micmiu – 软件开发+生活点滴

hadoop 2.2.0 单节点安装和测试

作者: Michael 日期: 2014 年 1 月 19 日 查看评论 发表评论 (6)

本文是详细记录Hadoop 2.2.0 在Mac OSX系统下单节点安装配置启动的详细步骤,并且演示运行一个简单的job。目录结构如下:

- 基础环境配置
- Hadoop安装配置
- 启动及演示

[一]、基础环境配置

1、OS: Mac OSX 10.9.1

2、JDK 1.6.0_65

不管是安装包还是自己编译源码安装都可以,这个就不多介绍了,搜索下有很多文章介绍的,只要确保环境变量配置正确即可,我的JAVA_HOME配置如下:

```
micmiu-mbp:~ micmiu$ echo $JAVA_HOME
/System/Library/Java/JavaVirtualMachines/1.6.0.jdk/Contents/Home
micmiu-mbp:~ micmiu$ java -version
java version "1.6.0_65"
Java(TM) SE Runtime Environment (build 1.6.0_65-b14-462-11M4609)
Java HotSpot(TM) 64-Bit Server VM (build 20.65-b04-462, mixed mode)
micmiu-mbp:~ micmiu$
```

3、无密码SSH登录

由于是单节点的应用,只要实现localhost 的无密码ssh登录即可,这个比较简单:

```
micmiu-mbp:~ micmiu$ cd ~
micmiu-mbp:~ micmiu$ ssh-keygen -t rsa -P ''
micmiu-mbp:~ micmiu$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

验证是否成功:

```
micmiu-mbp:~ micmiu$ ssh localhost
Last login: Sat Jan 18 10:17:19 2014
micmiu-mbp:~ micmiu$
```

这样就表示SSH无密码登录成功了。

有关SSH无密码登录的详细介绍可以参见:Linux(Centos)配置OpenSSH无密码登陆

[二]、**Hadoop**安装配置

1、下载发布包

打开官方下载链接 http://hadoop.apache.org/releases.html#Download ,选择2.2.0版本的发布包下载 后解压到指定路径下:`micmiu$ tar -zxf hadoop-2.2.0.tar.gz -C /usr/local/share`,那么本文中HADOOP_HOME = /usr/local/share/hadoop-2.2.0/。

近期文章

- Spring AOP拦截Struts1记录操作日志
- Baby
- Hadoop2.x在Ubuntu系统中编译源码
- Ubuntu设置默认jdk
- Ubuntu12.04中Eclipse安装svn插件问题
- eclipse中开发Hadoop2.x的Map/Reduce项目
- Hadoop2.x eclipse plugin插件编译安装配置
- maven编译Spark源码
- java.net.ConnectException: to 0.0.0.0:10020 failed on connection exception
- ERROR tool.ImportTool: Imported Failed: Attempted to generate class with no columns!

浏览

- 图文介绍openLDAP在windows上的安装配置 - 12,237 views
- Thrift入门及Java实例演示 - 11,665 views
- SSO之CAS单点登录实例演示 - 11,652 views
- java keytool证书工具使用小结 - 9,518 views
- java自带的MD5、SHA1算法演示 - 9,220 views
- VirtualBox安装Ubuntu图文介绍 - 8,792 views
- Openfire服务端源代码开发配置指南 - 8,720 views
- Eclipse+Maven构建web项目及部署时Maven lib依赖问题的解决 - 8,148 views
- Mac OSX 10.8安装SVN的方法 - 7,050 views

2、配置系统环境变量 `vi ~/.profile` ,添加如下内容:

```
# Hadoop settings by Michael@micmiu.com
export HADOOP_HOME="/usr/local/share/hadoop-2.2.0"
export HADOOP_PREFIX=${HADOOP_HOME}
export HADOOP_COMMON_HOME=${HADOOP_PREFIX}
export HADOOP_HDFS_HOME=${HADOOP_PREFIX}
export HADOOP_MAPRED_HOME=${HADOOP_PREFIX}
export HADOOP_YARN_HOME=${HADOOP_PREFIX}
export HADOOP_CONF_DIR="$HADOOP_HOME/etc/hadoop/"
export YARN_CONF_DIR=${HADOOP_CONF_DIR}

export PATH=$PATH:$HADOOP_PREFIX/bin:$HADOOP_PREFIX/sbin
```

3、修改 <HADOOP_HOME>/etc/hadoop/hadoop-env.sh

Mac OSX配置如下:

```
# The java implementation to use.
#export JAVA_HOME=${JAVA_HOME}
export JAVA_HOME=$(/usr/libexec/java_home -d 64 -v 1.6)
#找到HADOOP_OPTS 配置增加下面参数
export HADOOP_OPTS="$HADOOP_OPTS -Djava.security.krb5.realm=OX.AC.UK -
Djava.security.krb5.kdc=kdc0.ox.ac.uk:kdc1.ox.ac.uk"
```

跟多可以参见:$JAVA_HOME环境变量在Mac OS X中设置的问题

Linux|Unix 配置如下:

```
# The java implementation to use.
#export JAVA_HOME=${JAVA_HOME}
export JAVA_HOME=系统中JDK实际路径
```

4、修改 <HADOOP_HOME>/etc/hadoop/core-site.xml

在<configuration>节点下添加或者更新下面的配置信息:

```
<!-- 新变量f:s.defaultFS 代替旧的:fs.default.name |micmiu.com-->
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://localhost:9000</value>
    <description>The name of the default file system.</description>
  </property>

  <property>
    <name>hadoop.tmp.dir</name>
    <value>/Users/micmiu/tmp/hadoop</value>
    <description>A base for other temporary directories.</description>
  </property>

  <property>
    <name>io.native.lib.available</name>
    <value>false</value>
    <description>default value is true:Should native hadoop libraries, if
present, be used.</description>
  </property>
```

5、修改 <HADOOP_HOME>/etc/hadoop/hdfs-site.xml

在<configuration>节点下添加或者更新下面的配置信息:

```
<property>
        <name>dfs.replication</name>
        <value>1</value>
        <!-- 如果是单节点配置为1,如果是集群根据实际集群数量配置 | micmiu.com -->
</property>
```

6、修改 <HADOOP_HOME>/etc/hadoop/yarn-site.xml

在<configuration>节点下添加或者更新下面的配置信息:

- java.lang.OutOfMemoryError: GC overhead limit exceeded - 6,505 views

分类目录

- J2EE (41)
  - Hibernate (12)
  - Struts (5)
  - Spring (11)
  - JDBC (6)
  - JTA (3)
  - EJB (4)
- 海量数据 (28)
  - Hadoop (12)
  - Hive (8)
  - HBase (4)
  - Spark (2)
  - Sqoop (2)
- SOA (14)
  - webservice (12)
  - RPC (2)
- OpenSource (27)
  - openflashchart (2)
  - 生成文档 (5)
  - Openfire (4)
  - Security (1)
  - CORBA (4)
- 架构设计 (7)
  - 集群 (6)
  - 缓存 (1)
- 企业应用 (19)
  - SNMP (8)
  - SSO (4)
  - 服务器软件 (4)
- 编程语言 (34)
  - Java (24)
  - javascript (5)
  - Groovy (4)
  - nodejs (1)
- NoSQL (12)
  - Berkeley (6)
- 数据库 (10)
  - Oracle (7)
  - mysql (2)
- Android (2)
- web UI (1)
- 软件工具 (32)
  - 常用软件 (1)
  - 项目构建 (15)
  - 测试工具 (2)
- 操作系统 (21)
  - linux (15)
  - windows (1)
  - Mac (5)
- 小小 (12)
  - 心底软绵绵 (4)
  - 时间静悄悄 (5)
  - 肚皮圆滚滚 (3)

```
<!-- micmiu.com -->
<property>
        <name>yarn.nodemanager.aux-services</name>
        <value>mapreduce_shuffle</value>
    </property>

    <property>
        <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
        <value>org.apache.hadoop.mapred.ShuffleHandler</value>
    </property>
```

7、修改 <HADOOP_HOME>/etc/hadoop/mapred-site.xml

默认没有mapred-site.xml 文件,copy mapred-site.xml.template 一份为 mapred-site.xml即可

在<configuration>节点下添加或者更新下面的配置信息:

```
<!-- micmiu.com -->
<property>
        <name>mapreduce.framework.name</name>
        <value>yarn</value>
        <final>true</final>
</property>
```

[三]、启动及演示

1、启动Hadoop

首先执行`hdfs namenode -format`:

```
micmiu-mbp:~ micmiu$ hdfs namenode -format
14/01/18 23:07:07 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = micmiu-mbp.local/192.168.1.103
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.2.0
.................
.................
.................
14/01/18 23:07:08 INFO util.GSet: VM type       = 64-bit
14/01/18 23:07:08 INFO util.GSet: 0.029999999329447746% max memory = 991.7 MB
14/01/18 23:07:08 INFO util.GSet: capacity      = 2^15 = 32768 entries
Re-format filesystem in Storage Directory /Users/micmiu/tmp/hadoop/dfs/name ? (Y
or N) Y
14/01/18 23:07:26 INFO common.Storage: Storage directory
/Users/micmiu/tmp/hadoop/dfs/name has been successfully formatted.
14/01/18 23:07:26 INFO namenode.FSImage: Saving image file
/Users/micmiu/tmp/hadoop/dfs/name/current/fsimage.ckpt_0000000000000000000 using
no compression
14/01/18 23:07:26 INFO namenode.FSImage: Image file
/Users/micmiu/tmp/hadoop/dfs/name/current/fsimage.ckpt_0000000000000000000 of
size 198 bytes saved in 0 seconds.
14/01/18 23:07:27 INFO namenode.NNStorageRetentionManager: Going to retain 1
images with txid >= 0
14/01/18 23:07:27 INFO util.ExitUtil: Exiting with status 0
14/01/18 23:07:27 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at micmiu-mbp.local/192.168.1.103
************************************************************/
```

然后执行 `start-dfs.sh`:

```
micmiu-mbp:~ micmiu$ start-dfs.sh
Starting namenodes on [localhost]
localhost: starting namenode, logging to /usr/local/share/hadoop-
2.2.0/logs/hadoop-micmiu-namenode-micmiu-mbp.local.out
```

异常处理 (19)
杂谈 (2)

文章归档

2014年七月 (1)
2014年五月 (1)
2014年四月 (12)
2014年三月 (4)
2014年二月 (6)
2014年一月 (8)
2013年十二月 (1)
2013年十月 (9)
2013年九月 (7)
2013年八月 (8)
2013年六月 (4)
2013年五月 (4)
2013年四月 (3)
2013年三月 (10)
2013年二月 (1)
2013年一月 (2)
2012年十二月 (1)
2012年十一月 (7)
2012年十月 (2)
2012年九月 (5)
2012年八月 (13)
2012年七月 (14)
2012年六月 (12)
2012年五月 (16)
2012年四月 (20)
2012年三月 (22)
2012年二月 (88)

标签云(3D)

技术链接

10→me@iteye
21→nosqlfan
22→mysqlops
23→OpenCloudDB

```
localhost: starting datanode, logging to /usr/local/share/hadoop-2.2.0/logs/hadoop-micmiu-datanode-micmiu-mbp.local.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /usr/local/share/hadoop-2.2.0/logs/hadoop-micmiu-secondarynamenode-micmiu-mbp.local.out
micmiu-mbp:~ micmiu$ jps
1522 NameNode
1651 DataNode
1794 SecondaryNameNode
1863 Jps
micmiu-mbp:~ micmiu$
```

再执行 `start-yarn.sh` :

```
micmiu-mbp:~ micmiu$ start-yarn.sh
starting yarn daemons
starting resourcemanager, logging to /usr/local/share/hadoop-2.2.0/logs/yarn-micmiu-resourcemanager-micmiu-mbp.local.out
localhost: starting nodemanager, logging to /usr/local/share/hadoop-2.2.0/logs/yarn-micmiu-nodemanager-micmiu-mbp.local.out
micmiu-mbp:~ micmiu$ jps
2033 NodeManager
1900 ResourceManager
1522 NameNode
1651 DataNode
2058 Jps
1794 SecondaryNameNode
micmiu-mbp:~ micmiu$
```

启动日志没有错误信息,并确认上面的相关进程存在,就表示启动成功了。

2、演示

演示hdfs 一些常用命令,为wordcount演示做准备:

```
micmiu-mbp:~ micmiu$ hdfs dfs -ls /
micmiu-mbp:~ micmiu$ hdfs dfs -mkdir /user
micmiu-mbp:~ micmiu$ hdfs dfs -ls /
Found 1 items
drwxr-xr-x   - micmiu supergroup          0 2014-01-18 23:20 /user
micmiu-mbp:~ micmiu$ hdfs dfs -mkdir -p /user/micmiu/wordcount/in
micmiu-mbp:~ micmiu$ hdfs dfs -ls /user/micmiu/wordcount
Found 1 items
drwxr-xr-x   - micmiu supergroup          0 2014-01-18 23:21 /user/micmiu/wordcount/in
```

本地创建一个文件 micmiu-word.txt, 写入如下内容:

```
Hi Michael welcome to Hadoop
Hi Michael welcome to BigData
Hi Michael welcome to Spark
more see micmiu.com
```

把 micmiu-word.txt 文件上传到hdfs:

```
hdfs dfs -put micmiu-word.txt  /user/micmiu/wordcount/in
```

然后cd 切换到Hadoop的根目录下执行:

```
hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.2.0.jar wordcount  /user/micmiu/wordcount/in /user/micmiu/wordcount/out
```

ps: /user/micmiu/wordcount/out 目录不能存在 否则运行报错。

看到类似如下的日志信息:

```
micmiu-mbp:hadoop-2.2.0 micmiu$ hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.2.0.jar wordcount  /user/micmiu/wordcount/in /user/micmiu/wordcount/out
14/01/19 20:02:29 INFO client.RMProxy: Connecting to ResourceManager at
```

31→CrazyJvm
32→封神无度
33→董的博客

功能

登录
文章RSS
评论RSS
WordPress.org

```
/0.0.0.0:8032
14/01/19 20:02:29 INFO input.FileInputFormat: Total input paths to process : 1
14/01/19 20:02:29 INFO mapreduce.JobSubmitter: number of splits:1
.............
.............
.............
14/01/19 20:02:29 INFO mapreduce.JobSubmitter: Submitting tokens for job:
job_1390131922557_0001
14/01/19 20:02:30 INFO impl.YarnClientImpl: Submitted application
application_1390131922557_0001 to ResourceManager at /0.0.0.0:8032
14/01/19 20:02:30 INFO mapreduce.Job: The url to track the job: http://micmiu-
mbp.local:8088/proxy/application_1390131922557_0001/
14/01/19 20:02:30 INFO mapreduce.Job: Running job: job_1390131922557_0001
14/01/19 20:02:38 INFO mapreduce.Job: Job job_1390131922557_0001 running in
uber mode : false
14/01/19 20:02:38 INFO mapreduce.Job:  map 0% reduce 0%
14/01/19 20:02:43 INFO mapreduce.Job:  map 100% reduce 0%
14/01/19 20:02:50 INFO mapreduce.Job:  map 100% reduce 100%
14/01/19 20:02:50 INFO mapreduce.Job: Job job_1390131922557_0001 completed
successfully
14/01/19 20:02:51 INFO mapreduce.Job: Counters: 43
        File System Counters
                FILE: Number of bytes read=129
                FILE: Number of bytes written=158647
                FILE: Number of read operations=0
                FILE: Number of large read operations=0
                FILE: Number of write operations=0
                HDFS: Number of bytes read=228
                HDFS: Number of bytes written=83
                HDFS: Number of read operations=6
                HDFS: Number of large read operations=0
                HDFS: Number of write operations=2
        Job Counters
                Launched map tasks=1
                Launched reduce tasks=1
                Data-local map tasks=1
                Total time spent by all maps in occupied slots (ms)=3346
                Total time spent by all reduces in occupied slots (ms)=3799
        Map-Reduce Framework
                Map input records=4
                Map output records=18
                Map output bytes=179
                Map output materialized bytes=129
                Input split bytes=120
                Combine input records=18
                Combine output records=10
                Reduce input groups=10
                Reduce shuffle bytes=129
                Reduce input records=10
                Reduce output records=10
                Spilled Records=20
                Shuffled Maps =1
                Failed Shuffles=0
                Merged Map outputs=1
                GC time elapsed (ms)=30
                CPU time spent (ms)=0
                Physical memory (bytes) snapshot=0
                Virtual memory (bytes) snapshot=0
                Total committed heap usage (bytes)=283127808
        Shuffle Errors
                BAD_ID=0
                CONNECTION=0
                IO_ERROR=0
                WRONG_LENGTH=0
                WRONG_MAP=0
                WRONG_REDUCE=0
        File Input Format Counters
                Bytes Read=108
        File Output Format Counters
                Bytes Written=83
micmiu-mbp:hadoop-2.2.0 micmiu$
```

到此 wordcount的job已经执行完成,执行如下命令可以查看刚才job的执行结果:

```
micmiu-mbp:hadoop-2.2.0 micmiu$ hdfs dfs -ls /user/micmiu/wordcount/out
Found 2 items
-rw-r--r--   1 micmiu supergroup          0 2014-01-19 20:02
/user/micmiu/wordcount/out/_SUCCESS
-rw-r--r--   1 micmiu supergroup         83 2014-01-19 20:02
/user/micmiu/wordcount/oummmicmiu-mbp:hadoop-2.2.0 micmiu$ hdfs dfs -cat
/user/micmiu/wordcount/out/part-r-00000
BigData 1
Hadoop  1
```

```
Hi      3
Michael 3
Spark   1
micmiu.com      1
more    1
see     1
to      3
welcome 3
```

ps:这篇文章是两天完成的,所以演示中日志的时间前后有一天间隔。

————— EOF @Michael Sun —————

Hadoop Hadoop

← Hadoop 2.x build native library on Mac os x

hadoop 2.2.0 集群模式安装配置和测试 →

发表评论?

6 条评论。

hadoop少年 2014 年 7 月 4 日 在 下午 11:10 回复

博主请教你一下:
export HADOOP_OPTS="$HADOOP_OPTS -Djava.security.krb5.realm=OX.AC.UK -Djava.security.krb5.kdc=kdc0.ox.ac.uk:kdc1.ox.ac.uk"
这个参数的配置是什么意思啊

hadoop少年 2014 年 7 月 4 日 在 下午 5:02 回复

感谢博主的教程,终于弄好了!!哈哈

hadoop少年 2014 年 7 月 4 日 在 上午 10:33 回复

你好,为什么我的hadoop启动就提示localhost: env: bash: No such file or directory
,并且没有datanode,也没有secondarynamenode呢,logs里提示
ERROR org.apache.hadoop.security.UserGroupInformation: PriviledgedActionException as:UncleDrew cause:java.io.IOException: File /Users/UncleDrew/tmp/hadoop/mapred/system/jobtracker.info could only be replicated to 0 nodes, instead of 1
2014-07-03 11:02:09,449 INFO org.apache.hadoop.ipc.Server: IPC Server handler 6 on 9000, call addBlock(/Users/UncleDrew/tmp/hadoop/mapred/system/jobtracker.info, DFSClient_NONMAPREDUCE_-98620818_1, null) from 127.0.0.1:53375: error: java.io.IOException: File /Users/UncleDrew/tmp/hadoop/mapred/system/jobtracker.info could only be replicated to 0 nodes, instead of 1
java.io.IOException: File /Users/UncleDrew/tmp/hadoop/mapred/system/jobtracker.info could only be replicated to 0 nodes, instead of 1 这该怎么解决啊

hadoop少年 2014 年 7 月 2 日 在 下午 6:19 回复

博主,我的在网上下载了一个64位native包替换,为什么还是不行呢? 一样的警告。

Michael 2014 年 7 月 3 日 在 上午 10:55 回复

去百度网盘 http://yun.baidu.com/s/1c0rfIOo 下载

hadoop少年 2014 年 7 月 4 日 在 上午 9:52 回复

谢谢

发表评论

昵称

邮箱

网址

注意 - 你可以用以下 HTML tags and attributes:
<a href="" title=""> <abbr title=""> <acronym title=""> <b> <blockquote cite=""> <cite> <code> <del datetime=""> <i> <q cite=""> <strike>

提交

↑ 回到顶部