

情境二 Namenode 单节点故障的风险预防

Namenode 是 HDFS 系统中负责存储元数据的节点，主要功能是管理和调度 Datanode 的存储和运行。Namenode 单点故障是指，在 Hadoop 分布式文件系统工作过程中，由 Namenode 节点发生错误而引发系统崩溃的问题，这通常是由 Namenode 存储过量或节点故障所致。在工作场景中，此类问题一旦发生，可能会影响整个集群无法正常工作。

情境描述

针对 Namenode 单点故障的隐患，Hadoop 官方从 Hadoop2.0 版本开始，为用户提供了 HDFS High Availability(HDFS 高可用)架构来解决单点故障问题。HDFS HA 是针对 Namenode 单点元数据备份问题，提出的解决方案，其工作原理是，在传统的 Namenode 单节点基础上创建第二个 Namenode 节点，其中一个 Namenode 作为 Active Namenode，执行当前集群上的元数据服务，另一台 Standby Namenode 节点是备用节点，Standby Namenode 平时只作为 Slave 服务器执行简单的任务，在 Active Namenode 发生故障后切换为 Active 状态。两个节点都可访问、共享存储设备上的文件目录，然而只有 Active Namenode 对 Client 提供读写服务。

分析

在开始配置 HA 之前，要先检查准备环境，我们共有三个节点，主机名分别设为“bigdata1”、“bigdata2”、“bigdata3”，每个节点上已部署了 Hadoop、Java、Zookeeper，三台节点中，其中 bigdata1 和 bigdata2 作为 Name node，在 Hadoop 集群启动后，bigdata1 作为 Active Namenode，bigdata2 节点作为 Standby Namenode。

本节实施过程分以下几步：

- 一、配置环境变量
- 二、配置 Hadoop 参数
- 三、配置并启动 Zookeeper
- 四、启动 Hadoop HA 集群

实施过程

一、配置环境变量

由于主机名和 IP 地址不同，以下所有操作都须在每个节点上单独配置，在配置前建议先在每台虚拟机中执行“ifconfig”命令来查看和确定 IP 地址。

1、修改主机名

为方便集群节点角色的区分，需要修改各节点的主机名。使用“sudo gedit /etc/hostname”命令，或“vim /etc/hostname”命令，打开“/etc/hostname”文件，在其中输入想要设置的当前节点的主机名，图 6-2-1 中，主机名原本是“virtualBox”，我们将当前节点的主机名改为“bigdata1”。



图 6-2-1 设置 hostname

2、主机域名解析

打开 “/etc/hosts” 文件，将每个节点的 IP 地址和节点的主机名加入文件中。

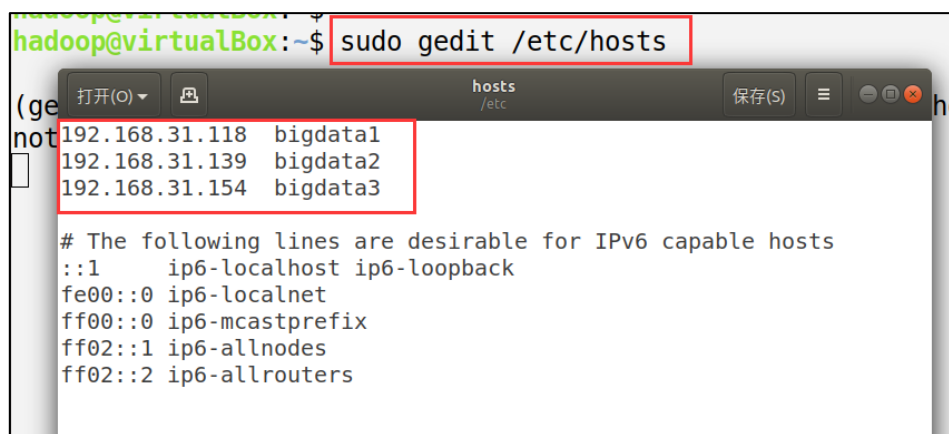


图 6-2-2 将节点 IP 地址和主机域名添加至 hosts 文件

3、配置环境变量

打开 “/etc/profile” 文件，将 Hadoop 和 Java 路径添加进去。注意，图 6-2-3 中，我们当前节点的 Hadoop 文件目录在 “/usr/local/hadoop” 中，Java 配置在 “/usr/lib/jvm/default-java” 目录中。

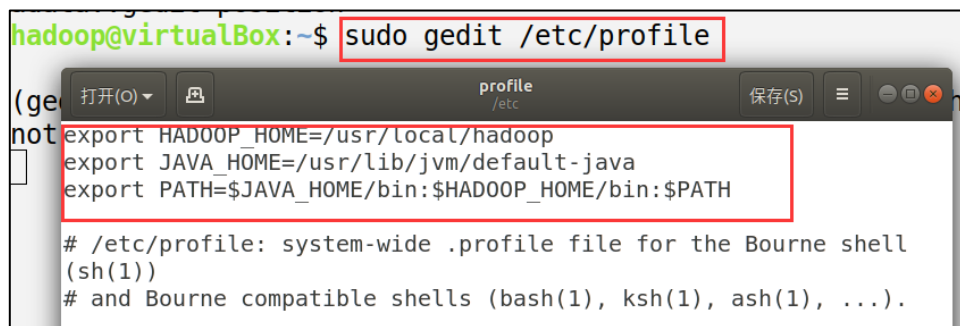


图 6-2-3 添加环境变量

4、配置各节点之间 ssh 免密登录

前面章节中我们已学过设置 ssh 免密登录，以及 ssh 秘钥生成的方法，这里不

再讲解。本节学习为 Hadoop 的两个 Namenode 节点添加远程登录其他节点的命令。由于当前集群内有两个 Namenode 节点, 所以要将两个节点的公钥文件都添加至所有节点中。在两个节点上都执行 “ssh-copy-id -f 用户名@主机名” 命令, 添加成功后, 会有图 6-2-4 中提示消息, 此时执行 “ssh 用户名@主机名” 命令, 可远程登录其他节点。

```
hadoop@virtualBox:~$ ssh-copy-id -f hadoop@bigdata2
Number of key(s) added: 1
Now try logging into the machine, with: "ssh 'hadoop@bigdata2'"
and check to make sure that only the key(s) you wanted were added.
```

图 6-2-4 设置各节点间 ssh 免密登录

二、配置 Hadoop 参数

Hadoop 的相关参数较多, 配置时容易出错, 可先在一台节点上配置, 最后通过 “scp” 命令将文件目录复制到其他节点上。以下步骤 1-8 都可在一台节点上设置, 然后根据步骤 9 的操作提示, 将 Hadoop 文件目录复制到其他节点中。

1、配置 slaves 文件

slaves 文件下存放的是部署 Datanode 的主机名, 我们可以将三台主机都部署 Datanode。slaves 文件存放在 Hadoop 目录下的 “./etc/hadoop” 中, 打开 slaves 文件并将三个主机名都添加进去。

```
hadoop@bigdata1:~$ cd /usr/local/hadoop/etc/hadoop/
hadoop@bigdata1:/usr/local/hadoop/etc/hadoop$ sudo gedit slaves
```

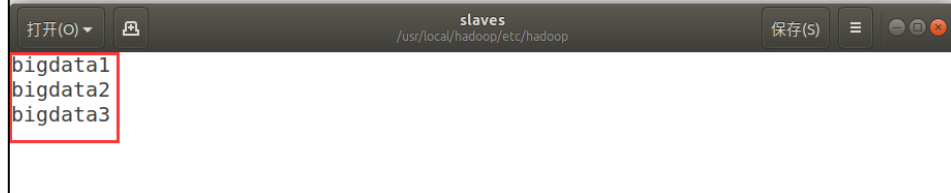


图 6-2-5 配置 slaves 文件

2、配置 core-site.xml 文件

找到并打开 core-site.xml, 加入如下配置。注意, 要将 fs.defaultFS 的值设置为将要作为 Master 节点的主机名地址, 如图 6-2-6 所示。

```
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://bigdata1:9000</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>file:/usr/local/hadoop/tmp</value>
    <description>Abase for other temporary directories.</description>
  </property>
</configuration>
```

图 6-2-6 配置 core-site.xml 文件

3、配置 hdfs-site.xml 文件

打开 hdfs-site.xml 文件, 将 dfs.namenode.secondary 地址与 core-site.xml 中的 fs.defaultFS 同步, 注意, 当前集群中只有 2 个 Slave 节点, 因此

dfs.replication 设置为 2。

```
<configuration>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>bigdata1:50090</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/usr/local/hadoop/tmp/dfs/name</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/usr/local/hadoop/tmp/dfs/data</value>
  </property>
</configuration>
```

图 6-2-7 配置 hdfs-site.xml 文件

4、配置 yarn-site.xml 文件

找到并打开 yarn-site.xml，加入图 6-2-1 中配置。将 yarn 部署在 bigdata2 节点中，因此要将 yarn.resourcemanager 的值设置为 bigdata2。

```
<configuration>
  <property>
    <name>yarn.resourcemanager.hostname</name>
    <value>bigdata2</value>
  </property>
  <property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
  </property>
</configuration>
```

图 6-2-8 yarn-site.xml 文件

5、配置 mapred-site.xml 文件

找到并打开 mapred-site.xml，加入如下配置。

```
<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.address</name>
    <value>bigdata1:10020</value>
  </property>
  <property>
    <name>mapreduce.jobhistory.webapp.address</name>
    <value>bigdata2:19888</value>
  </property>
</configuration>
```

图 6-2-9 mapred-site.xml 文件

6、复制 Hadoop 文件目录到其他节点

Hadoop 相关文件配置完成后，可使用“`scp -r` 当前 Hadoop 文件夹路径 要复制到的节点地址：目的路径”命令将 Hadoop 目录复制到 bigdata2 和 bigdata3 的服务器中。

```
hadoop@bigdata1:~$  
hadoop@bigdata1:~$ scp -r /usr/local/hadoop hadoop@bigdata2:/usr/local/hadoop  
libhdfs.so.0.0.0 100% 274KB 3.9MB/s 00:00  
libhdfs.so 100% 274KB 3.9MB/s 00:00  
libhadooppipes.a 100% 1889KB 4.3MB/s 00:00  
libhadoop.so.1.0.0 100% 780KB 4.4MB/s 00:00  
libhdfs.a 100% 433KB 4.4MB/s 00:00  
libhadoop.a 100% 1338KB 4.5MB/s 00:00  
libhadoop.so 100% 780KB 4.1MB/s 00:00
```

图 6-2-10 复制 Hadoop 文件目录到其他节点

7、为 Hadoop 文件目录授予 hadoop 用户权限

为了避免后面启动或访问 Hadoop 集群时因创建文件等授权问题影响整个 Hadoop 正常运行，为 3 个节点的 Hadoop 文件都授予 hadoop 用户权限。在每个节点上执行“`sudo chown -R 用户 授权路径`”命令。

```
hadoop@bigdata1:~$ sudo chown -R hadoop /usr/local/hadoop  
[sudo] hadoop 的密码：  
hadoop@bigdata1:~$
```

图 6-2-11 为 Hadoop 文件目录授予 hadoop 用户权限

三、配置并启动 Zookeeper

为搭建 Zookeeper 集群，以下 Zookeeper 的配置和命令操作要在集群中所有 3 台服务器上都执行。

1、配置 zoo.cfg 文件

使用“`cp 旧文件名 新文件名`”命令，将“`zookeeper/conf`”路径下的“`zoo_sample.cfg`”文件名修改为“`zoo.cfg`”。

```
hadoop@bigdata1:/usr/local/zookeeper/zookeeper-3.4.12/conf$ cp zoo_sample.cfg zoo.cfg
```

图 6-2-12 修改 zoo.cfg 文件名

打开“`zoo.cfg`”文件，将所有节点的 IP 地址和主机名信息添加到 `zoo.cfg` 中，添加信息如图 6-2-13 所示。

```
dataDir=/usr/local/zookeeper/zkdata  
server.1=bigdata1:2888:3888  
server.2=bigdata2:2888:3888  
server.3=bigdata3:2888:3888
```

图 6-2-13 编辑 zoo.cfg

2、配置 myid 文件

切换到“`./zkdata`”目录中（该目录是我们手动创建），创建名为“`myid`”的文件，在其中编辑 zookeeper 节点编号，注意，此编号要与“`zoo.cfg`”中编辑的“`server.`”后的编号对应。例如图 6-2-13 中，当前主机为 bigdata1，为其分配的服务器为“`server.1`”，因此在“`myid`”文件中要输入“1”。同样，要为另外两台服务器创建并编辑“`myid`”文件。

3、启动 Zookeeper 集群

分别在三台服务器上的 zookeeper 的 bin 目录下，使用“zkServer.sh start”命令，启动 Zookeeper 集群。

```
hadoop@bigdata1:/usr/local/zookeeper/zookeeper-3.4.12/bin$ ./zkServer.sh start
ZooKeeper JMX enabled by default
Using config: /usr/local/zookeeper/zookeeper-3.4.12/bin/../conf/zoo.cfg
Starting zookeeper ... STARTED
```

图 6-2-14 启动 Zookeeper 集群

四、启动 Hadoop HA 集群

1、启动 journalnode

切换到“./Hadoop /sbin”目录，输入“./hadoop_daemon.sh start journalnode”命令，来启动 journalnode。

```
hadoop@bigdata1:/usr/local/hadoop/sbin$ ./hadoop-daemon.sh start journalnode
starting journalnode, logging to /usr/local/hadoop/logs/hadoop-hadoop-journalnode-bigdata1.out
```

图 6-2-15 启动 journalnode

2、格式化 Namenode

若是初次部署并启动 Hadoop 集群，需要先格式化 namenode。

```
hadoop@bigdata1:/usr/local/hadoop/sbin$ hdfs namenode -format
```

图 6-2-16 格式化 Namenode

3、同步两个 Namenode 节点

在 bigdata1 节点上启动 namenode，如前面所介绍，bigdata2 节点将作为 Active Namenode 存放的服务器。

```
hadoop@bigdata1:/usr/local/hadoop$ cd/sbin/
hadoop@bigdata1:/usr/local/hadoop/sbin$ ./hadoop-daemon.sh start namenode
```

图 6-2-17 在 bigdata1 节点上启动 namenode

在 bigdata2 节点上执行 Standby Namenode 的同步机制。

```
hadoop@bigdata2:/usr/local/hadoop/bin$ hdfs namenode -bootstrapStandby
19/08/09 05:03:38 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = bigdata2/192.168.31.139
STARTUP_MSG: args = [-bootstrapStandby]
STARTUP_MSG: version = 2.6.5
```

图 6-2-18 执行 Standby Namenode 的同步机制

同样，bigdata2 节点上启动 namenode。

```
hadoop@bigdata2:/usr/local/hadoop/sbin$ ./hadoop-daemon.sh start namenode
```

图 6-2-19 在 bigdata2 节点上启动 namenode

4、启动 Hadoop 集群

切换到“./hadoop/sbin”目录下，执行“start-all.sh”命令，启动 Hadoop 集群。

五、查看服务启动情况

本节主要讲解如何大家 Hadoop HA，启动 Hadoop 集群后，可在浏览器登入 <http://bigdata1:50070> 和 <http://bigdata2:50070> 查看两个 Namenode 节点状态。如

图 6-2- 20 和图 6-2- 21 所示。

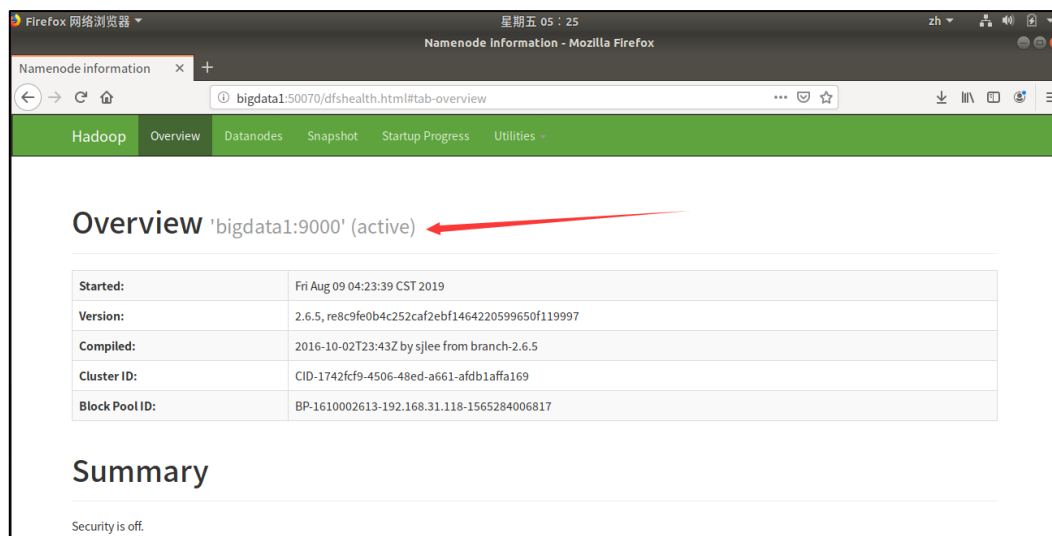


图 6-2- 20 查看 Active Namenode 状态

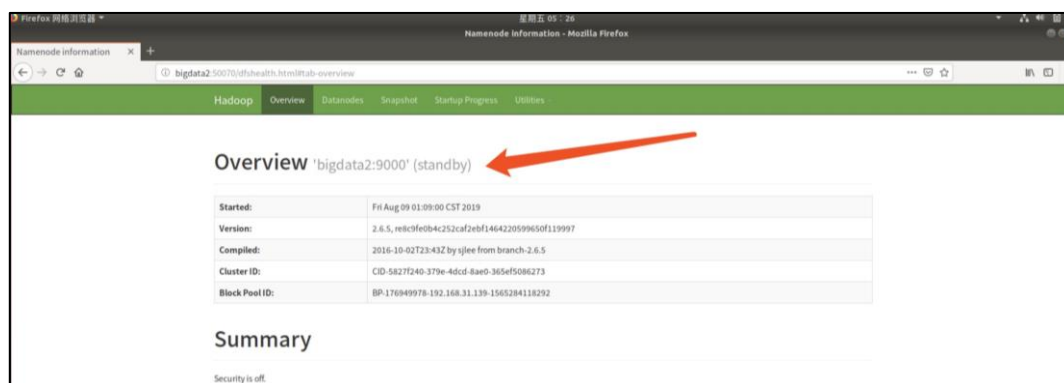


图 6-2- 21 查看 Standby Namenode 状态