

个人资料



ggzone

访问: 225240次

积分: 3780

等级: BLOG > 5

排名: 第4693名

原创: 156篇

转载: 13篇

译文: 0篇

评论: 71条

文章搜索

文章分类

C++ (14)

WTL (2)

boost (3)

vs (3)

opencv (3)

VC (10)

ubuntu (7)

FineUI (1)

Asp.net (3)

matlab (1)

STL-03 (12)

STL-0X (0)

CodeBlocks (2)

GCC (1)

Hadoop (11)

java (3)

SQL SERVER (12)

gSOAP (3)

C# (5)

WP (13)

networking (8)

OS (2)

cocos2d-x (1)

Lua (6)

SQL (1)

design-patterns (1)

Ubuntu14.04下hadoop-2.6.0单机配置和伪分布式配置

标签: [Ubuntu](#) [Hadoop](#) [配置](#)

2015-01-05 18:44 16835人阅读 评论

分类: [Hadoop \(10\)](#)

版权声明：本文为博主原创文章，未经博主允许不得转载。

目录(?)

注意：安装之前最好删除hadoop-2.6.0/dfs/data下的所有文件，避免出现各种问题。尤其安装包不是官方现下载的。  
安装错误记录：<http://blog.csdn.net/ggz631047367/article/details/49979641>  
需要重新编译的教程：<http://blog.csdn.net/ggz631047367/article/details/42460589>  
在Ubuntu下创建hadoop用户组和用户

hadoop的管理员最好就是以后要登录桌面环境运行eclipse的用户，否则后面会有拒绝读写的问题出现。当然不是也有办法办法解决。

1. 创建hadoop用户组;

[html]

01. sudo addgroup hadoop

2. 创建hadoop用户;

[html]

01. sudo adduser -ingroup hadoop hadoop

3. 给hadoop用户添加权限，打开/etc/sudoers文件;

[html]

01. sudo gedit /etc/sudoers

在root ALL=(ALL:ALL) ALL下添加hadoop ALL=(ALL:ALL) ALL.

在Ubuntu下安装JDK

具体见：<http://blog.csdn.net/ggz631047367/article/details/42366687>  
  
//JAVA\_HOME=/usr/lib/jvm/jdk1.8.0\_25

安装ssh服务

[html]

01. sudo apt-get install ssh openssh-server

建立ssh无密码登录本机

http://blog.csdn.net/ggz631047367/article/details/42426391

1/8

MongoDB (1)

Erlang (1)

Docker (5)

ansible (6)

Shell (2)

zabbix (5)

Hive (5)

Sqoop (5)

ZooKeeper (1)

Git (1)

Spark (6)

Django (1)

Thrift (1)

Python (1)

Redis (1)

ELK (6)

文章存档

2016年01月 (7)

2015年12月 (22)

2015年11月 (20)

2015年10月 (10)

2015年09月 (15)

展开

阅读排行

Ubuntu14.04下hadoop-2 (16834)

hadoop2.6.0的eclipse插 (15719)

vs2013配置opencv2.4.9 (11121)

vs2013错误解决方法 (7349)

VS2013配置WTL90\_414 (6306)

win8.1系统vs2013中boo (6128)

32位ubuntu14.04手动编 (4997)

Vector的小知识点 (4808)

WTL在Win8.1系统WM\_I (4780)

STL容器能力一览表和各 (4757)

评论排行

32位ubuntu14.04手动编 (20)

hadoop2.6.0的eclipse插 (15)

Ubuntu14.04下hadoop-2 (10)

vs2013配置opencv2.4.9 (5)

VS2013配置WTL90\_414 (5)

win8.1系统vs2013中boo (5)

vs2013错误解决方法 (4)

C++获取Lua全局变量和 (3)

xgCalendar在ASP.NET中 (2)

Win10 恢复后退键 (1)

推荐文章

\*机器学习与数据挖掘网上资源搜罗——良心推荐

\*架构设计：系统间通信（17）——服务治理与Dubbo 中篇（分析）

\*拉开大变革序幕（中）：Docker 场景化尝试

切换到hadoop用户，执行以下命令：

```
[html]
01. su - hadoop
```

ssh生成密钥有rsa和dsa两种生成方式，默认情况下采用rsa方式。

1. 创建ssh-key，， 这里我们采用rsa方式；

```
[html]
01. ssh-keygen -t rsa -P "" （注：回车后会在~/.ssh/下生成两个文件：id_rsa和id_rsa.pub这两个文件是成对出现的）
```

2. 进入~/.ssh/目录下，将id\_rsa.pub追加到authorized\_keys授权文件中，开始是没有authorized\_keys文件的；

```
[html]
01. cd ~/.ssh
02. cat id_rsa.pub >> authorized_keys （完成后就可以无密码登录本机了。）
```

3. 登录localhost;

```
[html]
01. ssh localhost
```

4. 执行退出命令;

```
[html]
01. exit
```

安装hadoop

下载地址：<http://apache.fayea.com/hadoop/common/stable/hadoop-2.6.0.tar.gz>

因多次修改，可能会有Hadoop实际路径和配置文件中Hadoop路径不同，可能会产生错误

1. 把hadoop解压到/usr/local下：

```
[html]
01. sudo tar -zxvf hadoop-2.6.0.tar.gz
02. sudo mv hadoop-2.6.0 /usr/local/hadoop
03. sudo chmod -R 775 /usr/local/hadoop
04. sudo chown -R hadoop:hadoop /usr/local/hadoop //否则ssh会拒绝访问
```

2.配置

修改bashrc的配置：

```
[html]
01. sudo gedit ~/.bashrc
```

在文件末尾添加：

```
[html]
01. #HADOOP VARIABLES START
02.
03. export JAVA_HOME=/usr/lib/jvm/jdk1.8.0_25
04.
05. export HADOOP_INSTALL=/usr/local/hadoop
06.
07. export PATH=$PATH:$HADOOP_INSTALL/bin
08. export PATH=$PATH:$JAVA_HOME/bin
09. export PATH=$PATH:$HADOOP_INSTALL/sbin
10. export HADOOP_MAPRED_HOME=$HADOOP_INSTALL
11. export HADOOP_COMMON_HOME=$HADOOP_INSTALL
```

\*Android应用开发allowBackup敏感信息泄露的一点反思

\*Linux多线程实践（四） 线程的特定数据

\*Android点击Button水波纹效果

最新评论

- 32位ubuntu14.04手动编译hadoc  
ggzone: @zhangxueyang1:时间太长我也记不清怎么解决出错问题了，但我有编译好的http://pa...
- 32位ubuntu14.04手动编译hadoc  
zhangxueyang1: ./hadoop-project/target/antrun/build-main.xml (No s...
- 32位ubuntu14.04手动编译hadoc  
zhangxueyang1: 官网都是64位的，好像之前的32为全都给更没了
- C++获取Lua全局变量和执行Lua  
ggzone: @ggz631047367:如果不是你写的lua脚本，最起码别人交给你要有个文档说明吧。
- C++获取Lua全局变量和执行Lua  
ggzone: @tjyuanxi:我不太清除你什么意思，如果lua脚本是你写的话，那你肯定知道返回几个值。如果你说...
- C++获取Lua全局变量和执行Lua  
tjyuanxi: c调用lua后，怎么知道lua有几个返回值？如果不知道返回值，当然也就不知道怎么取返回值了。
- 32位ubuntu14.04手动编译hadoc  
ggzone: @jileidilei:你jdk是不是用的jdk1.8.0，我当时也不行，切换到1.7编译成功的
- 32位ubuntu14.04手动编译hadoc  
狂矢追夕: 我的别的都没问题了，就是最后进行编译的时候openjdk报错说找不到lib目录下面的一个jar包，我...
- Ubuntu14.04下hadoop-2.6.0单  
cuilll: @a1006570862:不用重新编译，下载64位的native覆盖32位的native就可以了参考...
- Win10 恢复后退键  
倒霉蛋儿: 也是，win10又不是没有移动端，把这个直接去掉放在扩展里面，巨硬实在有点坑

```
12. export HADOOP_HDFS_HOME=$HADOOP_INSTALL
13. export YARN_HOME=$HADOOP_INSTALL
14. export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_I
15. export HADOOP_OPTS="-
Djava.library.path=$HADOOP_INSTALL/lib"
16. #HADOOP VARIABLES END
```

如果不知道JAVA\_HOME可以通过命令获得：

```
[html]
01. update-alternatives --config java
```

目录取到java根目录即可。

执行下面命令使改动生效：

```
[html]
01. source ~/.bashrc
```

修改hadoop-env.sh的配置：

```
[html]
01. sudo gedit /usr/local/hadoop/etc/hadoop/hadoo
env.sh
```

找到JAVA\_HOME改为上面的值。

测试

- 通过执行hadoop自带实例WordCount验证是否安装成功

/usr/local/hadoop路径下创建input文件夹

```
[html]
01. mkdir input
02. cp README.txt input
```

在hadoop目录下执行WordCount：

```
[html]
01. bin/hadoop jar share/hadoop/mapreduce/sources/hadoop-mapreduce-
examples-2.6.0-sources.jar
02. org.apache.hadoop.examples.WordCount input output
```

Hadoop伪分布式配置

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

```
[html]
01. <configuration>
02.   <property>
03.     <name>hadoop.tmp.dir</name>
04.     <value>/usr/local/hadoop/tmp</value>
05.     <description>Abase for other temporary directories.
</description>
06.   </property>
07.   <property>
08.     <name>fs.defaultFS</name>
09.     <value>hdfs://localhost:9000</value>
10.   </property>
```

```
11. </configuration>
```

```
sudo gedit /usr/local/hadoop/etc/hadoop/yarn-site.xml
```

```
[html]
01. <property>
02.     <name>yarn.resourcemanager.hostname</name>
03.     <value>master</value>
04. </property>
05. <property>
06.     <description>The address of the applications manager interface :
07.     <name>yarn.resourcemanager.address</name>
08.     <value>${yarn.resourcemanager.hostname}:8032</value>
09. </property>
10. <property>
11.     <description>The address of the scheduler interface.
12.     <name>yarn.resourcemanager.scheduler.address</name>
13.     <value>${yarn.resourcemanager.hostname}:8030</value>
14. </property>
15. <property>
16.     <description>The http address of the RM web application.
17.     <name>yarn.resourcemanager.webapp.address</name>
18.     <value>${yarn.resourcemanager.hostname}:8088</value>
19. </property>
20. <property>
21.     <description>The https address of the RM web application.
22.     <name>yarn.resourcemanager.webapp.https.address</name>
23.     <value>${yarn.resourcemanager.hostname}:8090</value>
24. </property>
25. <property>
26.     <name>yarn.resourcemanager.resource-tracker.address</name>
27.     <value>${yarn.resourcemanager.hostname}:8031</value>
28. </property>
29. <property>
30.     <description>The address of the RM admin interface.
31.     <name>yarn.resourcemanager.admin.address</name>
32.     <value>${yarn.resourcemanager.hostname}:8033</value>
33. </property>
34. <property>
35.     <name>yarn.nodemanager.aux-services</name>
36.     <value>mapreduce_shuffle</value>
37. </property>
```

```
sudo gedit /usr/local/Hadoop/etc/Hadoop/mapred-site.xml //伪分布式不用配
```

```
[html]
01. <property>
02.     <name>mapreduce.framework.name</name>
03.     <value>yarn</value>
04. </property>
05. <property>
06.     <name>mapreduce.jobhistory.address</name>
07.     <value>master:10020</value>
08.     <description>MapReduce JobHistory Server IPC host:port</description>
09. </property>
10. <property>
11.     <name>mapreduce.jobhistory.webapp.address</name>
12.     <value>master:19888</value>
13.     <description>MapReduce JobHistory Server Web UI host:port</description>
14. </property>
```

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

```
[html]
01. <configuration>
02. <property>
03.     <name>dfs.replication</name>
04.     <value>1</value>
05. </property>
06. <property>
07.     <name>dfs.namenode.name.dir</name>
08.     <value>file:/usr/local/hadoop/dfs/name</value>
09. </property>
10. <property>
11.     <name>dfs.datanode.data.dir</name>
12.     <value>file:/usr/local/hadoop/dfs/data</value>
13. </property>
14. <property>                                //这个属性节点是为了防止后面eclopse存在拒
    绝读写设置的
15.     <name>dfs.permissions</name>
16.     <value>>false</value>
17. </property>
18. </configuration>
```

```
sudo gedit /usr/local/hadoop/etc/hadoop/masters 添加: localhost
```

```
sudo gedit /usr/local/hadoop/etc/hadoop/slaves 添加: localhost
```

关于配置的一点说明：上面只要配置 `fs.defaultFS` 和 `dfs.replication` 就可以运行，不过有个说法是如没有配置 `hadoop.tmp.dir` 参数，此时 Hadoop 默认的使用的临时目录为 `/tmp/hadoop-hadoop`，而这个目录在每次重启后都会被干掉，必须重新执行 `format` 才行（未验证），所以伪分布式配置中最好还是设置一下。

配置完成后，首先在 Hadoop 目录下创建所需的临时目录：

```
[html]
01. cd /usr/local/hadoop
02. mkdir tmp dfs dfs/name dfs/data
```

修改pid目录位置（可以不做）

```
vi hadoop-env.sh
```

```
export HADOOP_PID_DIR=/usr/local/hadoop-2.6.0/pid
```

```
vi mapred-env.sh
```

```
export HADOOP_MAPRED_PID_DIR=/usr/local/hadoop-2.6.0/pid
```

接着初始化文件系统HDFS。

```
[html]
01. bin/hdfs namenode -format //每次执行此命令要把dfs/data/文件清空
```

成功的话，最后的提示如下，Exiting with status 0 表示成功，Exiting with status 1: 则是出错。

```
[html]
01. sbin/start-dfs.sh
02. sbin/start-yarn.sh
```

开启Jobhistory

```
[html]
```

```
01. sbin/mr-jobhistory-daemon.sh start historyserver
02. <a target=_blank href="http://master:19888/">http://master:19888/</a>
```

Unable to load native-hadoop library for your platform这个提示,解决方式:

1、重新编译源码后将新的lib/native替换到集群中原来的lib/native

2、修改hadoop-env.sh, 增加

```
export HADOOP_OPTS="-Djava.library.path=$HADOOP_PREFIX/lib:$HADOOP_PREFIX/lib/native"
```

Namenode information:<http://localhost:50070>来查看Hadoop的信息。

All Applications: <http://http://2xx.81.8x.1xx:8088/>, 将其中的2xx.81.8x.1xx替换为你的实际IP地址。

运行例子:

1.先在hdfs上建个文件夹 bin/hdfs dfs -mkdir -p /user/ha1/input

```
bin/hdfs dfs -mkdir -p /user/ha1/output
```

2.上传一些文件: bin/hdfs dfs -put etc/hadoop/ /user/ha1/input 把etc/hadoop文件上传到hdfs的/user/ha1/input中

3.执行指令

```
bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.6.0.jar grep
/user/ha1/input/hadoop /user/ha1/output/temp 'dfs[a-z.]+'
```

错误1:

WARN hdfs.DFSClient: DataStreamer Exception

org.apache.hadoop.ipc.RemoteException(java.io.IOException): File /user/ha1/input/hadoop/yarn-env.sh.\_COPYING\_ could only be replicated to 0 nodes instead of minReplication (=1). There are 0 datanode(s) running and no node(s) are excluded in this operation.

//删除dfs/data/目录所有文件, 关闭所有服务重新执行bin/hdfs namenode -format以及1, 2步骤

错误2:

org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.hdfs.server.namenode.SafeModeException):

Cannot delete /user/root/grep-temp-23493900. Name node is in safe mode.

The reported blocks 188 has reached the threshold 0.9990 of total blocks 188. The number of live datanodes 2 has reached the minimum number 0. In safe mode extension. Safe mode will be turned off automatically in 10 seconds.

at

org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkNameNodeSafeMode(FSNamesystem.java:1

操作太快, 未退出安全模式, 等退出安全模式

4.查看结果

```
bin/hdfs dfs -cat /user/ha1/output/temp/*
```

```
[html]
```

```
01. 8 dfs.audit.logger
02. 4 dfs.class
03. 3 dfs.server.namenode.
04. 2 dfs.audit.log.maxbackupindex
```

```
05. 2 dfs.period
06. 2 dfs.audit.log.maxfilesize
07. 1 dfsmetrics.log
08. 1 dfsadmin
09. 1 dfs.servers
10. 1 dfs.replication
11. 1 dfs.file
12. 1 dfs.datanode.data.dir
13. 1 dfs.namenode.name.dir
```

终止一个job:

```
hadoop job -kill job_1447903602796_0001
```

顶

0

踩

0

上一篇

Ubuntu14.04下codeblocks手动编译配置bost\_1\_57\_0

下一篇

32位ubuntu14.04手动编译hadoop2.6.0

我的同类文章

Hadoop (10)

• hadoop2.6.0的eclipse插件编译和设置

• ubuntu搭建分布式hadoop-2.6.0概略和错误

• Hadoop 2.6.0动态添加节点

• Hdfs增量导入小文件合并的思路

• Hadoop2.6.0错误

• 32位ubuntu14.04手动编译hadoop2.6.0

• mac下配置eclipse的hadoop环境

• Hadoop修改SSH端口号

• eclipse打jar包步骤

主题推荐

ubuntu

eclipse

html

hadoop

分布式

桌面环境

管理

猜你在找

- Hadoop入门学习

Hadoop 2.0新特性剖析及应用

HDFS精讲

Hadoop2.X 应用解析

Hadoop 2.X大数据平台基础
- 新手推荐Hadoop安装教程\_单机伪分布式配置

ubuntu12环境下hadoop112伪分布式配置


ubuntu1304下配置hadoop-121伪分布式环境

Ubuntu 141032位上搭建Hadoop250单机伪分布式环

CentOS 65 配置hadoop 260伪分布式


查看评论

6楼 [wsrspirit](#) 2015-06-03 11:31发表



你好~我按照你的过程安装，单机完成之后wordcount执行成功，但是安装伪集群之后在执行wordcount就是这个错误：  
Exception in thread "main" java.net.ConnectException: Call From ubuntu/116.255.235.199 to localhost:9000 failed on connection exception: java.net.ConnectException: Connection refused; For more details see:  
<http://wiki.apache.org/hadoop/ConnectionRefused>  
请问这是怎么回事，hdfs和yarn能够启动

5楼 [WhoisPo](#) 2015-06-01 17:09发表



etc/hadoop/ 下面没有masters配置文件

4楼 [地下室森林](#) 2015-04-27 14:09发表

重新编译源码后将新的lib/native替换到集群中原来的lib/native



你好，问一下，怎么重新编译？

Re: [cuilll](#) 2015-10-04 17:27发表



回复地下室森林：不用重新编译，下载64位的native覆盖32位的native就可以了参考：  
<http://www.secdocor.com/html/yyjs/31101.html>

Re: [ggzone](#) 2015-04-27 19:01发表



回复地下室森林：不知道你说的是不是这个<http://blog.csdn.net/ggz631047367/article/details/42460589>

3楼 [ytwu\\_ucas](#) 2015-04-03 14:22发表



你好，请问，我在hadoop用户下创建的hadoop环境，但是我在另外的账户A安装的Hbase，平时使用的就是其他账户A。  
现在的问题是：  
1.A账户无法启动Hadoop。  
2.Hadoop账户可以启动Hadoop，但是无法启动Hbase账户。  
3.A账户由于Hadoop无法启动，导致Hbase也无法启动。

~~~~~纠结ing~~~~~  
期待回复！！！！

Re: [ggzone](#) 2015-04-03 17:45发表



回复ytwu\_ucas：没遇到过这个问题，不敢乱讲

2楼 [wlntuer](#) 2015-01-30 17:38发表



我觉得写的非常好，而且我基本上都能运行成功，实在是太棒啦，对于我个小白来讲，需要这种资料。

1楼 [完美妖姬](#) 2015-01-21 10:53发表



我到“在hadoop目录下执行WordCount”这块了就显示不对。显示是connection refused。这是怎么回事呢，急急

Re: [ggzone](#) 2015-03-14 09:19发表



回复完美妖姬：不好意思，近段时间没网。是不是忘了前面的某个步骤

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

\* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

核心技术类目

全部主题

Hadoop

AWS

移动游戏

Java

Android

iOS

Swift

智能硬件

Docker

OpenStack

VPN

Spark

ERP

IE10

Eclipse

CRM

JavaScript

数据库

Ubuntu

NFC

WAP

jQuery

BI

HTML5

Spring

Apache

.NET

API

HTML

SDK

IIS

Fedora

XML

LBS

Unity

Splashtop

UML

components

Windows Mobile

Rails

QEMU

KDE

Cassandra

CloudStack

FTC

coremail

OPhone

CouchBase

云计算

iOS6

Rackspace

Web App

SpringSide

Maemo

Compuware

大数据

aptech

Perl

Tornado

Ruby

Hibernate

ThinkPHP

HBase

Pure

Solr

Angular

Cloud Foundry

Redis

Scala

Django

Bootstrap

