欢迎点击这里的链接进入精彩的Linux公社网站

Linux公社(<u>www.Linuxidc.com</u>)于2006年9月25日注册并开通网站,Linux现在已经成为一种广受关注和支持的一种操作系统,IDC是互联网数据中心,LinuxIDC就是关于Linux的数据中心。

<u>Linux公社</u>是专业的Linux系统门户网站,实时发布最新Linux资讯,包括Linux、Ubuntu、Fedora、RedHat、红旗Linux、Linux教程、Linux认证、SUSE Linux、Android、Oracle、Hadoop、CentOS、MySQL、Apache、Nginx、Tomcat、Python、Java、C语言、OpenStack、集群等技术。

Linux公社(<u>LinuxIDC.com</u>)设置了有一定影响力的Linux专题栏目。

Linux公社 主站网址: <u>www.linuxidc.com</u> Linux公社资源站网址: <u>linux.linuxidc.com</u> Linux公社手机站网址: m.linuxidc.com

旗下网站: www.linuxidc.net

包括: <u>Ubuntu 专题 Fedora 专题 Android 专题 Oracle 专题 Hadoop 专题</u> RedHat 专题 SUSE 专题 红旗 Linux 专题 CentOS 专题



Linux 公社微信公众号: linuxidc com



微信扫一扫

Linxidc.com

订阅专业的最新Linux资讯及开源技术教程。

搜索微信公众号:linuxidc_com

Hadoop2.7.3+HBase1.2.5+ZooKeeper3.4.6搭建分布式集群环境详解

一、环境说明

个人理解:

ZooKeeper 可以独立搭建集群,HBase 本身不能独立搭建集群需要和 Hadoop 和 HDFS 整合。

集群环境至少需要 3 个节点(也就是 3 台服务器设备): 1 个 Master, 2 个 Slave, 节点之间局域网连接,可以相互 ping 通,下面举例说明,配置节点 IP 分配如下:

IP 角色

10.10.50.133 master

10.10.125.156 slave1

10.10.114.112 slave2

三个节点均使用 CentOS 6.5 系统,为了便于维护,集群环境配置项最好使用相同用户名、用户密码、相同 Hadoop、Hbase、zookeeper 目录结构。

注:

主机名和角色最好保持一致,如果不同也没关系,只需要在/etc/hosts 中配置好对应关系即可可以通过编辑/etc/sysconfig/network 文件来修改 hostname

软件包下载准备:

hadoop-2.7.3.tar.gz hbase-1.2.5-bin.tar.gz zookeeper-3.4.6.tar.gz jdk-8u111-linux-x64.rpm

因为是测试环境此次都使用 root 来操作,如果是生产环境建议使用其他用户如 hadoop,需要给目录授权为 hadoop chown -R hadoop.hadoop /data/yunva

二、准备工作

2.1 安装 JDK

在三台机器上配置 JDK 环境,下载 jdk-8u111-linux-x64.rpm 文件直接安装:

rpm -ivh jdk-8u111-linux-x64.rpm

修改配置文件 vim /etc/profile:

export JAVA_HOME=/usr/java/jdk1.8.0_111 # 不同的 jdk 路径需要修改此项

export PATH=\$JAVA_HOME/bin:\$PATH

export HADOOP_HOME=/data/yunva/hadoop-2.7.3

export HADOOP_INSTALL=\$HADOOP_HOME

export HADOOP_MAPRED_HOME=\$HADOOP_HOME

export HADOOP COMMON HOME=\$HADOOP HOME

 $\stackrel{\cdot}{\mathsf{export}}\,\mathsf{HADOOP_HDFS_HOME} = \mathsf{\$HADOOP_HOME}$

export YARN_HOME=\$HADOOP_HOME

export HADOOP_COMMON_LIB_NATIVE_DIR=\$HADOOP_HOME/lib/native

export PATH=\$PATH:\$HADOOP_HOME/sbin:\$HADOOP_HOME/bin

export HADOOP_SSH_OPTS="-p 48490" # 非默认 ssh 的 22 号端口需要添加此项,表示端口为 48490

因本次部署环境,jdk不同所以需要单独修改配置:

master

export JAVA_HOME=/usr/java/jdk1.8.0_111

slave1

export JAVA_HOME=/usr/java/jdk1.8.0_65

slave2 export JAVA_HOME=/usr/java/jdk1.8.0_102

然后重新加载配置文件使之生效:

source /etc/profile

2.2 添加 Hosts 映射关系

分别在三个节点上添加 hosts 映射关系:

vim /etc/hosts

添加的内容如下:

10.10.50.133 master 10.10.125.156 slave1 10.10.114.112 slave2

2.3 集群之间 SSH 无密码登陆

CentOS 默认安装了 ssh,如果没有你需要先安装 ssh。

集群环境的使用必须通过 ssh 无密码登陆来执行,本机登陆本机必须无密码登陆,主机与从机之间必须可以双向无密码登陆,从机与从机之间无限制。

2.3.1 设置 master 无密码自动登陆 slave1 和 slave2

主要有三步:

- ①生成公钥和私钥
- ②导入公钥到认证文件
- ③更改权限

ssh-keygen -t rsa

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

chmod 700 ~/.ssh && chmod 600 ~/.ssh/*

测试,第一次登录可能需要 yes 确认,之后就可以直接登录了:

ssh master

ssh slave1

ssh slave2

对于 slave1 和 slave2,进行无密码自登陆设置,操作同上。

也有个快捷的操作方式,当所有的服务器都 ssh-keygen -t rsa 生成公钥后,在 master 上操作无密码登陆 master/slave1/slave2 成功后,直接拷贝给其他主机即可

然后,将证书文件复制到其他机器的用户主目录下

scp -P 48490 authorized_keys master:/root/.ssh/

scp -P 48490 authorized_keys slave1:/root/.ssh/

scp -P 48490 authorized_keys slave2:/root/.ssh/

三、Hadoop 集群安装配置

这里会将 hadoop、hbase、zookeeper 的安装包都解压到/data/yunva/文件夹下,并重命名 安装目录如下:

/data/yunva/hadoop-2.7.3 /data/yunva/hbase-1.2.5 /data/yunva/zookeeper-3.4.6

3.1 修改 hadoop 配置

配置文件都在/data/yunva/hadoop-2.7.3/etc/hadoop/目录下

```
<configuration>
  property>
    <name>fs.default.name</name>
     <value>hdfs://master:9000</value>
  </property>
</configuration>
3.1.2 hadoop-env.sh
添加 JDK 路径,如果不同的服务器 jdk 路径不同需要单独修改:
export JAVA_HOME=/usr/java/jdk1.8.0_111
3.1.3 hdfs-site.xml
# 创建 hadoop 的数据和用户目录
# mkdir -p /data/yunva/hadoop-2.7.3/hadoop/name
# mkdir -p /data/yunva/hadoop-2.7.3/hadoop/data
<configuration>
  property>
    <name>dfs.name.dir</name>
    <value>/data/yunva/hadoop-2.7.3/hadoop/name</value>
  </property>
  property>
    <name>dfs.data.dir</name>
     <value>/data/yunva/hadoop-2.7.3/hadoop/data</value>
  </property>
  cproperty>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
</configuration>
3.1.4 mapred-site.xml
# mv mapred-site.xml.template mapred-site.xml
<configuration>
  cproperty>
    <name>mapred.job.tracker</name>
     <value>master:9001</value>
  </property>
</configuration>
3.1.5 修改 slaves 文件, localhost 改为
# cat /data/yunva/hadoop-2.7.3/etc/hadoop/slaves
slave1
slave2
注意: 三台机器上都进行相同的配置,都放在相同的路径下(如果 jdk 路径不同需要单独修改)
使用 scp 命令进行从本地到远程(或远程到本地)的轻松文件传输操作:
scp -r /data/yunva/hadoop-2.7.3/
                              slave1:/data/yunva
scp -r /data/yunva/hadoop-2.7.3/
                               slave2:/data/yunva
3.2 启动 hadoop 集群
进入 master 的/data/yunva/hadoop-2.7.3/目录,执行以下操作:
# bin/hadoop namenode -format
格式化 namenode,第一次启动服务前执行的操作,以后不需要执行。
然后启动 hadoop:
```

3.1.1 core-site.xml

sbin/start-all.sh

更多教程请到Linux公社 www.linuxidc.com

```
通过 jps 命令能看到除 jps 外有 3 个进程:
# jps
30613 NameNode
30807 SecondaryNameNode
887 Jps
30972 ResourceManager
hbase-env.sh(java 路径不同需要修改)
master
export JAVA HOME=/usr/java/jdk1.8.0 111
export HBASE_CLASSPATH=/data/yunva/hadoop-2.7.3/etc/hadoop/
export HBASE_MANAGES_ZK=false
export HBASE_SSH_OPTS="-p 48490" # 非默认 ssh 的 22 端口需要添加此项表示 ssh 为 48490
slave1
export JAVA_HOME=/usr/java/jdk1.8.0_65
export HBASE_CLASSPATH=/data/yunva/hadoop-2.7.3/etc/hadoop/
export HBASE_MANAGES_ZK=false
export HBASE SSH OPTS="-p 48490"
slave2
export JAVA_HOME=/usr/java/jdk1.8.0_102
export HBASE_CLASSPATH=/data/yunva/hadoop-2.7.3/etc/hadoop/
export HBASE_MANAGES_ZK=false
export HBASE_SSH_OPTS="-p 48490"
四、ZooKeeper 集群安装配置
可参考 CentOS 6.5 环境下 Zookeeper-3.4.6 集群环境部署及单机部署详解 https://www.linuxidc.com/Linux/2018-
03/151439.htm
五、HBase 集群安装配置
配置文件目录/data/yunva/hbase-1.2.5/conf
5.1 hbase-env.sh
export JAVA HOME=/usr/java/jdk1.8.0 111 # 如果 jdk 路径不同需要单独配置
export HBASE CLASSPATH=/data/yunva/hadoop-2.7.3/etc/hadoop/
export HBASE_MANAGES_ZK=false
export HBASE_SSH_OPTS="-p 48490" # ssh 端口非默认 22 需要修改
5.2 hbase-site.xml(保持一致)
<configuration>
  cproperty>
    <name>hbase.rootdir</name>
     <value>hdfs://master:9000/hbase</value>
  </property>
  property>
    <name>hbase.master</name>
     <value>master</value>
  </property>
  cproperty>
    <name>hbase.cluster.distributed</name>
    <value>true</value>
  </property>
  property>
    <name>hbase.zookeeper.property.clientPort</name>
     <value>2181</value>
  </property>
```

cproperty>

<name>hbase.zookeeper.quorum</name>

```
<value>master,slave1,slave2</value>
  </property>
  property>
    <name>zookeeper.session.timeout</name>
    <value>60000000</value>
  </property>
  cproperty>
    <name>dfs.support.append</name>
     <value>true</value>
  </property>
</configuration>
5.3 更改 regionservers
在 regionservers 文件中添加 slave 列表:
slave1
slave2
5.4 分发并同步安装包
将整个 hbase 安装目录都拷贝到所有 slave 服务器:
$ scp -P 48490 -r /data/yunva/hbase-1.2.5 slave1:/data/yunva/
$ scp -P 48490 -r /data/yunva/hbase-1.2.5 slave2:/data/yunva/
六、启动集群
1. 启动 ZooKeeper
/data/yunva/zookeeper-3.4.6/bin/zkServer.sh start
2. 启动 hadoop
/data/yunva/hadoop-2.7.3/sbin/start-all.sh
3. 启动 hbase
/data/yunva/hbase-1.2.5/bin/start-hbase.sh
4. 启动后, master 上进程和 slave 进程列表
[root@master ~]# jps
Jps
SecondaryNameNode # hadoop 进程
               # hadoop master 进程
NameNode
ResourceManager # hadoop 进程
             # hbase master 进程
HMaster
ZooKeeperMain
                # zookeeper 进程
[root@slave1 ~]# jps
Jps
ZooKeeperMain
                 # zookeeper 进程
DataNode
               # hadoop slave 进程
HRegionServer
                 # hbase slave 进程
```

欢迎点击这里的链接进入精彩的Linux公社网站

Linux公社(<u>www.Linuxidc.com</u>)于2006年9月25日注册并开通网站,Linux现在已经成为一种广受关注和支持的一种操作系统,IDC是互联网数据中心,LinuxIDC就是关于Linux的数据中心。

<u>Linux公社</u>是专业的Linux系统门户网站,实时发布最新Linux资讯,包括Linux、Ubuntu、Fedora、RedHat、红旗Linux、Linux教程、Linux认证、SUSE Linux、Android、Oracle、Hadoop、CentOS、MySQL、Apache、Nginx、Tomcat、Python、Java、C语言、OpenStack、集群等技术。

Linux公社(LinuxIDC.com)设置了有一定影响力的Linux专题栏目。

Linux公社 主站网址:www.linuxidc.comLinux公社资源站网址:linux.linuxidc.comLinux公社手机站网址:m.linuxidc.com

旗下网站: www.linuxidc.net

包括: <u>Ubuntu 专题 Fedora 专题 Android 专题 Oracle 专题 Hadoop 专题 RedHat 专题 SUSE 专题 红旗 Linux 专题 CentOS 专题</u>



Linux 公社微信公众号: linuxidc_com



微信扫一扫

Linxidc.com

订阅专业的最新Linux资讯及开源技术教程。

搜索微信公众号:linuxidc_com

```
5. 进入 hbase shell 进行验证
# cd /data/yunva/hbase-1.2.5/
[root@test6_vedio hbase-1.2.5]# bin/hbase shell
2017-04-28 09:51:51,479 WARN [main] util.NativeCodeLoader: Unable to load native-hadoop library for your
platform... using builtin-java classes where applicable
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/data/yunva/hbase-1.2.5/lib/slf4j-log4j12-
1.7.5.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/data/yunva/hadoop-2.7.3/share/hadoop/common/lib/slf4j-log4j12-
1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.5, rd7b05f79dee10e0ada614765bb354b93d615a157, Wed Mar 1 00:34:48 CST 2017
hbase(main):001:0> list
TABLE
0 row(s) in 0.2620 seconds
=>[]
hbase(main):003:0> create 'scores', 'grade', 'course'
0 row(s) in 1.3300 seconds
=> Hbase::Table - scores
hbase(main):004:0> list
TABLE
scores
1 row(s) in 0.0100 seconds
=> ["scores"]
6. 进入 zookeeper shell 进行验证
[root@test6 vedio zookeeper-3.4.6]# bin/zkCli.sh -server
2017-04-28 10:04:33,083 [myid:] - INFO [main:Environment@100] - Client
environment:zookeeper.version=3.4.6-1569965, built on 02/20/2014 09:09 GMT
2017-04-28 10:04:33,088 [myid:] - INFO [main:Environment@100] - Client
environment:host.name=test6_vedio
2017-04-28 10:04:33,088 [myid:] - INFO [main:Environment@100] - Client
environment:java.version=1.8.0_111
2017-04-28 10:04:33,091 [myid:] - INFO [main:Environment@100] - Client environment:java.vendor=Oracle
Corporation
2017-04-28 10:04:33,091 [myid:] - INFO [main:Environment@100] - Client
environment:java.home=/usr/java/jdk1.8.0 111/jre
2017-04-28 10:04:33,091 [myid:] - INFO [main:Environment@100] - Client
environment:java.class.path=/data/yunva/zookeeper-3.4.6/bin/../build/classes:/data/yunva/zookeeper-
3.4.6/bin/../build/lib/*.jar:/data/yunva/zookeeper-3.4.6/bin/../lib/slf4j-loq4j12-1.6.1.jar:/data/yunva/zookeeper-
3.4.6/bin/../lib/slf4j-api-1.6.1.jar:/data/yunva/zookeeper-3.4.6/bin/../lib/netty-
3.7.0.Final.jar:/data/yunva/zookeeper-3.4.6/bin/../lib/log4j-1.2.16.jar:/data/yunva/zookeeper-
3.4.6/bin/../lib/jline-0.9.94.jar:/data/yunva/zookeeper-3.4.6/bin/../zookeeper-3.4.6.jar:/data/yunva/zookeeper-
3.4.6/bin/../src/java/lib/*.jar:/data/yunva/zookeeper-3.4.6/bin/../conf:
2017-04-28 10:04:33,091 [myid:] - INFO [main:Environment@100] - Client
environment:java.library.path=/usr/java/packages/lib/amd64:/usr/lib64:/lib64:/lib:/usr/lib
2017-04-28 10:04:33,091 [myid:] - INFO [main:Environment@100] - Client environment:java.io.tmpdir=/tmp
2017-04-28 10:04:33,091 [myid:] - INFO [main:Environment@100] - Client environment:java.compiler=<NA>
2017\text{-}04\text{-}28\ 10\text{:}04\text{:}33\text{,}092\ [\text{myid:}]\text{- INFO}\ [\text{main:Environment}@100]\text{- Client environment:os.name} = \text{Linux}
2017-04-28 10:04:33,092 [myid:] - INFO [main:Environment@100] - Client environment:os.arch=amd64
2017-04-28 10:04:33,092 [myid:] - INFO [main:Environment@100] - Client environment:os.version=2.6.32-
```

2017-04-28 10:04:33,092 [myid:] - INFO [main:Environment@100] - Client environment:user.name=root

431.11.25.el6.ucloud.x86 64

2017-04-28 10:04:33,092 [myid:] - INFO [main:Environment@100] - Client

environment:user.dir=/data/yunva/zookeeper-3.4.6

2017-04-28 10:04:33,094 [myid:] - INFO [main:ZooKeeper@438] - Initiating client connection,

connectString=localhost:2181 sessionTimeout=30000

watcher=org.apache.zookeeper.ZooKeeperMain\$MyWatcher@446cdf90

Welcome to ZooKeeper!

2017-04-28 10:04:33,128 [myid:] - INFO [main-SendThread(localhost:2181):ClientCnxn\$SendThread@975] - Opening socket connection to server localhost/127.0.0.1:2181. Will not attempt to authenticate using SASL (unknown error)

JLine support is enabled

2017-04-28 10:04:33,209 [myid:] - INFO [main-SendThread(localhost:2181):ClientCnxn\$SendThread@852] - Socket connection established to localhost/127.0.0.1:2181, initiating session

 $2017-04-28\ 10:04:33,218\ [myid:]$ - INFO [main-SendThread(localhost:2181):ClientCnxn\$SendThread@1235] - Session establishment complete on server localhost/127.0.0.1:2181, sessionid = 0x35bb23d68ba0003, negotiated timeout = 30000

WATCHER::

WatchedEvent state:SyncConnected type:None path:null

[zk: localhost:2181(CONNECTED) 0] ls /

zookeeper hbase

[zk: localhost:2181(CONNECTED) 0] ls /hbase

[replication, meta-region-server, rs, splitWAL, backup-masters, table-lock, flush-table-proc, region-in-transition, online-snapshot, master, running, recovering-regions, draining, namespace, hbaseid, table]

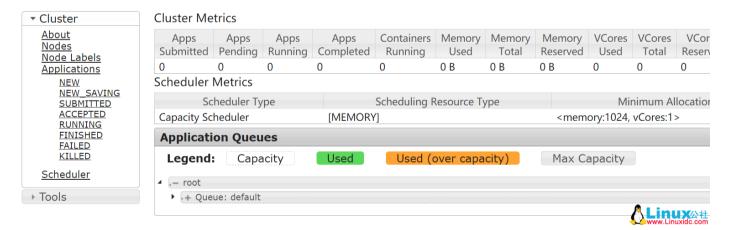
如果访问默认的 http 管理端口页面可以看到集群的情况

hadoop:

http://IP:8088/cluster/cluster







hbase:

http://IP:16010/master-status



Master test6_vedio

Region Servers

Base Stats	Memory	Requests	Storefiles	Compactions					
ServerName				Start time	;	Version	Requests Per Second		
test3,16020,1493344230464				Fri Apr 28	09:50:30 CST 2017	1.2.5	0		
test4_haili_dev,16020,1493344230453				Fri Apr 28	09:50:30 CST 2017	1.2.5	0		
Total:2							0		
								Linux公社 www.Linuxidc.com	

hdfs:

http://IP:50070/dfshealth.html#tab-overview



Datanode Information

In operation

Node	Last contact	Admin State	Capacity	Used	Non DFS Used	Remaining	Blocks	Block pool used	Failed Vol
test3:50010 (10.10.125.156:50010)	0	In Service	19.69 GB	184 KB	13.99 GB	5.7 GB	14	184 KB (0%)	0
test4_haili_dev:50010 (10.10.114.112:50010)	0	In Service	196.86 GB	184 KB	31.37 GB	165.49 GB	14	184 KB (0%)	0



完整文档可以到 Linux 公社资源站下载:

免费下载地址在 http://linux.linuxidc.com/

用户名与密码都是 www.linuxidc.com

具体下载目录在 /2018 年资料/3 月/19 日/Hadoop2.7.3+HBase1.2.5+ZooKeeper3.4.6 搭建分布式集群环境详解/

下载方法见 http://www.linuxidc.com/Linux/2013-07/87684.htm

Hadoop2.3-HA 高可用集群环境搭建 https://www.linuxidc.com/Linux/2017-03/142155.htm
Hadoop 项目之基于 CentOS7 的 Cloudera 5.10.1(CDH)的安装部署 https://www.linuxidc.com/Linux/2017-04/143095.htm

Hadoop2.7.2 集群搭建详解(高可用) https://www.linuxidc.com/Linux/2017-03/142052.htm
使用 Ambari 来部署 Hadoop 集群 (搭建内网 HDP 源) https://www.linuxidc.com/Linux/2017-02/140783.htm
Ubuntu 14.04 下 Hadoop 2.7.2 https://www.linuxidc.com/Linux/2017-08/146232.htm
Ubuntu 16.04 上构建分布式 Hadoop-2.7.3 集群 https://www.linuxidc.com/Linux/2017-06/144932.htm
Hadoop2.7.3+Spark2.1.0 完全分布式集群搭建过程 https://www.linuxidc.com/Linux/2017-06/144926.htm

更多 Hadoop 相关信息见 Hadoop 专题页面 http://www.linuxidc.com/topicnews.aspx?tid=13

本文永久更新链接地址: https://www.linuxidc.com/Linux/2018-03/151438.htm