**hadoop2.7.4格式化遇到的问题**

1 概述

  解决hadoop启动hdfs时，datanode无法启动的问题。错误为：

**[plain]** [view plain](http://blog.csdn.net/liuxinghao/article/details/40121843) [copy](http://blog.csdn.net/liuxinghao/article/details/40121843)

1. FATAL org.apache.hadoop.hdfs.server.datanode.DataNode: Initialization failed for Block pool <registering> (Datanode Uuid unassigned) service to test/192.168.1.104:9000. Exiting.

2 问题描述

  执行start-all.sh后，根据打印日志，可以看到分别执行了NameNode、DataNode的操作。

**[plain]** [view plain](http://blog.csdn.net/liuxinghao/article/details/40121843) [copy](http://blog.csdn.net/liuxinghao/article/details/40121843)

1. Starting namenodes on [test]
2. test: starting namenode, logging to /usr/local/hadoop/log/hadoop/hadoop-hadoop-namenode-test.out
3. test: starting datanode, logging to /usr/local/hadoop/log/hadoop/hadoop-hadoop-datanode-test.out
4. test02: starting datanode, logging to /usr/local/hadoop/log/hadoop/hadoop-hadoop-datanode-test02.out

  但是执行jps查看启动结果时，返现DataNode并没有启动。

**[plain]** [view plain](http://blog.csdn.net/liuxinghao/article/details/40121843) [copy](http://blog.csdn.net/liuxinghao/article/details/40121843)

1. 10256 ResourceManager
2. 29634 NameNode
3. 29939 SecondaryNameNode
4. 30054 Jps
5. 10399 NodeManager

3 查找问题

  很是费解，刚刚还能够正常运行，并且执行了wordcount的测试程序。于是回想了一下刚才的操作，执行了dfs格式化(hdfs namenode -format和hdfs datanode -format)，然后重新启动就出现了这个情况。难道与格式化有关？于是查看日志：

**[plain]** [view plain](http://blog.csdn.net/liuxinghao/article/details/40121843) [copy](http://blog.csdn.net/liuxinghao/article/details/40121843)

1. FATAL org.apache.hadoop.hdfs.server.datanode.DataNode: Initialization failed for Block pool <registering> (Datanode Uuid unassigned) service to test/192.168.1.104:9000

  根据日志描述，原因是datanode的uuid未分配。

  原因找到，看看是否如日志描述的这样。

  打开hdfs-site.xml中关于datanode和namenode对应的目录，分别打开其中的current/VERSION文件，进行对比。

${datanode}/current/VERSION:

**[plain]** [view plain](http://blog.csdn.net/liuxinghao/article/details/40121843) [copy](http://blog.csdn.net/liuxinghao/article/details/40121843)

1. storageID=DS-be8dfa2b-17b1-4c9f-bbfe-4898956a39ed
2. clusterID=CID-200e6206-98b5-44b2-9e48-262871884eeb
3. cTime=0
4. datanodeUuid=406b6d6a-0cb1-453d-b689-9ee62433b15d
5. storageType=DATA\_NODE
6. layoutVersion=-55

${namenode}/current/VERSION:

**[plain]** [view plain](http://blog.csdn.net/liuxinghao/article/details/40121843) [copy](http://blog.csdn.net/liuxinghao/article/details/40121843)

1. namespaceID=670379
2. clusterID=CID-a3938a0b-57b5-458d-841c-d096e2b7a71c
3. cTime=0
4. storageType=NAME\_NODE
5. blockpoolID=BP-325596647-127.0.1.1-1407429078192
6. layoutVersion=-56

  发现datanode与namenode中记录的clusterid不一致，于是修改datanode的VERSION文件中的clusterID，使与namenode保持一致，然后启动dfs（执行start-dfs.sh），在执行jps查看启动情况，发现全部正常启动。

**[plain]** [view plain](http://blog.csdn.net/liuxinghao/article/details/40121843) [copy](http://blog.csdn.net/liuxinghao/article/details/40121843)

1. 10256 ResourceManager
2. 30614 NameNode
3. 30759 DataNode
4. 30935 SecondaryNameNode
5. 31038 Jps
6. 10399 NodeManager

4 分析问题原因

  执行hdfs namenode -format后，current目录会删除并重新生成，其中VERSION文件中的clusterID也会随之变化，而datanode的VERSION文件中的clusterID保持不变，造成两个clusterID不一致。

  所以为了避免这种情况，可以再执行的namenode格式化之后，**删除**datanode的current文件夹，或者**修改**datanode的VERSION文件中出clusterID与namenode的VERSION文件中的clusterID一样，然后重新启动dfs。