**HA机制配置**

关闭已经启动的所有进程

# 配置hdfs-site.xml

指定命名空间的名称：

<property>

<name>dfs.nameservices</name>

<value>ns1</value>

</property>

指定两个namenodeID名称：

<property>

<name>dfs.ha.namenodes.ns1</name>

<value>nn1,nn2</value>

</property>

指定namenode所在机器位置以及交互端口号

<property>

<name>dfs.namenode.rpc-address.ns1.nn1</name>

<value>MyDream1:8020</value>

</property>

<property>

<name>dfs.namenode.rpc-address.ns1.nn2</name>

<value>MyDream2:8020</value>

</property>

指定namenode所在机器位置以及外部交互端口号

<property>

<name>dfs.namenode.http-address.ns1.nn1</name>

<value>MyDream1:50070</value>

</property>

<property>

<name>dfs.namenode.http-address.ns1.nn2</name>

<value>MyDream2:50070</value>

</property>

指定日志节点的机器以及端口号还有命名空间名称

<property>

<name>dfs.namenode.shared.edits.dir</name

<value>qjournal:// MyDream1:8485; MyDream2:8485; MyDream3:8485/ns1</value>

</property>

指定并创建日志节点存储的位置路径：

<property>

<name>dfs.journalnode.edits.dir</name>

<value>/home/hadoop/hadoop-2.5.0/data/dfs/jn</value>

</property>

指定代理访问文件系统

<property>

<name>dfs.client.failover.proxy.provider.ns1</name> <value>org.apache.hadoop.hdfs.server.namenode.ha.ConfiguredFailoverProxyProvider</value>

</property>

选择指定隔离的方式，选择的是SSH远程协议

<property>

<name>dfs.ha.fencing.methods</name>

<value>sshfence</value>

</property>

<property>

<name>dfs.ha.fencing.ssh.private-key-files</name>

<value>/home/hadoop/.ssh/id\_rsa</value>

</property>

# 配置core-site.xml

指定命名空间

<property>

<name>fs.defaultFS</name>

<value>hdfs://ns1</value>

</property>

# 将配置分发到各个节点

$ scp –r /home/hadoop/hadoop-2.5.0/etc/hadoop MYDream2:/home/hadoop/hadoop-2.5.0/etc/

$ scp –r /home/hadoop/hadoop-2.5.0/etc/hadoop MYDream3:/home/hadoop/hadoop-2.5.0/etc/

# 启动服务

注意先启动ZK和日志节点

$zkServer.sh start

$ sbin/hadoop-daemon.sh start journalnode

# 格式化namenode

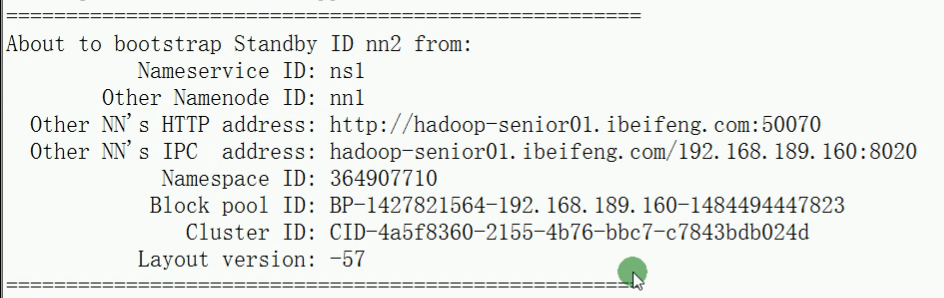
$ hdfs namenode –format

启动namenode

# 同步namenode

在另一台指定的namenode执行

$hdfs namenode –bootstrapStandby



在该机器启动namenode

# 切换namenode状态对外提供服务

在选定的namenode机器执行

Namenode1执行$hdfs haadmin transitionToActive nn1

Or

Namenode2执行$hdfs haadmin transitionToActive nn2

# 自动故障转移

1、ZKFC：故障转移监听器

NameNode(Active) NameNode(Standby)

ZKFC ZKFC

## 配置hdfs-site.xml

指定是否开启自动故障转移功能

<property>

<name>dfs.ha.automatic-failover.enabled</name>

<value>true</value>

</property>

指定zookeeper节点机器位置以及端口号

<property>

<name>ha.zookeeper.quorum</name> <value>MyDream1:2181,MyDream2:2181,MyDream3:2181</value>

</property>

## 将配置分发到各个节点

$ scp –r /home/hadoop/hadoop-2.5.0/etc/hadoop MYDream2:/home/hadoop/hadoop-2.5.0/etc/

$ scp –r /home/hadoop/hadoop-2.5.0/etc/hadoop MYDream3:/home/hadoop/hadoop-2.5.0/etc/

## 初始化HA状态

先要启动zookeeper服务

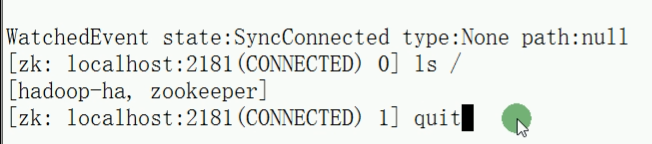
zkServer start

初始化HA

hdfs zkfc –formatZK

查看状态

zkCli.Sh



5、启动HDFS相关的服务进程

sbin/start-dfs.sh -》同时将ZKFC监听器启动的

$ sbin/hadoop-daemon.sh start zkfc

6、测试自动跳转

-》可以测试运行一个Wordcount在yarn上，关闭其中一个namenode的进程，观察是否切换成功

# 自动切换流程

1、ZKFC通过RPC请求获取namenode状态

2、ZKFC链接ZK，完成选举，保持心跳

3、另一台的ZKFC监控本机的namenode状态

4、由于某种因素，对外提供服务的namenode当机出现异常

5、ZKFC会在下一个监控的周期中检测到namenode失效的信息，触发后续的故障处理的操作

6、将失效的信息给到ZK

7、ZK会与另一台的节点ZKFC进行通信

8、standby节点ZKFC进入切换的流程，隔离，将失效的namenode隔离出集群之外

9、ZKFC进行选举，确认standby节点作为一个新的active节点

10、完成选举之后，standby节点就切换到了active节点