Hadoop+Hbase+Solr(集成中文分词)分布式部署流程_V0.1.2

集群节点规划及示例说明

1. 主节点namenode:

```
hadoop120 192.168.138.120
```

2. 数据节点datanode*2：

```
hadoop121 192.168.138.121
hadoop122 192.168.138.122
```

3. 示例中 Hbase 和 Solr 都部署在主节点hadoop120 上面（这个不是固定的、自行决定部署在哪个节点，也可以单独部署）
4. 所有节点的hadoop用户名都一样，密码按需设置；下面示例中用户密码都是hadoop/hadoop
5. 虚拟机安装suse系统时候尽量不要用简易安装，简易安装后是动态IP获取而且是默认hostname，修改hostname比较麻烦（后附修改方法）；自定义安装过程中可以配置IP和hostname以及其他配置（比如关闭防火墙等等）
6. 所有节点关闭防火墙
7. 部署hadoop前先同步时间(最好是连接统一的NTPserver)，不然会影响hbase的集群启动
8. 以下示例代码命令行中 `#` 代表是超级管理员用户（root），`$` 代表普通用户（hadoop用户);`sudo` 是用普通用户执行管理员权限的动作，需要输入管理员密码
9. 配置文件中的 `#` 标识注释 （profile、*.sh）
10. 阅读及操作以下内容需要一定的Linux的基础，部分命令没有解释请自行google用法

示例中用到的工具和依赖包

- jdk-7u79-linux-x64.tar.gz
- hadoop-2.5.0.tar.gz
- apache-tomcat-7.0.56.tar.gz (部署solr的中间件)
- hbase-0.98.7-hadoop2-bin.tar.gz
- solr-4.10.1.tgz
- jcseg-1.9.5-src-jar-dict.zip（中文分词）
- commons-logging-api-1.0.4.jar + slf4j-api-1.7.12.jar （在tomcat中启动solr的时候需要这两个jar包，添加到tomcat的lib文件中）
- hadoop-client-2.5.0.jar（Hbase中用来替换hadoop-client-2.2.0.jar的包，/hadoop-2.x.x/share/中没有这个包，从maven中下载的）
- Xmanager(包括ssh工具、ftp工具等)

过程简述（默认hadoop用户已经添加且hostname、IP设置完成）

1. 配置hosts文件，即按照规划的IP和节点hostname将其写入hosts文件，方便各节点直接解析以及hadoop相关文件配置；
2. 解压安装JDK，示例中配置的全局的环境变量`/etc/profile`，所以用的是root用户；
3. 配置SSH免密码登录，配置的所有节点之间hadoop用户免密码登录，不是root用户的免密登录，所以用的是hadoop用户的权限；
4. 解压安装hadoop，添加到全局环境变量`/etc/profile`用的是root用户权限，修改配置文件、打包和远程传输配置好的hadoop文件及启动关闭hadoop集群都是在hadoop用户权限下完成；所有节点的hadoop配置都一样；
5. 解压安装Hbase,hadoop用户权限下完成修改配置文件、打包和远程拷贝hbase文件及启动关闭Hbase集群，不管用什么权限拷贝的jar包到Hbase文件夹中，最后一定要将owner修改为hadoop用户；所有节点的hbase配置都一样；
6. 解压安装solr,全部操作都是在hadoop用户权限下，将solr的工程war包部署在tomcat容器中，修改启动tomcat后自动解压的war包中的配置文件，拷贝相关依赖jar包到tomcat容器中，此时solr已经部署完成；然后拷贝中文分词jar包并修改相关配置文件，重启tomcat测试中文分词。

1.hosts文件修改(root用户权限下)

- 配置hosts文件（下图中最下面三行）

```
# vim /etc/hosts
```

```
hadoop@hadoop120:~> sudo vim /etc/hosts
root's password:

#
# hosts         This file describes a number of hostname-to-address
#               mappings for the TCP/IP subsystem.  It is mostly
#               used at boot time, when no name servers are running.
#               On small systems, this file can be used instead of a
#               "named" name server.
# Syntax:
#
# IP-Address  Full-Qualified-Hostname  Short-Hostname
#

127.0.0.1       localhost

# special IPv6 addresses
::1             localhost ipv6-localhost ipv6-loopback

fe00::0         ipv6-localnet

ff00::0         ipv6-mcastprefix
ff02::1         ipv6-allnodes
ff02::2         ipv6-allrouters
ff02::3         ipv6-allhosts

192.168.138.120 hadoop120
192.168.138.121 hadoop121
192.168.138.122 hadoop122
~
```

2.JDK 安装 (root用户权限下)

- 解压

```
# mkdir /usr/lib/java
# cp /{FTP_PATH}/jdk-7u79-linux-x64.tar.gz  /usr/lib/java
# tar -zxvf jdk-7u79-linux-x64.tar.gz
# chown root:root jdk1.7.0_79 -R
# rm jdk-7u79-linux-x64.tar.gz
```

- 配置文件

```
# vim /etc/profile
```

- 在profile末尾添加以下内容（文本内容中#是注释的意思）

```
# set java environment
export JAVA_HOME=/usr/lib/java/jdk1.7.0_79
export JRE_HOME=/usr/lib/java/jdk1.7.0_79/jre
export CLASSPATH=.:$JAVA_HOME/lib:$JRE_HOME/lib:$CLASSPATH
export PATH=$JAVA_HOME/bin:$JRE_HOME/bin:$JAVA_HOME:$PATH
# end of java
```

- 保存profile后执行编译操作，让变更生效

```
# source /etc/profile
```

- 测试是否安装成功

```
# java -version
```

- 输出刚才安装的java版本号，说明安装成功；输出找不到java命令或者不是刚才配置的版本则说明安装失败

3.SSH免密码登录 （hadoop用户权限下)

- 一定不能把root的公钥拷贝给其他机器;SSH密钥生成过程在每个节点上面都需要做，把所有节点的公钥相互拷贝即所有节点都可以通过hadoop用户免密登录其他节点
- 检测系统是否安装SSH server

```
$ ssh localhost
```

  - 如果弹出来输入yes/no的选项，然后要求输入密码，则说明ssh server 已安装，输入密码进入后再输入exit退出。如下图所示

```
hadoop@hadoop120:~> ssh localhost
The authenticity of host 'localhost (::1)' can't be established.
ECDSA key fingerprint is 1a:c9:e0:0b:29:97:82:08:fb:3c:0a:ff:a5:88:e9:64 [MD5].
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Password:
Last login: Mon Feb 27 10:12:07 2017 from 192.168.138.1
hadoop@hadoop120:~> exit
logout
Connection to localhost closed.
```

- 配置ssh免密码登录：生成密钥

```
$ cd ~/.ssh
$ ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
```

```
hadoop@hadoop120:~/.ssh> ssh-keygen -t rsa -P '' -f ~/.ssh/id_rsa
Generating public/private rsa key pair.
Your identification has been saved in /home/hadoop/.ssh/id_rsa.
Your public key has been saved in /home/hadoop/.ssh/id_rsa.pub.
The key fingerprint is:
56:c0:3a:98:8b:f2:7b:d8:f1:b4:03:f7:36:c2:30:68 [MD5] hadoop@hadoop120
The key's randomart image is:
+--[ RSA 2048]----+
|       ..        |
|        ..       |
|     o .  .      |
|    o o  .       |
|   o . .S        |
|. E * o.         |
| + o X o         |
|  o o * +        |
|  .o   + .       |
+--[MD5]----------+
```

- ~/.ssh目录下内容： **私钥 id_rsa 的权限是600**

```
hadoop@hadoop120:~/.ssh> ll
total 12
-rw------- 1 hadoop users 1675 Feb 27 10:58 id_rsa
-rw-r--r-- 1 hadoop users  398 Feb 27 10:58 id_rsa.pub
-rw-r--r-- 1 hadoop users  171 Feb 27 10:18 known_hosts
```

- 配置ssh免密码登录：添加公钥到受信任列表

```
$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
hadoop@hadoop120:~/.ssh> cat id_rsa.pub >> authorized_keys
hadoop@hadoop120:~/.ssh> ll
total 16
-rw-r--r-- 1 hadoop users  398 Feb 27 11:57 authorized_keys
-rw------- 1 hadoop users 1675 Feb 27 10:58 id_rsa
-rw-r--r-- 1 hadoop users  398 Feb 27 10:58 id_rsa.pub
-rw-r--r-- 1 hadoop users  171 Feb 27 10:18 known_hosts
hadoop@hadoop120:~/.ssh> cat authorized_keys
ssh-rsa AAAAB3NzaC1yc2EAAAADAQABAAABAQCV1pJ7mlfBGDTyniKhvGSCAlxwvI6RqDAvYI+uQLbYOYH6wCny9GjG4+s8vNpumqKFdy9as323QLqZZmIWQikFsdnx1OC5IgV7s7vHscr
VOKr3uCOWW0I18LbgJxJkWIUwEibKTYKuSd9h8cUjtyvw3AZ5Gml2qMk/mR0I82vSTK4s8CArNab7AedmcTC/YYMFG4HzucObPS/39JnWACZuiCvq0MKDdAvzFSBWyh3bUpxdkwn3j7SyUv
BaHbgmnIMa1JnEr1dkCznPvKmxJLYpoManbLfA0yZKjA1A5XJWsnpVmzxp4CrI3t62923WTCs9m5yV5frJec7dcrlDGulv hadoop@hadoop120
```

- 测试是否可以免密码连接本机

```
$ ssh localhost
```

```
hadoop@hadoop120:~/.ssh> ssh localhost
Last login: Mon Feb 27 10:18:29 2017 from localhost
hadoop@hadoop120:~> exit
logout
Connection to localhost closed.
```

- 复制公钥到其他节点(hadoop120/hadoop121/haddoop122在hosts中已经配置)

```
$ scp ~/.ssh/id_rsa.pub hadoop@hadoop121:~/.ssh/id_rsa.pub120
$ ssh hadoop121
$ cd ~/.ssh/
$ cat ~/.ssh/id_rsa.pub120 >> ~/.ssh/authorized_keys
```

  - **此时hadoop120在hadoop121的受信任列表中即在hadoop120机器上可以无密码访问hadoop121；在hadoop121上重复生成密钥的过程，将生成的公钥添加到hadoop120的受信任列表即可让hadoop121无密码ssh连接hadoop120**
  - **所有节点重复以上过程，将自己的公钥拷贝给其他节点**

4.安装Hadoop（hadoop用户权限下）

- root角色用mkdir命令在`/usr/`下创建hadoop文件夹，变更拥有者为hadoop用户，`/usr/hadoop`**目录下的所有文件的拥有者都是hadoop用户，**以下不再说明；
- 解压hadoop-2.5.0.tar.gz到/usr/hadoop/路径下
- 配置文件（编辑profile需要root权限）

```
# vim /etc/profile
```

- 在profile末尾添加以下内容

```
# set hadoop environment
export HADOOP_HOME=/usr/hadoop/hadoop-2.5.0
export PATH=.:$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$PATH
export HADOOP_HOME_WARN_SUPPRESS=1
export HADOOP_COMMON_LIB_NATIVE_DIR=$HADOOP_HOME/lib/native
export HADOOP_OPTS="$HADOOP_OPTS -Djava.library.path=$HADOOP_HOME/lib"
# end of hadoop
```

- 保存profile后执行编译操作，让变更生效

```
# source /etc/profile
```

- 测试是否安装成功

```
$ hadoop version
```

```
hadoop@hadoop120:~> hadoop version
Hadoop 2.5.0
Subversion http://svn.apache.org/repos/asf/hadoop/common -r 1616291
Compiled by jenkins on 2014-08-06T17:31Z
Compiled with protoc 2.5.0
From source with checksum 423dcd5a752eddd8e45ead6fd5ff9a24
This command was run using /usr/hadoop/hadoop-2.5.0/share/hadoop/common/hadoop-common-2.5.0.jar
```

- 配置${HADOOP_HOME}/etc/hadoop路径下的五个配置文件slaves、core-site.xml、hdfs-site.xml、mapred-site.xml、yarn-site.xml

```
hadoop@hadoop120:~> cd /usr/hadoop/hadoop-2.5.0/etc/hadoop/
hadoop@hadoop120:/usr/hadoop/hadoop-2.5.0/etc/hadoop> ls
capacity-scheduler.xml  hadoop-env.sh               httpfs-env.sh           mapred-env.cmd             ssl-client.xml.example
configuration.xsl       hadoop-metrics2.properties  httpfs-log4j.properties mapred-env.sh              ssl-server.xml.example
container-executor.cfg  hadoop-metrics.properties   httpfs-signature.secret mapred-queues.xml.template yarn-env.cmd
core-site.xml           hadoop-policy.xml           httpfs-site.xml         mapred-site.xml.template   yarn-env.sh
hadoop-env.cmd          hdfs-site.xml               log4j.properties        slaves                     yarn-site.xml
```

- 6.5-1 salves
  - **删除localhost(如果保留localhost，则将namenode也同时作为datanode)**
  - **每行添加一个datanode的hostname即hosts文件配置的名称**

```
hadoop@hadoop120:/usr/hadoop/hadoop-2.5.0/etc/hadoop> vim slaves
hadoop@hadoop120:/usr/hadoop/hadoop-2.5.0/etc/hadoop> cat slaves
hadoop121
hadoop122
hadoop@hadoop120:/usr/hadoop/hadoop-2.5.0/etc/hadoop>
```

- 6.5-2 core-site.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>
        <property>
                <name>fs.defaultFS</name>
                <value>hdfs://hadoop120:9000</value>
        </property>
        <property>
                <name>hadoop.tmp.dir</name>
                <value>file:/usr/hadoop/hadoop-2.5.0/tmp</value>
                <description>Abase for other temporary directories.</description>
        </property>
</configuration>
```

- 6.5-3 hdfs-site.xml

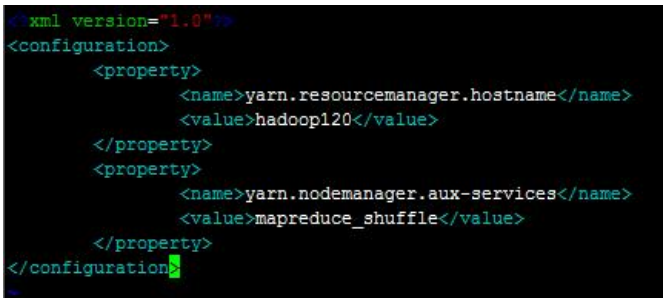
```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<configuration>
        <property>
                <name>dfs.namenode.secondary.http-address</name>
                <value>hadoop120:50090</value>
        </property>
        <property>
                <name>dfs.replication</name>
                <value>1</value>
        </property>
        <property>
                <name>dfs.namenode.name.dir</name>
                <value>file:/usr/hadoop/hadoop-2.5.0/tmp/dfs/name</value>
        </property>
        <property>
                <name>dfs.datanode.data.dir</name>
                <value>file:/usr/hadoop/hadoop-2.5.0/tmp/dfs/data</value>
        </property>
</configuration>
```

- 6.5-4 mapred-site.xml

```
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>

<configuration>
        <property>
                <name>mapreduce.framework.name</name>
                <value>yarn</value>
        </property>
        <property>
                <name>mapreduce.jobhistory.address</name>
                <value>hadoop120:10020</value>
        </property>
        <property>
                <name>mapreduce.jobhistory.webapp.address</name>
                <value>hadoop120:19888</value>
        </property>
</configuration>
~
```

- 6.5-5 yarn-site.xml



```
<?xml version="1.0"?>
<configuration>
        <property>
                <name>yarn.resourcemanager.hostname</name>
                <value>hadoop120</value>
        </property>
        <property>
                <name>yarn.nodemanager.aux-services</name>
                <value>mapreduce_shuffle</value>
        </property>
</configuration>
```

- 打包已经配置好的hadoop文件夹然后复制到其他节点

```
$ cd /usr/hadoop/
$ tar -zcf ~/hadoop.master.tar.gz ./hadoop-2.5.0
$ cd ~
$ scp hadoop.master.tar.gz hadoop@hadoop121:/usr/hadoop
```

  - **记得更改拷贝过去的文件权限，然后解压($ tar -zxvf)**
- 安装其他节点的jdk并在profile中增加java及hadoop环境变量（见4.JDK安装）；
- 配置SSH无密码登录（见5.SSH免密码登录）
- 所有节点的hadoop和java都配置无误后，第一次启动hadoop集群要在namenode进行初始化

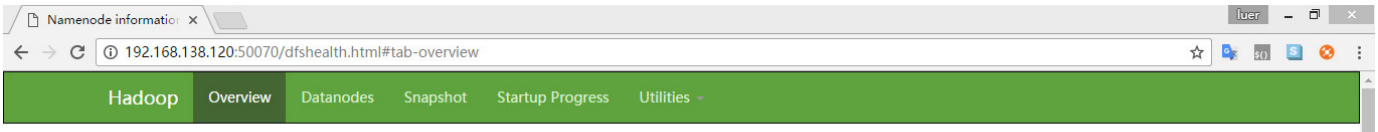
```
$ hdfs namenode -format
```

- 启动hadoop集群

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

- 通过`jps`命令查看各个节点所启动的进程
  - hadoop120即namenode上应该有`NameNode、ResourceManager、SecondrryNameNode、JobHistoryServer、jps`进程
  - hadoop121和hadoop122即datanode上应该有`DataNode、NodeManager、jps`进程
  - **以上所有进程少一不可**
- 在namenode上即hadoop120上 通过命令 `hdfs dfsadmin -report` 查看 DataNode 是否正常启动
  - 如果 Live datanodes 不为 0 ，则说明集群启动成功
- 通过访问web页面查看集群状态

```
http://hadoop120:50070/
http://hadoop120:8088/
```



Overview 'hadoop120:9000' (active)

Started:	Wed Mar 08 10:08:17 CST 2017
Version:	2.5.0, r1616291
Compiled:	2014-08-06T17:31Z by jenkins from branch-2.5.0
Cluster ID:	CID-09c3853e-d5b2-4516-b05d-1fc6ea2b0297
Block Pool ID:	BP-1882334979-192.168.138.120-1488870009555

Summary

Security is off.

Safemode is off.

45 files and directories, 15 blocks = 60 total filesystem object(s).

Heap Memory used 32.06 MB of 52.63 MB Heap Memory. Max Heap Memory is 966.69 MB.

Non Heap Memory used 29.82 MB of 31.31 MB Commited Non Heap Memory. Max Non Heap Memory is 130 MB.

Configured Capacity:	37.26 GB
DFS Used:	160 KB
Non DFS Used:	13.02 GB
DFS Remaining:	24.24 GB
DFS Used%:	0%
DFS Remaining%:	65.05%
Block Pool Used:	160 KB
Block Pool Used%:	0%
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	2 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0

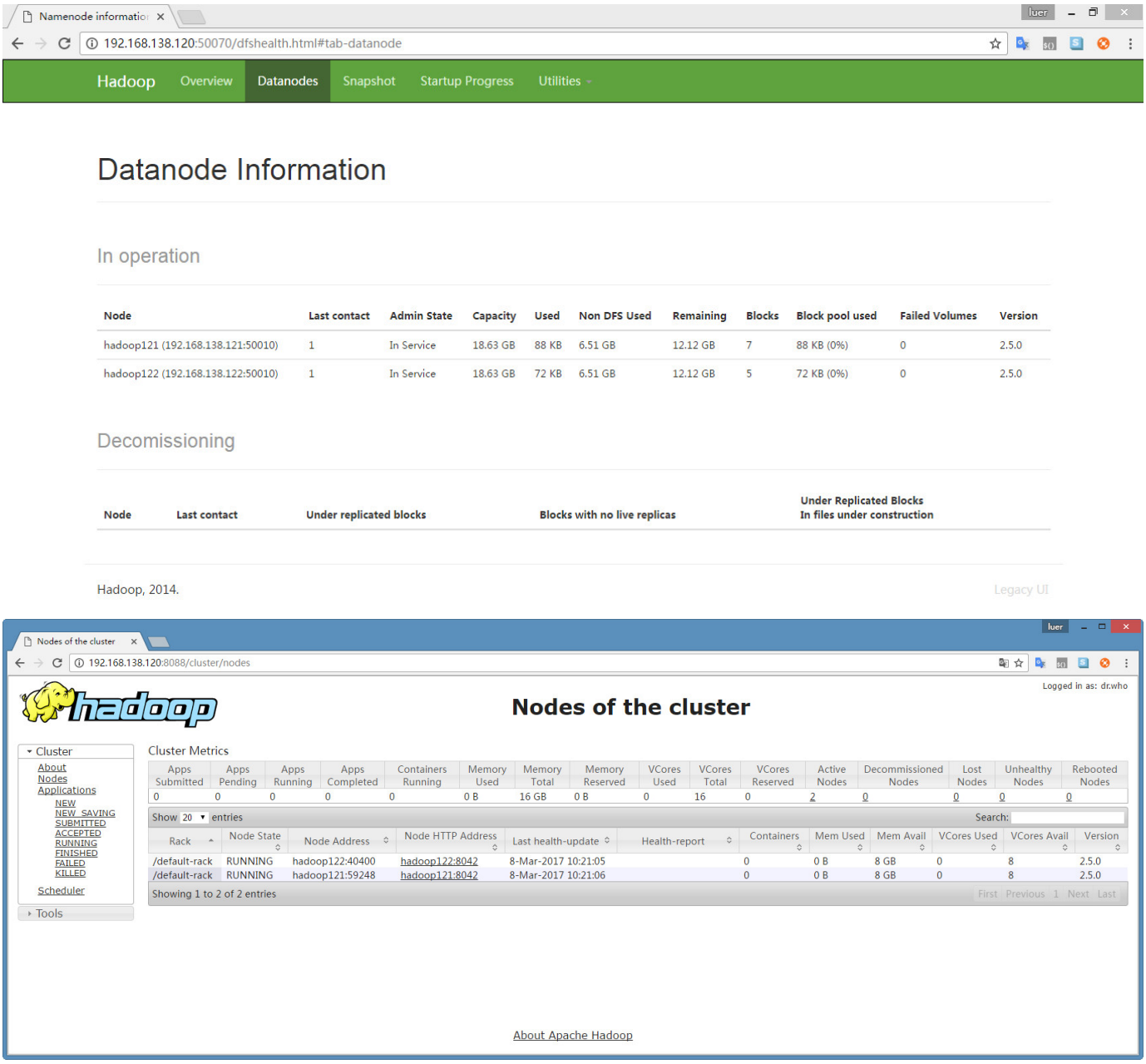
NameNode Journal Status

Current transaction ID: 206

Journal Manager	State
FileJournalManager(root=/usr/hadoop/hadoop-2.5.0/tmp/dfs/name)	EditLogFileOutputStream(/usr/hadoop/hadoop-2.5.0/tmp/dfs/name/current/edits_inprogress_0000000000000000168)

NameNode Storage

Storage Directory	Type	State
/usr/hadoop/hadoop-2.5.0/tmp/dfs/name	IMAGE_AND_EDITS	Active



- 关闭hadoop集群

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

5.安装Hbase（hadoop用户权限下）

- 用tar命令解压hbase的包到`/usr/hadoop/`目录下
- 修改`/conf/hbase-env.sh`中(该行是注释掉的，去掉前面的“#”即可)

```
export HBASE_MANAGES_ZK=true
```

- 修改`/conf/hbase-site.xml`:

```
<configuration>
    <property>
            <name>hbase.master</name>
            <value>hadoop120:60000</value>
    </property>
```

```
    <property>
            <name>hbase.rootdir</name>
            <value>hdfs://hadoop120:9000/hbase</value>
            <description>
                    The directory sharedbyRegionServers.
            </description>
    </property>
    <property>
            <name>hbase.cluster.distributed</name>
            <value>true</value>
    </property>
    <property>
            <name>hbase.zookeeper.property.clientPort</name>
            <value>2181</value>
    </property>
    <property>
            <name>hbase.zookeeper.quorum</name>
            <value>hadoop121,hadoop122</value>
    </property>
</configuration>
```

- 修改`/conf/regionservers`文件（一行一个数据节点）

```
hadoop121
hadoop122
```

- 在HDFS上创建，在hbase-site.xml中设置的HBASE的rootdir路径，该路径不是linux系统路径，是hadoop中的路径

```
hadoop  fs  -mkdir  /hbase
```

- 替换/lib/下面中hadoop开头的jar包（15个）
  - hbase-0.98.7里面的是2.2版本的，为了保证一致性需要替换成2.5（其他版本同理）的hadoop的jar包
  - 这些jar分散在/haodoop-2.5.0/share/hadoop/中的各个子文件夹，只是版本号不一样，找不到的jar包（比如hadoop-client-2.5.0.jar这个我就没找到）去maven中央库中下载
- 打包配置好的hbase并复制到数据节点对应的目录下解压（同配置好的hadoop文件夹的拷贝解压过程，不再详述）
- 切回namenode即hadoop120,启动HBase

```
$ /.../bin/start-hbase.sh
```

- 查看节点进程(jps命令）
  - namenode（hadoop120）中应该多了`HMaster`线程
  - datanode (hadoop121\hadoop122) 中应该多了`HQuorumPeer、HRegionServer`两个线程
- 通过web页面查看HBase状态

```
http://hadoop120:60010/
```

Master hadoop120

Region Servers

Dead Region Servers

	ServerName	Stop time
	hadoop121,60020,1488873861627	Tue Mar 07 17:04:42 CST 2017
	hadoop122,60020,1488873860426	Tue Mar 07 17:04:39 CST 2017
Total:	servers: 2	

Backup Masters

ServerName	Port	Start Time
Total:0		

Tables

User Tables | System Tables | Snapshots

Regions in Transition

Region	State	RIT time (ms)
1588230740	hbase:meta,,1.1588230740 state=FAILED_OPEN, ts=Tue Mar 07 16:04:34 CST 2017 (9096s ago), server=hadoop122,60020,1488873860426	9096079
Total number of Regions in Transition for more than 60000 milliseconds	1	
Total number of Regions in Transition	1	

Tasks

Show All Monitored Tasks | Show non-RPC Tasks | Show All RPC Handler Tasks | Show Active RPC Calls | Show Client Operations | View as JSON

Start Time	Description	State	Status
Tue Mar 07 17:04:39 CST 2017	Doing distributed log split in [hdfs://hadoop120:9000/hbase/WALs/hadoop122,60020,1488873860426-splitting]	RUNNING (since 1hrs, 31mins, 30sec ago)	Waiting for distributed tasks to finish. scheduled=1 done=0 error=0 (since 0sec ago)
Tue Mar 07 16:04:19 CST 2017	Master startup	RUNNING (since 2hrs, 31mins, 50sec ago)	Assigning hbase:meta region (since 2hrs, 31mins, 38sec ago)

Software Attributes

Attribute Name	Value	Description
HBase Version	0.98.7-hadoop2, r800c23e2207aa3f9bddb7e9514d8340bcfb89277	HBase version and revision
HBase Compiled	Wed Oct 8 15:58:11 PDT 2014, apurtell	When HBase version was compiled and by whom
Hadoop Version	2.5.0, r1616291	Hadoop version and revision
Hadoop Compiled	2014-08-06T17:31Z, jenkins	When Hadoop version was compiled and by whom
Zookeeper Quorum	hadoop122:2181,hadoop121:2181	Addresses of all registered ZK servers. For more, see zk dump.
HBase Root Directory	hdfs://hadoop120:9000/hbase	Location of HBase home directory
HMaster Start Time	Tue Mar 07 16:04:20 CST 2017	Date stamp of when this HMaster was started
HMaster Active Time	Tue Mar 07 16:04:22 CST 2017	Date stamp of when this HMaster became active
HBase Cluster ID	fbbdc6a1-8cd4-4058-b852-f565edb56a27	Unique identifier generated for each HBase cluster
Load average	NaN	Average number of regions per regionserver. Naive computation.
Coprocessors	[]	Coprocessors currently loaded by the master

- 关闭HBase

```
$ /.../bin/stop-hbase.sh
```

6.安装Solr（hadoop用户权限下）

- 将apache-tomcat和solr的压缩包解压至namenode即hadoop120的/usr/hadoop目录下
- 复制`/.../solr4.10.1/dist/solr-4.10.1.war`到`/.../apache-tomcat-7.0.56/webapps`目录下
- 启动apache-tomcat-7.0.56，webapps下的solr的war包会自动解压，进入`/.../apache-tomcat-7.0.56/webapps/solr-4.10.1/WEB-INF`下，修改web.xml文件：

```
<env-entry>
   <env-entry-name>solr/home</env-entry-name>
   <env-entry-value>/usr/hadoop/solr-4.10.1/example/solr</env-entry-value>
   <!--Your solr path -->
   <env-entry-type>java.lang.String</env-entry-type>
</env-entry>
```

- 将 `commons-logging-api-1.0.4.jar` 和 `slf4j-api-1.7.12.jar` 添加到tomcat的`/.../apache-tomcat-7.0.56/lib/`目录下
- 重启tomcat，打开`http://hadoop120:8080/solr-4.10.1`测试是否可打开solr主页
  - 此时的solr还不能进行中文分词，需要添加对应的依赖包并更新配置文件
- 修改 `/.../solr-4.10.1/example/solr/collection1/conf/`下的`schema.xml`文件，添加如下代码到末尾：

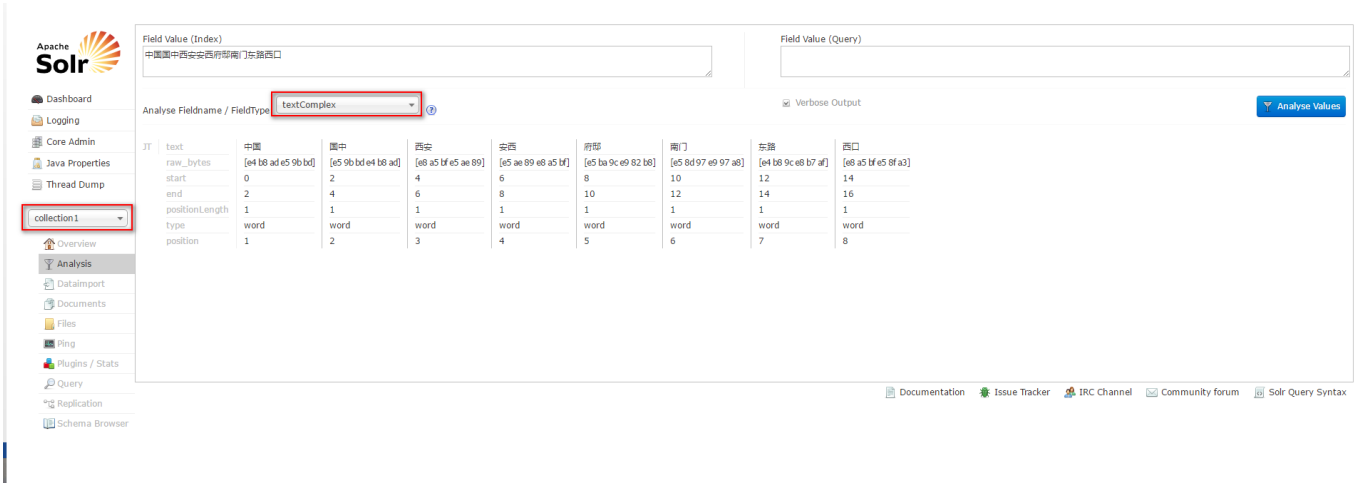
```
<fieldType name="textComplex" class="solr.TextField">
        <analyzer>
                <tokenizer class="org.lionsoul.jcseg.solr.JcsegTokenizerFactory" mode="complex"/>
        </analyzer>
</fieldType>
<fieldType name="textSimple" class="solr.TextField">
        <analyzer>
                <tokenizer class="org.lionsoul.jcseg.solr.JcsegTokenizerFactory" mode="simple"/>
        </analyzer>
</fieldType>
```

- 解压jcseg-1.9.5-src-jar-dict.zip：

```
unzip  jcseg-1.9.5-src-jar-dict.zip
```

> 可以直接把解压出来的jar包拷贝进去，就不用在suse中用命令解压了

- 将`/.../jcseg-1.9.5/output`文件夹中的`jcseg-core-1.9.5.jar`和`jcseg-solr-1.9.5.jar` 拷贝到`/.../apache-tomcat-7.0.56/webapps/solr/WEB-INF/lib/`文件夹下
- 将文件夹`/.../solr-4.10.1/dist/solrj-lib/`中的jar包拷贝到`/.../apache-tomcat-7.0.56/webapps/solr/WEB-INF/lib/`目录下
- 将`/.../jcseg-1.9.5/`文件夹中的词库文件夹`lexicon`拷贝到文件夹`/.../apache-tomcat-7.0.56/webapps/solr/WEB-INF/lib`下
- 打开`http://hadoop120:8080/solr-4.10.1`，选`collection1，textComplex`，填一段中文进行测试，是否成功分词



Tips:Suse11 修改hostname

- 改当前的hostname可以用

```
# hostname  myname
```

  - 此时只是修改了内存中的，下次启动又恢复了
- 机器在启动的时候会调用`/etc/rc.d/boot.localnet`

> 这个脚本会判断当前的$HOSTNAME变量是否设置，没有设置就会读取/etc/HOSTNAME来设置主机名。

- 因此要修改hostname需要完成以下操作

  1. 修改/etc/HOSTNAME文件
  2. 修改当前的$HOSTNAME变量
  3. 停掉/etc/rc.d/boot.localnet
  4. 启动 /etc/rc.d/boot.localnet
- 执行脚本如下：

```
# export HOSTNAME=myname
# echo $HOSTNAME>/etc/HOSTNAME
# /etc/rc.d/boot.localnet stop
# /etc/rc.d/boot.localnet start
```

参考网站

- https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html#YARN*on*Single_Node
- https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/ClusterSetup.html
- http://www.powerxing.com/install-hadoop-cluster/
- https://allanche.github.io/2016/01/11/hadoop%E6%95%85%E9%9A%9C%E8%AE%B0%E5%BD%95%E9%9B%86/
- http://abloz.com/hbase/book.html
- http://luchunli.blog.51cto.com/2368057/1672186
- https://www.cnblogs.com/yjmyzz/p/4280069.html
- https://hadoop.apache.org/docs/r2.6.0/hadoop-project-dist/hadoop-hdfs/hdfs-default.xml
- http://roclinux.cn/?page_id=3759
- http://linux.vbird.org/linux_basic/0310vi.php

Produced by luer@TJ2017.