ./etc/hadoop/mapred-site.xml
```

然后再进行编辑，同样使用 `gedit` 编辑会比较方便些 `gedit ./etc/hadoop/mapred-site.xml`：

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
```

接着修改配置文件 **yarn-site.xml**：

```
<configuration>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

然后就可以启动 YARN 了（需要先执行过 `./sbin/start-dfs.sh`）：

```
$ ./sbin/start-yarn.sh          $ 启动YARN
$ ./sbin/mr-jobhistory-daemon.sh start historyserver # 开启历史
服务器，才能在Web中查看任务运行情况
```

开启后通过 `jps` 查看，可以看到多了 NodeManager 和 ResourceManager 两个后台进程，如下图所示。

```
hadoop@DBLab-XMU: /usr/local/hadoop
hadoop@DBLab-XMU:/usr/local/hadoop$ ./sbin/start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hadoop-resource
manager-DBLab-XMU.out
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hadoop-n
odemanager-DBLab-XMU.out
hadoop@DBLab-XMU:/usr/local/hadoop$ jps
13519 NameNode
13829 SecondaryNameNode
15284 NodeManager
15317 Jps
13633 DataNode
15163 ResourceManager
```

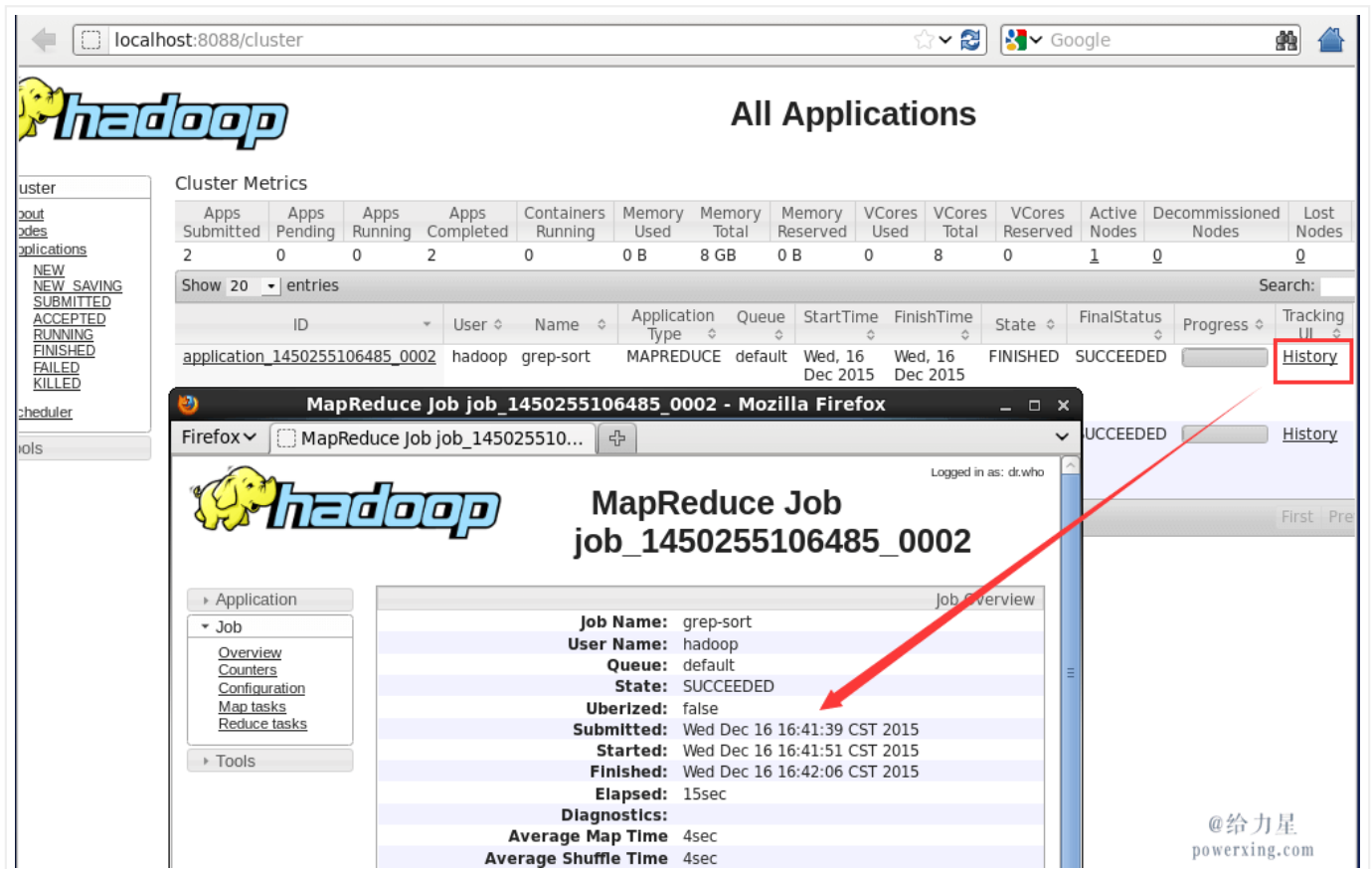
启动YARN的输出信息

成功启动后多了 NodeManager 和 ResourceManager

@给力星
powerxing.com

开启YARN

启动 YARN 之后，运行实例的方法还是一样的，仅仅是资源管理方式、任务调度不同。观察日志信息可以发现，不启用 YARN 时，是“mapred.LocalJobRunner”在跑任务，启用 YARN 之后，是“mapred.YARNRunner”在跑任务。启动 YARN 有个好处是可以通过 Web 界面查看任务的运行情况：<http://localhost:8088/cluster> (<http://localhost:8088/cluster>)，如下图所示。



开启YARN后可以查看任务运行信息

但 YARN 主要是为集群提供更好的资源管理与任务调度，然而这在单机上体现不出价值，反而会使程序跑得稍慢些。因此在单机上是否开启 YARN 就看实际情况了。

不启动 YARN 需重命名 `mapred-site.xml`

如果不想启动 YARN，务必把配置文件 `mapred-site.xml` 重命名，改成 `mapred-site.xml.template`，需要用时改回来就行。否则在该配置文件存在，而未开启 YARN 的情况下，运行程序会提示 “Retrying connect to server: 0.0.0.0/0.0.0.0:8032” 的错误，这也是为何该配置文件初始文件名为 `mapred-site.xml.template`。

同样的，关闭 YARN 的脚本如下：

```
$ ./sbin/stop-yarn.sh
$ ./sbin/mr-jobhistory-daemon.sh stop historyserver
```

自此，你已经掌握 Hadoop 的配置和基本使用了。

附加教程：配置 PATH 环境变量

在这里额外讲一下 PATH 这个环境变量（可执行 `echo $PATH` 查看，当中包含了多个目录）。例如我们在主文件夹 `~` 中执行 `ls` 这个命令时，实际执行的是 `/bin/ls` 这个程序，而不是 `~/ls` 这个程序。系统是根据 PATH 这个环境变量中包含的目录位置，逐一进行查找，直至在这些目录位置下找到匹配的程序（若没有匹配的则提示该命令不存在）。

上面的教程中，我们都是先进入到 `/usr/local/hadoop` 目录中，再执行 `sbin/hadoop`，实际上等同于运行 `/usr/local/hadoop/sbin/hadoop`。我们可以将 Hadoop 命令的相关目录加入到 PATH 环境变量中，这样就可以直接通过 `start-dfs.sh` 开启 Hadoop，也可以直接通过 `hdfs` 访问 HDFS 的内容，方便平时的操作。

同样我们选择在 `~/.bashrc` 中进行设置（`vim ~/.bashrc`，与 `JAVA_HOME` 的设置相似），在文件最前面加入如下单独一行：

```
export PATH=$PATH:/usr/local/hadoop/sbin:/usr/local/hadoop/bin
```

添加后执行 `source ~/.bashrc` 使设置生效，生效后，在任意目录中，都可以直接使用 `hdfs` 等命令了，读者不妨现在就执行 `hdfs dfs -ls input` 查看 HDFS 文件试试看。

安装 Hadoop 集群

在平时的学习中，我们使用伪分布式就足够了。如果需要安装 Hadoop 集群，请查看 Hadoop集群安装配置教程 (<http://www.powerxing.com/install-hadoop-cluster/>)。

相关教程

- 使用Eclipse编译运行MapReduce程序 (<http://www.powerxing.com/hadoop-build-project-using-eclipse/>): 使用 Eclipse 可以方便的开发、运行 MapReduce 程序，还可以直接管理 HDFS 中的文件。
- 使用命令行编译打包运行自己的MapReduce程序 (<http://www.powerxing.com/hadoop-build-project-by-shell/>): 有时候需要直接通过命令来编译、打包 MapReduce 程序。

参考资料

- <http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html> (<http://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html>)
- <http://www.cnblogs.com/xia520pi/archive/2012/05/16/2503949.html> (<http://www.cnblogs.com/xia520pi/archive/2012/05/16/2503949.html>)
- <http://www.micmiu.com/bigdata/hadoop/hadoop-2x-ubuntu-build/> (<http://www.micmiu.com/bigdata/hadoop/hadoop-2x-ubuntu-build/>)

🔗 <http://www.powerxing.com/install-hadoop/> (<http://www.powerxing.com/install-hadoop/>)

📖 笔记 (<http://www.powerxing.com/notes/>) 🔖

Hadoop (<http://www.powerxing.com/tag/hadoop/>)

相关文章

- **Hadoop集群安装配置教程_Hadoop2.6.0_Ubuntu/CentOS**
(<http://www.powerxing.com/install-hadoop-cluster/>)
- **使用命令行编译打包运行自己的MapReduce程序 Hadoop2.4.1**
(<http://www.powerxing.com/hadoop-build-project-by-shell/>)
- **使用Eclipse编译运行MapReduce程序 Hadoop2.6.0_Ubuntu/CentOS**
(<http://www.powerxing.com/hadoop-build-project-using-eclipse/>)
- **Hadoop-in-Practice第四章MapReduce-Join代码运行问题**

(<http://www.powerxing.com/hadoop-in-practice-joins/>)

➤ Hadoop安装配置简略教程 (<http://www.powerxing.com/install-hadoop-simplify/>)

➤ Hadoop安装教程_单机/伪分布式配置_CentOS6.4/Hadoop2.6.0

(<http://www.powerxing.com/install-hadoop-in-centos/>)

136 条评论

最新 最早 最热



Blue

很不错的Hadoop教程，按照步骤来很顺利的就装好了。

1月6日 回复 顶 转发



ss

真的是最好的教程

1月4日 回复 顶 转发



peter

我见过的最给力的教程，感谢给力兄，对我帮助很大

2015年12月23日 回复 顶 转发



xuanyeyu

我为什么总是提示找不到.bin/hdfs :not found

2015年12月21日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

1. 是 ./bin/hdfs 2. 注意区分相对路径与绝对路径，教程中类似 ./bin/hdfs 、 ./etc/hadoop 的路径均是相对路径，需要在 /usr/local/hadoop 目录下才正确（文章中有说明这一点的，请仔细阅读）。

2015年12月21日 回复 顶 转发



wantrd

楼主，你这个hadoop2.6 伪分布配好后为什么只有 SecondaryNameNode DataNode DataNode 没有 JobTracker 和 TaskTracker 是版本问题吗？还有，我配好了伪分布后，使用eclipse做开发环境，为什么在web上监控不到我的job，而且我观察控制台有“ mapred.Task: Task 'attempt_local1469050249_0001_m_000000_0' done.” 这样的语句输出，这个“local”感觉是本地模式，我是按照 Run Configurations 方式运行的。这样运行算是伪分布吗？谢谢

2015年12月13日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

新版 Hadoop 使用了 MapReduce V2 架构，也叫做 YARN，跟以往的不一样，你可以自己去了解下。要在 Web 上监控 job，需要开启 YARN 和 historyserver。伪分布式是只有一台机器，同时运行 namenode 跟 datanode，并且访问的是 HDFS 上的文件。

2015年12月17日 回复 顶 转发


```
at
org.apache.hadoop.fs.shell.CommandWithDestination.copyStreamToTarget(CommandWithDestin
ation.java:391)
at
org.apache.hadoop.fs.shell.CommandWithDestination.copyFileToTarget(CommandWithDestinati
on.java:328)
at
org.apache.hadoop.fs.shell.CommandWithDestination.processPath(CommandWithDestination.ja
va:263)
at
org.apache.hadoop.fs.shell.CommandWithDestination.processPath(CommandWithDestination.ja
va:248)
at org.apache.hadoop.fs.shell.Command.processPaths(Command.java:306)
at org.apache.hadoop.fs.shell.Command.processPathArgument(Command.java:278)
at
org.apache.hadoop.fs.shell.CommandWithDestination.processPathArgument(CommandWithDes
tination.java:243)
at org.apache.hadoop.fs.shell.Command.processArgument(Command.java:260)
at org.apache.hadoop.fs.shell.Command.processArguments(Command.java:244)
at
org.apache.hadoop.fs.shell.CommandWithDestination.processArguments(CommandWithDestina
tion.java:220)
at
org.apache.hadoop.fs.shell.CopyCommands$Put.processArguments(CopyCommands.java:267)
at org.apache.hadoop.fs.shell.Command.processRawArguments(Command.java:190)
at org.apache.hadoop.fs.shell.Command.run(Command.java:154)
at org.apache.hadoop.fs.FsShell.run(FsShell.java:287)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:70)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:84)
at org.apache.hadoop.fs.FsShell.main(FsShell.java:340)
```

请问，这怎么解决？

另外我在ubuntu 14.04 64位上安装hadoop 2.6.2伪分布式版，却没有出现这样的问题。

2015年12月4日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 这个好像是JDK的问题，你可以网上搜索看看，我也不是很清楚，例如
<http://tieba.baidu.com/p/518443482> 16楼提到的原因与解决方法。

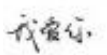
2015年12月4日 回复 顶 转发



zestinc

教程非常给力， thanks

2015年12月4日 回复 顶 转发



SXT加油 (<http://weibo.com/3969591038>)

运行jps命令，datanode没有啊，你这是只需要配置本身自己电脑就够了吗

(<http://weibo.com/3969591038>) 2015年12月3日 回复 顶 转发

给力星 (<http://www.powerxing.com/>)



是的，这是伪分布式配置，是在单机上运行的。建议再检查一下步骤。

(<http://www.powerxing.com/>) 2015年12月8日 回复 顶 转发



CX

回复 给力星: 这是说伪分布式的设置中，运行jps命令，没有出现datanode是正常的吗

2015年12月13日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 回复 CX: 是只需要配置本身自己的电脑。在伪分布式运行模式下，没有出现 datanode 说明配置有错，程序是跑不起来的。

2015年12月15日 回复 顶 转发



陈小予 (<http://www.diefishfish.com/?ds>)

过来踩踩

(<http://www.diefishfish.com/?ds>) 2015年11月30日 回复 顶 转发



gxz

```
hadoop@ubuntu:/usr/local/hadoop$ bin/hdfs dfs -ls input
```

OpenJDK Client VM warning: You have loaded library

/usr/local/hadoop/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.

It's highly recommended that you fix the library with 'execstack -c ', or link it with '-z noexecstack'.

15/11/28 18:51:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

虽然楼主说警报并影响正常使用，但是我想查看hdfs中的文件，它却只显示这个警报，而并没有显示给我hdfs中有些什么，我该怎么办呢？ 460087935@qq.com

2015年11月29日 回复 顶 转发



gxz

```
hadoop@ubuntu:/usr/local/hadoop$ bin/hdfs dfs -cat output/*
```

OpenJDK Client VM warning: You have loaded library

/usr/local/hadoop/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.

It's highly recommended that you fix the library with 'execstack -c ', or link it with '-z noexecstack'.

15/11/28 18:58:46 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

输入这个cat命令也是一样，它只显示警报，而并不显示 cat出来的结果。此时我的心是崩溃的。

2015年11月29日 回复 顶 转发



gxz

回复 gxz:

```
hadoop@ubuntu:/usr/local/hadoop/etc/hadoop$ hdfs dfs -rm -r /user/hadoop/output
```


Exception in thread "main" java.lang.RuntimeException: core-site.xml not found
at org.apache.hadoop.conf.Configuration.loadResource(Configuration.java:2566)
at org.apache.hadoop.conf.Configuration.loadResources(Configuration.java:2492)
at org.apache.hadoop.conf.Configuration.getProps(Configuration.java:2405)

```
at org.apache.hadoop.conf.Configuration.set(Configuration.java:1143)
at org.apache.hadoop.conf.Configuration.set(Configuration.java:1115)
at org.apache.hadoop.conf.Configuration.setBoolean(Configuration.java:1451)
at
org.apache.hadoop.util.GenericOptionsParser.processGeneralOptions(GenericOptionsParser.java:321)
at
org.apache.hadoop.util.GenericOptionsParser.parseGeneralOptions(GenericOptionsParser.java:487)
at org.apache.hadoop.util.GenericOptionsParser.(GenericOptionsParser.java:170)
at org.apache.hadoop.util.GenericOptionsParser.(GenericOptionsParser.java:153)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:64)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:84)
at org.apache.hadoop.fs.FsShell.main(FsShell.java:340)
而后再出现了一个异常。唉~
```

2015年11月29日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

你试试 `bin/hdfs dfs -ls /input`("代表当前目录")
(<http://www.douban.com/people/103654556/>)

2015年11月30日 回复 顶 转发



zyx

回复 gxz: 1.Exception in thread "main" java.lang.RuntimeException: core-site.xml not found
2.节点都启动了吗?

2015年12月4日 回复 顶 转发



zz

请教下，按照上面配置，core-site.xml配置defaultFs为hdfs://main:9000(main 为我主机名)，为什么启动hadoop成功，能看到namenode,datanode等进程，日志也没打印错误，但是9000端口就是拒绝连接呢？我看了下，9000端口没在监听状态，知道的麻烦告知下，谢谢啦~~~~~

2015年11月25日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

将main改成localhost呢？localhost不就是指代了自己的电脑吗
(<http://www.douban.com/people/103654556/>)

2015年11月26日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

应为 `hdfs://localhost:9000`
(<http://www.powerxing.com/>)

2015年11月30日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

给力星，能不能建立一个关于hadoop学习的群，大家有什么问题可以一起讨论？意下如何
(<http://www.douban.com/people/103654556/>)

2015年11月23日 回复 顶 转发

给力星 (<http://www.powerxing.com/>)



很遗憾，虽然我研究生期间接触的是Hadoop相关的东西，但我校招找的还是前端方向的工作... 在还没毕业之前还会帮老师做些相关事情，等毕业上班了应该不会再有接触了...

(<http://www.powerxing.com/>)

2015年11月30日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

回复 给力星: 恩，好吧，不过不相关也可以了解一下，我觉得大数据这个方向还是很有发展前景的。一同学习，话说工作找到互联网公司了吗？

(<http://www.douban.com/people/103654556/>)

2015年11月30日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

回复 benjamin: 签了美团。

(<http://www.powerxing.com/>)

2015年11月30日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

回复 给力星: 恩，美团在厦门也有研发，厦门城市不错。美团同学也蛮多

(<http://www.douban.com/people/103654556/>)

2015年12月1日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

各位在启用分布式，使用命令 `sbin/start-dfs.sh` 时出现了什么问题吗？为什么我运行这个命令要等待很长的时间，老是卡到这个地方

```
The: ssh: Could not resolve hostname the: Name or service not known
highly: ssh: Could not resolve hostname highly: Name or service not known
have: ssh: Could not resolve hostname have: Name or service not known
It's: ssh: Could not resolve hostname it's: Name or service not known
recommended: ssh: Could not resolve hostname recommended: Name or service not known
fix: ssh: Could not resolve hostname fix: Name or service not known
link: ssh: Could not resolve hostname link: No address associated with hostname
with: ssh: Could not resolve hostname with: Name or service not known
with: ssh: Could not resolve hostname with: Name or service not known
or: ssh: Could not resolve hostname or: Name or service not known
which: ssh: connect to host which port 22: Connection refused
stack: ssh: connect to host stack port 22: Connection refused
might: ssh: connect to host might port 22: Connection refused
try: ssh: connect to host try port 22: Connection refused
VM: ssh: connect to host vm port 22: Connection refused
fix: ssh: connect to host fix port 22: Connection refused
OpenJDK: ssh: connect to host openjdk port 22: Connection refused
you: ssh: connect to host you port 22: Connection refused
the: ssh: connect to host the port 22: Connection refused
stack: ssh: connect to host stack port 22: Connection refused
library: ssh: connect to host library port 22: Connection refused
loaded: ssh: connect to host loaded port 22: Connection refused
```

2015年11月23日 回复 顶 转发



Wang

我也在这里卡住了 请问你的问题解决了没有

2015年11月26日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

回复 Wang: 没有啊, 现在开启不了HDFS伪分布模式, 我的邮箱llxjtu@126.com, 有什么问题可以一起交流下

2015年11月26日 回复 顶 转发



gxz

输入 yes就好了

2015年11月28日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

回复 gxz: 恩, 已经安装成功了

2015年11月29日 回复 顶 转发



benjamin (<http://www.douban.com/people/103654556/>)

请问各位, 为什么我配置ssh无密码登录不成功, 每次进入运行 ssh localhost命令, 都要输入hadoop用户的登录密码?

2015年11月22日 回复 顶 转发



daidaixiong

最棒的文档 🇨🇳 👍 👍

2015年11月14日 回复 顶 转发



hunter (<http://t.qq.com/taojiaen115>)

请问我在Red Hat 4.8.2-16 搭建了hdfs,在使用httpfs的时候, 日志中出现了

log4j:WARN No appenders could be found for logger

(org.apache.hadoop.util.Shell).

log4j:WARN Please initialize the log4j system properly.

log4j:WARN See <http://logging.apache.org/log4j/1.2/faq.html#noconfig> for more info.

请问这是为什么, 而且我需要把hadoop-env.sh 中的JAVA_HOME改成/home/java....jdk才能用, 是不是和这个问题有关联啊??

现在不管做什么操作json都会显示NullPointerException 求教求教

2015年11月11日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

log4j是用来记录日志信息的。你这两问题我不是很清楚, 得你自行搜索一下了...

2015年11月30日 回复 顶 转发



凯 (<http://t.qq.com/k729427690>)

你好: 我在做格式化时

hadoop@martinpoint4:~\$ bin/hdfs namenode -format

bash: bin/hdfs: 没有那个文件或目录

请问怎么解决呀? 谢谢!

2015年11月10日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

教程中涉及到的 `bin/hdfs ...`, `sbin/start-dfs.sh ...` 等命令, 都是以 `/usr/local/hadoop` 为当前目录来执行的。你把目录切换到 `/usr/local/hadoop` 下, 再去执行就好了。

2015年11月30日 回复 顶 转发



geek

同为厦大学生! 良心之作!

2015年11月7日 回复 顶 转发



同

给力兄, 我在`sbin/start-dfs.sh`这一步卡住了, 显示没有创建文件夹的权限, 比如`logs`文件夹, `namenode.out`

2015年11月4日 回复 顶 转发



同

因为普通用户没有权限, 但是如果加`sudo`的话, 又得`root`密码连接`ssh`

2015年11月4日 回复 顶 转发



同

回复 同: 细细看了一下, 发现漏了一步, 修改`hadoop`文件夹的权限, 不过还是请教那条命令的作用?

2015年11月4日 回复 顶 转发



CXY

为什么 `bin/hdfs dfs -mkdir input` 失败? `mkdir: `input': No such file or directory`

2015年10月18日 回复 顶 转发




给力星 (<http://www.powerxing.com/>)

在这一步之前你没有先执行 `bin/hdfs dfs -mkdir -p /user/hadoop` 吧。需要先创建 `hdfs` 上的用户目录。

2015年10月18日 回复 顶 转发



CXY

回复 给力星: 谢谢楼主呢~成功啦 

2015年10月19日 回复 顶 转发



dal

回复 给力星: 我以`root`身份执行`hdfs dfs -mkdir -p /user/hadoop`时, 就出现了`/user/hadoop: No such file or directory`
怎破?

2015年10月22日 回复 顶 转发



小样 (<http://t.qq.com/zhang1992511521>)

`bin/hdfs dfs -mkdir /input` 这样就对了
(<http://t.qq.com/zhang1992511521>)

2015年11月3日 回复 顶 转发



wang

whx@whx-desktop:/usr/local/hadoop\$ bin/hdfs dfs -put etc/hadoop/*.xml input
 15/10/17 21:31:52 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
 put: Call From whx-desktop/127.0.1.1 to localhost:9000 failed on connection exception:
 java.net.ConnectException: Connection refused; For more details see:
<http://wiki.apache.org/hadoop/ConnectionRefused>
 whx@whx-desktop:/usr/local/hadoop\$
 你能帮我解决一下不呢? 十分感谢
 email:wang010366@vip.qq.com

2015年10月17日 回复 顶 转发

孙毛毛蛋蛋 (<http://weibo.com/3608800131>)

(<http://weibo.com/3608800131>) WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable, 我也遇到这个问题了, 你的解决了吗? 肿么办? 万分感谢!!!

2015年11月19日 回复 顶 转发

鹏霖 (<http://t.qq.com/penglin5442>)

你好, 学长, 请问我在 `sudo sbin/start-dfs.sh`
 这步之后出现了
 Starting namenodes on [localhost]

root@localhost's password:

我该怎么办, 问题是出在哪里

注: 我刚开始学不到一个月, 麻烦您说的详细点, 谢谢 🐼🐼

2015年10月13日 回复 顶 转发



大大

应该是ssh没有配置好

2015年10月14日 回复 顶 转发



faker

不用加sudo, 因为是创建的用户是hadoop, 加sudo是用root身份

2015年10月20日 回复 顶 转发

鹏霖 (<http://t.qq.com/penglin5442>)

(<http://t.qq.com/penglin5442>) 回复大大就是这个原因 已经解决 谢谢 🐼

2015年10月21日 回复 顶 转发

鹏霖 (<http://t.qq.com/penglin5442>)

(<http://t.qq.com/penglin5442>) 回复faker是hadoop权限设置的问题 和 ssh 没配好 🐼

2015年10月21日 回复 顶 转发

visitor



wonderful tutorial! thanks for your help

2015年9月28日 回复 顶 转发



lq

比一些书上讲的好多了

2015年9月27日 回复 顶 转发



专攻鸡兔同笼等一系列学术问题 (<http://weibo.com/2657119114>)

(<http://weibo.com/2657119114>)

楼主你好，请问下我关掉终端，然后再重新打开终端输入sbin/start-dfs.sh，
结果提示-bash:/sbin/start-dfs.sh: No such file or directory这是为什么呢？
(我想关掉重新打开hadoop试试)

2015年9月21日 回复 顶 转发



专攻鸡兔同笼等一系列学术问题 (<http://weibo.com/2657119114>)

(<http://weibo.com/2657119114>) 哦，明白了，要先cd路径到hadoop的文件夹

2015年9月21日 回复 顶 转发



专攻鸡兔同笼等一系列学术问题 (<http://weibo.com/2657119114>)

成功了 谢谢楼主！！希望自己能坚持下去，能进一步学习

(<http://weibo.com/2657119114>) 2015年9月21日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 嗯，加油~

2015年9月21日 回复 顶 转发



victor

想請問一下，解壓縮後，執行cd ./hadoop 與 ./bin/hadoop 都顯示bash:./bin/hadoop:No such file or directory

2015年9月8日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 解压是解压到/usr/local 目录下，并且执行 cd ./hadoop 之前，还执行了如下操作，更改了解压后的文件名：

```
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
```

这样操作后，最终的目录是 /usr/local/hadoop，请自行检查。

2015年9月8日 回复 顶 转发



victor

回复 给力星: 感謝樓主，可以了!!

2015年9月8日 回复 顶 转发



在这一段“在这一步以及后面启动 Hadoop 时若提示 Error: JAVA_HOME is not set and could not be found. 的错误, 则需要在文件 ./etc/hadoop/hadoop-env.sh 。。。 ”中地址好像写错了, 根本没有那个文件应该是/usr/local/hadoop/etc/hadoop/hadoop-env.sh

2015年9月2日 回复 顶 转发



clgsnail

mapreduce能否直接读取本地目录文件?

2015年8月25日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

Hadoop单机配置(非分布式) 就是读取本地目录文件。
(<http://www.powerxing.com/>)

2015年8月25日 回复 顶 转发



欧祥钦 (<http://t.qq.com/ouxiangqin>)

用虚拟机安装的, 执行这句话的时候 bin/hdfs dfs -mkdir -p /user/hadoop , 出现OpenJDK Client VM warning: You have loaded library /usr/local/hadoop/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.

It's highly recommended that you fix the library with 'execstack -c ', or link it with '-z noexecstack'.

15/08/18 02:29:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

2015年8月18日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

WARN 提示不影响正常功能, 可以放心使用。若要消除该提醒, 可自行搜索网上解决方法~
(<http://www.powerxing.com/>)

2015年8月25日 回复 顶 转发



骆元 (<http://t.qq.com/luoyuan347184259>)

楼主你好, 按照你的教程安装了伪分布式后, localhost:50070页面能够正常显示结点信息。但运行mapreduce程序后, <http://localhost:8088/cluster> 却看不到运行程序的相关信息。。。这个问题困扰了我好几天, 请不吝赐教~

2015年7月30日 回复 顶(2) 转发



zz

楼主问题终于弄好了, hadoop2.6.0伪分布式运行mapreduce程序, 需要添加yarn-site.xml配置文件。

yarn-site.xml:

mapreduce.framework.nameyarn.nodemanager.aux-servicesmapreduce_shuffle

这样<http://localhost:8088/cluster> 才能看到运行程序的相关信息 😊

2015年7月31日 回复 顶 转发



zz

回复 zz: 参考网址: <http://www.centoscn.com/hadoop/2015/0118/4525.html>

2015年7月31日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 回复 zz: 目前单机教程只是涉及如何启动Hadoop, 还没有涉及通过 <http://localhost:8088/cluster> 查看 MapReduce 任务。谢谢你的反馈, 后续考虑补充上。

2015年8月4日 回复 顶 转发



王雪洁 (<http://t.qq.com/wangxuejie333>)

(<http://t.qq.com/wangxuejie333>) 楼主, jobtracker和tasktracker是旧版本Hadoop才有的。那现在这个mapreduce是怎么管理的那? 并且mapred-site.sh里面再配置jobtracker是无效的吧

2015年7月29日 回复 顶(2) 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 新版本采用的是Yarn框架, 你可以自行搜索一下相关介绍。新版没有jobtracker, 再去配置自然是无效的。

2015年7月29日 回复 顶 转发



贾睿 (<http://t.qq.com/jiarui19950205>)

真的很棒 谢谢你

(<http://t.qq.com/jiarui19950205>) 2015年7月27日 回复 顶 转发



贾睿 (<http://t.qq.com/jiarui19950205>)

(<http://t.qq.com/jiarui19950205>) 哈哈

2015年7月27日 回复 顶 转发



正在学习Hadoop的包包

补充一点点内容。第一次使用start-all.sh启动Hadoop时, 若出现报错: localhost:Error: JAVA_HOME is not set and could not be found.则是因为JAVA_HOME环境变量没有设置, 需要在/etc/hadoop/hadoop-env.sh中设置JAVA_HOME, export JAVA_HOME=\$JAVA_HOME 这样设置仍然会报错, 只有使用绝对路径才不会报错。如: export JAVA_HOME=/usr/lib/jvm/jdk1.8.0_51

2015年7月25日 回复 顶 转发



czb

我也是。

2015年7月25日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 估计是通过 ~/.bashrc 来设置 \$JAVA_HOME 还是不能保证有效, 之前是在 /etc/environment 中进行设置, 不过需要重启才能保证生效, 比较麻烦。我已在文章中作了相应补充。

2015年7月27日 回复 顶 转发

czb



每次重启ubuntu后, \$ echo \$HADOOP_HOME时发现都是空的, 现在只好每次重启后手工执行source命令生效:

```
$source /usr/local/hadoop/etc/hadoop/hadoop-env.sh
```

这应该属linux知识, 但我不知如何解决, 望各位给予指导。

2015年7月25日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>)

在启动Hadoop时会运行hadoop-env.sh这个脚本让\$HADOOP_HOME变量生效的, 应该是不需要你预先使其生效的。重启ubuntu后 echo \$HADOOP_HOME 为空是正常的, 不会影响启动。

2015年7月27日 回复 顶 转发



正在学习Hadoop的包包

赞! 这是我见过的最棒的安装文档! 很多小问题都想到了! 负责任的作者!

2015年7月25日 回复 顶 转发



liulifeng201 (<http://t.qq.com/liulifeng2010123>)

楼主的文档是有多给力我是不知道, 反正我是用上了。

谢谢楼主, 用其他的教程都快哭了, 楼主好人啊。真是感谢

(<http://t.qq.com/liulifeng2010123>)

2015年7月25日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>)

谢谢支持。我也是被其他教程坑过, 深感不靠谱, 才自己写了一个较为详细的。

2015年7月27日 回复 顶 转发



JRE不需要吗?

楼主, 有一个疑问, 不需要安装JRE了吗? 我不太懂openjdk的jvm

2015年7月22日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>)

请注意看安装Java环境那给出的命令: `sudo apt-get install openjdk-7-jre openjdk-7-jdk` 是有安装jre的

2015年7月22日 回复 顶 转发



czb

回复 给力星: 谢谢楼主, 单机版本搞定了。

楼主还可以提供其他例子供我们学习吗? 在hadoop只找到楼主文章所讲的这么一个例子, 其他跟mapreduce目录平行的如

hdfs等下面都没找到其他*example*.tar例子了。

2015年7月25日 回复 顶 转发



陈思桥 (<http://t.qq.com/xuanyuanqiao>)

很感谢楼主的教程使我顺利的用上了hadoop。但是有个问题要说下,

JAVA_HOME这个环境变量在/usr/local/hadoop/etc/hadoop/hadoop-env.sh 里面也得加上 `export JAVA_HOME=/usr/lib/jvm/java-7-openjdk-amd64`

不然的话namenode格式化的会出现JAVA_HOME not set 的问题

2015年7月12日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

谢谢反馈。hadoop-env.sh中对JAVA_HOME变量的设置为 `export JAVA_HOME=${JAVA_HOME}`，也就是说如果 \$JAVA_HOME 原本有设置好，那就不必再重复设置了。当然直接在 hadoop-env.sh 设置JAVA_HOME变量能保证可运行。

2015年7月12日 回复 顶 转发



陈思桥 (<http://t.qq.com/xuanyuanqiao>)

回复 给力星: 是的，我的hadoop必须设置，不知道其他人的怎么样。

2015年7月12日 回复 顶 转发



Hon

回复 陈思桥: 我也遇到了同样的情况，之前是在/etc/environment里添加的环境变量，一直提示 not set。只好也在hadoop-env.sh里加上了。

2015年7月25日 回复 顶 转发



czb

回复 陈思桥: 我也是

2015年7月25日 回复 顶 转发



小明

请问前辈：我安装到伪分布式那里，老是提示Java HotSpot(TM) Client VM warning: You have loaded library /usr/Hadoop/hadoop-2.7.0/lib/native/libhadoop.so.1.0.0 which might have disabled stack guard. The VM will try to fix the stack guard now.

然后bin/hdfs dfs -mkdir input 提示mkdir: `input': No such file or directory。机子是32位 ubuntu12.04.5, hadoop2.7.0

2015年7月10日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

warning不会影响正常执行的。mkdir遇到的错误，应该是你没有先执行"bin/hdfs dfs -mkdir -p /user/hadoop"吧。要去掉warn可以搜索一下网上的资料，例如可以在etc/hadoop/hadoop-env.sh 中，尝试加上如下两行：

```
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="-Djava.library.path=$HADOOP_HOME/lib"
```

2015年7月10日 回复 顶 转发



at

我的datanode 总是开不了
弄了下面几个日志：

```
2015-07-05 20:02:47,536 INFO
```

```
org.apache.hadoop.hdfs.server.blockmanagement.DatanodeManager:
```

```
dfs.namenode.datanode.registration.ip-hostname-check=true
```

```
2015-07-05 20:02:47,751 INFO org.apache.hadoop.hdfs.server.namenode.FSNamesystem:
```

```
dfs.namenode.safemode.min.datanodes = 0
```

```
2015-07-05 20:02:48,306 INFO org.apache.hadoop.hdfs.StateChange: STATE* Network topology
has 0 racks and 0 datanodes
```

```
2015-07-05 20:55:11,573 INFO
```

```
org.apache.hadoop.hdfs.server.blockmanagement.DatanodeManager:
dfs.namenode.datanode.registration.ip-hostname-check=true
2015-07-05 20:55:11,737 INFO org.apache.hadoop.hdfs.server.namenode.FSNamesystem:
dfs.namenode.safemode.min.datanodes = 0
2015-07-05 20:55:12,196 INFO org.apache.hadoop.hdfs.StateChange: STATE* Network topology
has 0 racks and 0 datanodes
2015-07-05 20:00:13,003 INFO
org.apache.hadoop.hdfs.server.blockmanagement.DatanodeManager:
dfs.namenode.datanode.registration.ip-hostname-check=true
2015-07-05 20:00:13,192 INFO org.apache.hadoop.hdfs.server.namenode.FSNamesystem:
dfs.namenode.safemode.min.datanodes = 0
2015-07-05 20:02:52,481 INFO
org.apache.hadoop.hdfs.server.blockmanagement.DatanodeManager:
dfs.namenode.datanode.registration.ip-hostname-check=true
2015-07-05 20:02:52,704 INFO org.apache.hadoop.hdfs.server.namenode.FSNamesystem:
dfs.namenode.safemode.min.datanodes = 0
2015-07-05 20:55:16,430 INFO
org.apache.hadoop.hdfs.server.blockmanagement.DatanodeManager:
dfs.namenode.datanode.registration.ip-hostname-check=true
2015-07-05 20:55:16,607 INFO org.apache.hadoop.hdfs.server.namenode.FSNamesystem:
dfs.namenode.safemode.min.datanodes = 0
```

2015年7月5日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 你看一下日志文件最后面的内容，看下有带 Fatal、Error 或者 Java Exception 的信息。

2015年7月6日 回复 顶 转发



xmu2012级

官方提供的预编译版本只有32bit的吧，上面的教程是基于32bit还是已经自己编译好了？

2015年6月29日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 直接用的官方编译版本。不管32位还是64位，安装配置过程都是一样的。

2015年6月29日 回复 顶 转发



uiero

唯一照着做成功安装的好贴！顶

2015年6月26日 回复 顶 转发



aaa

硬是把一个grep search说成word count。。。

2015年6月26日 回复 顶 转发



给力星 (<http://www.powerxing.com/>)

(<http://www.powerxing.com/>) 多谢提示。一直误以为 hadoop-mapreduce-examples-*.jar 是 word count 实例，只是还可以通过正则限定要统计的单词... 查阅了一下，examples.jar 里包含了很多例子。这两天有时间我再修正下教程。

2015年6月27日 回复 顶 转发

**梁佳宾** (<http://liangjiabin.com/>)

在用bin/hdfs操作时总显示如下警告是怎么回事？

(http://liangjiabin.com/)

WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

谢谢！

2015年6月23日 回复 顶 转发

**给力星** (<http://www.powerxing.com/>)

(http://www.powerxing.com/)

这个只是WARN，不会影响正常执行。你用的是2.4.1版本？换成2.6.0就没有这个问题了。

2015年6月24日 回复 顶 转发

**白然**

回复 给力星: Ubuntu下32位跟64位都没有这个告警吗？我用Cent OS 6.x 64位的，会有这个告警的？

2015年7月28日 回复 顶 转发

**郑士明** (<http://t.qq.com/shiming-beijing>)

我在 /etc/profile中 export

(http://t.qq.com/shiming-beijing)

```
PATH=":/usr/local/hadoop/sbin:/usr/local/hadoop/bin:$JAVA_HOME/bin:$JAVA_HOME/jre/bin:bin:$PATH" 然后zsm@zhengshiming-Aspire-E1-471G:~$ sbin/start-dfs.sh 提示bash: sbin/start-dfs.sh: 没有那个文件或目录，请问高手这是咋了？
```

2015年6月17日 回复 顶 转发

**给力星** (<http://www.powerxing.com/>)

(http://www.powerxing.com/)

首先，应该是 `PATH="/usr/local..."`，你的PATH中第一个冒号是多余的；然后，应该是直接运行 `start-dfs.sh`，前面不需要加目录，如果包含目录，运行的命令 `sbin/start-dfs.sh` 等同于 `./sbin/start-dfs.sh`。

2015年6月17日 回复 顶 转发

**给力星** (<http://www.powerxing.com/>)

(http://www.powerxing.com/)

另外，你可以运行命令 `echo $PATH`，检查一下目录是否正确。

2015年6月17日 回复 顶 转发

**郑士明** (<http://t.qq.com/shiming-beijing>)

(http://t.qq.com/shiming-beijing) 回复 给力星: 明白了！sbin/是多余的！谢谢了

2015年6月18日 回复 顶 转发

**游客**

mapred-site.xml的配置在2.0以后就木有了麼

2015年6月16日 回复 顶 转发

**给力星** (<http://www.powerxing.com/>)

2015年6月16日

回复

顶

转发



郑士明 (<http://t.qq.com/shiming-beijing>)

给力兄, 多谢你了, 我是你的忠实粉丝

(<http://t.qq.com/shiming-beijing>)

2015年6月15日

回复

顶

转发



给力星 (<http://www.powerxing.com/>)

谢谢支持~

(<http://www.powerxing.com/>)

2015年6月16日

回复

顶

转发



smile

不错。我按装这篇文章安装成功了

2015年6月15日

回复

顶

转发

社交帐号登录: 微信 微博 QQ 人人 [更多»](#)



说点什么吧...

发布

多说 (<http://duoshuo.com>)