Hadoop配置过程

环境：

Ubuntu 16.04 + Hadoop-2.9.0 + Jdk1.8.0\_162

机器配置：

|  |  |  |  |
| --- | --- | --- | --- |
| **主机** | **CPU及内存** | **硬盘** | **功能** |
| NameNode | 1核1G | 20G+80G | NameNode结点 |
| Rscmng | 1核1G | 20G+80G | ResourceManager结点  以及MapReduce JobHistory Server结点 |
| Datanode1 | 1核4G | 20G+80G | 作为DataNode结点，NodeManager结点，用于hdfs的存储以及mapreduce任务的执行 |
| Datanode2 | 1核4G | 20G+80G | 同上 |
| Datanode3 | 1核4G | 20G+80G | 同上 |
| Datanode4 | 1核4G | 20G+80G | 同上 |
| MongoDB | 1核1G | 20G | 作为MongoDB结点存储数据 |

1. Java环境安装与配置

下载jdk并解压到/opt/目录下，配置环境变量

vim /etc/profile.d/hadoop.sh

1. export JAVA\_HOME=/usr/local/jdk1.8.0\_162
2. export JRE\_HOME=${JAVA\_HOME}/jre
3. export CLASSPATH=.:${JAVA\_HOME}/lib:${JRE\_HOME}/lib:$CLASSPATH
4. export PATH=${JAVA\_HOME}/bin:${JRE\_HOME}/bin:$PATH

source /etc/profile.d/hadoop.sh使配置文件生效

输入java –version判断java是否安装成功

1. Hadoop环境安装与配置

下载hadoop-2.9.0安装包并解压到/opt/目录下

配置hadoop环境变量

vim /opt/hadoop-2.9.0/etc/hadoop/hadoop-env.sh

1. # The java implementation to use.
2. export JAVA\_HOME=/usr/local/jdk1.8.0\_162
3. export HADOOP\_PID\_DIR=/opt/hadoop
4. export HADOOP\_SECURE\_DN\_PID\_DIR=/opt/hadoop

修改配置文件core-site.xml

vim /opt/hadoop-2.9.0/etc/hadoop/core-site.xml

1. **<configuration>**
2. **<property>**
3. **<name>**hadoop.tmp.dir**</name>**
4. **<value>**/hadoop/hdfs/**</value>**
5. **<description>**Abase for other temporary directories.**</description>**
6. **</property>**
8. **<property>**
9. **<name>**fs.defaultFS**</name>**
10. **<value>**hdfs://namenode:9000**</value>**
11. **</property>**
13. **<property>**
14. **<name>**io.file.buffer.size**</name>**
15. **<value>**131072**</value>**
16. **<description>**Size of read/write buffer used in SequenceFiles.**</description>**
17. **</property>**
18. **</configuration>**

修改配置文件hdfs-site.xml

vim /opt/hadoop-2.9.0/etc/hadoop/hdfs-site.xml

1. **<configuration>**
2. <!-- Configurations for NameNode:  -->
4. **<property>**
5. **<name>**hadoop.tmp.dir**</name>**
6. **<value>**/hadoop/hdfs/**</value>**
7. **<description>**Abase for other temporary directories.**</description>**
8. **</property>**
10. **<property>**
11. **<name>**dfs.namenode.safemode.extension**</name>**
12. **<value>**30000**</value>**
13. **</property>**
15. **<property>**
16. **<name>**dfs.replication**</name>**
17. **<value>**1**</value>**
18. **</property>**
20. **<property>**
21. **<name>**dfs.namenode.name.dir**</name>**
22. **<value>**file:/hadoop/hdfs/dfs/name,file:/hadoop/hdfs2/dfs/name**</value>**
23. **</property>**
25. **<property>**
26. **<name>**dfs.blocksize**</name>**
27. **<value>**268435456**</value>**
28. **</property>**
30. **<property>**
31. **<name>**dfs.namenode.handler.count**</name>**
32. **<value>**10**</value>**
33. **</property>**
35. <!-- Configurations for DataNode: -->
37. **<property>**
38. **<name>**dfs.datanode.data.dir**</name>**
39. **<value>**file:/hadoop/hdfs/dfs/data**</value>**
40. **</property>**
42. **<property>**
43. **<name>**dfs.permissions**</name>**
44. **<value>**false**</value>**
45. **</property>**
47. **</configuration>**

修改配置文件yarn-site.xml

vim /opt/hadoop-2.9.0/etc/hadoop/yarn-site.xml

1. **<configuration>**
3. **<property>**
4. **<name>**yarn.acl.enable**</name>**
5. **<value>**false**</value>**
6. **</property>**
8. **<property>**
9. **<name>**yarn.admin.acl**</name>**
10. **<value>**\***</value>**
11. **</property>**
13. **<property>**
14. **<name>**yarn.log-aggregation-enable**</name>**
15. **<value>**false**</value>**
16. **</property>**
18. **<property>**
19. **<name>**yarn.resourcemanager.address**</name>**
20. **<value>**rscmng:8032**</value>**
21. **</property>**
23. **<property>**
24. **<name>**yarn.resourcemanager.scheduler.address**</name>**
25. **<value>**rscmng:8030**</value>**
26. **</property>**
28. **<property>**
29. **<name>**yarn.resourcemanager.resource-tracker.address**</name>**
30. **<value>**rscmng:8031**</value>**
31. **</property>**
33. **<property>**
34. **<name>**yarn.resourcemanager.admin.address**</name>**
35. **<value>**rscmng:8033**</value>**
36. **</property>**
38. **<property>**
39. **<name>**yarn.resourcemanager.webapp.address**</name>**
40. **<value>**rscmng:8088**</value>**
41. **</property>**

44. **<property>**
45. **<name>**yarn.resourcemanager.scheduler.class**</name>**
46. **<value>**org.apache.hadoop.yarn.server.resourcemanager.scheduler.capacity.CapacityScheduler**</value>**
47. **</property>**
49. **<property>**
50. **<name>**yarn.scheduler.minimum-allocation-mb**</name>**
51. **<value>**1024**</value>**
52. **</property>**
54. **<property>**
55. **<name>**yarn.scheduler.maximum-allocation-mb**</name>**
56. **<value>**3584**</value>**
57. **</property>**
59. **<property>**
60. **<name>**yarn.nodemanager.resource.memory-mb**</name>**
61. **<value>**4096**</value>**
62. **</property>**
64. **<property>**
65. **<name>**yarn.nodemanager.vmem-pmem-ratio**</name>**
66. **<value>**5**</value>**
67. **</property>**
69. **<property>**
70. **<name>**yarn.nodemanager.local-dirs**</name>**
71. **<value>**/hadoop/yarn/nm-local-dir**</value>**
72. **</property>**
74. **<property>**
75. **<name>**yarn.nodemanager.log-dirs**</name>**
76. **<value>**/hadoop/yarn/userlogs**</value>**
77. **</property>**
79. **<property>**
80. **<name>**yarn.nodemanager.log.retain-seconds**</name>**
81. **<value>**10800**</value>**
82. **</property>**

85. <!--t Site specific YARN configuration properties -->
86. **<property>**
87. **<name>**yarn.nodemanager.aux-services**</name>**
88. **<value>**mapreduce\_shuffle**</value>**
89. **</property>**
90. **</configuration>**

修改配置文件mapred-site.xml

vim /opt/hadoop-2.9.0/etc/hadoop/mapred-site.xml

1. **<configuration>**
2. **<property>**
3. **<name>**mapreduce.framework.name**</name>**
4. **<value>**yarn**</value>**
5. **</property>**
7. **<property>**
8. **<name>**mapred.job.tracker**</name>**
9. **<value>**master:9001**</value>**
10. **<description>**Host or IP and port of JobTracker.**</description>**
11. **</property>**
13. **<property>**
14. **<name>**mapreduce.map.memory.mb**</name>**
15. **<value>**2048**</value>**
16. **</property>**
18. **<property>**
19. **<name>**mapreduce.map.java.opts**</name>**
20. **<value>**-Xmx2048M**</value>**
21. **</property>**
23. **<property>**
24. **<name>**mapreduce.reduce.memory.mb**</name>**
25. **<value>**3584**</value>**
26. **</property>**
28. **<property>**
29. **<name>**mapreduce.reduce.java.opts**</name>**
30. **<value>**-Xmx2560M**</value>**
31. **</property>**
33. **<property>**
34. **<name>**mapreduce.task.io.sort.mb**</name>**
35. **<value>**512**</value>**
36. **</property>**
38. **<property>**
39. **<name>**mapreduce.task.io.sort.factor**</name>**
40. **<value>**100**</value>**
41. **</property>**
43. **<property>**
44. **<name>**mapreduce.reduce.shuffle.parallelcopies**</name>**
45. **<value>**50**</value>**
46. **</property>**
48. **<property>**
49. **<name>**mapreduce.jobhistory.address**</name>**
50. **<value>**rscmng:10020**</value>**
51. **</property>**
53. **<property>**
54. **<name>**mapreduce.jobhistory.webapp.address**</name>**
55. **<value>**rscmng:19888**</value>**
56. **</property>**
58. **<property>**
59. **<name>**mapreduce.jobhistory.intermediate-done-dir**</name>**
60. **<value>**/mr-history/tmp**</value>**
61. **</property>**
63. **<property>**
64. **<name>**mapreduce.jobhistory.done-dir**</name>**
65. **<value>**/mr-history/done**</value>**
66. **</property>**
67. **</configuration>**

修改配置文件slaves

vim /opt/hadoop-2.9.0/etc/hadoop/slaves

1. datanode1
2. datanode2

添加环境变量、用户、用户组等

vim /etc/profile.d/hadoop.sh

1. export HADOOP\_HOME=/opt/hadoop-2.9.0
2. export JAVA\_HOME=/usr/local/jdk1.8.0\_162
3. export JRE\_HOME=${JAVA\_HOME}/jre
4. export CLASSPATH=.:${JAVA\_HOME}/lib:${JRE\_HOME}/lib:$CLASSPATH
5. export PATH=${JAVA\_HOME}/bin:${JRE\_HOME}/bin:${HADOOP\_HOME}/bin:$PATH
7. export HADOOP\_PREFIX=/opt/hadoop-2.9.0
8. export HADOOP\_CONF\_DIR=/opt/hadoop-2.9.0/etc/hadoop

切换到root权限，创建hadoop用户组，创建hdfs,yarn用户

1. groupadd hadoop
2. useradd -d /opt/hdfs -g hadoop -s /bin/bash -m hdfs
3. useradd -d /opt/yarn -g hadoop -s /bin/bash -m yarn

创建目录，修改权限

1. mkdir /hadoop
2. mkdir /opt/hadoop-2.9.0/logs
3. mkdir /opt/hadoop
4. chmod -R 775 /hadoop/  /opt/hadoop/  /opt/hadoop-2.9.0/   /usr/local/jdk1.8.0\_162

将namenode虚拟机克隆为rscmng和datanode1，重新生成mac地址，修改主机名。配置namenode与datanode1在hdfs下的免密登录，配置rscmng与datanode1在yarn下的免密登陆：

1. ssh-keygen -t rsa -P '' -f ~/.ssh/id\_rsa
2. cat ~/.ssh/id\_rsa.pub **>>** ~/.ssh/authorized\_keys

在datanode1上拷贝namenode的公钥添加进authorized\_keys即可

检验免密登录是否成功

配置datanode1，创建/hadoop/hdfs及/hadoop/yarn目录，修改权限

1. mkdir /hadoop/hdfs
2. mkdir /hadoop/yarn
3. chown -R hdfs:hadoop /hadoop/hdfs/ /hadoop/yarn/
4. chmod -R 775 /hadoop/hdfs/ /hadoop/yarn/

将datanode1克隆为datanode2，分别修改/opt/hadoop-2.9.0/etc/hadoop/hdfs-site.xml文件，添加dfs.datanode.data.dir属性，指定各datanode块的存放位置

1. **<property>**
2. **<name>**dfs.datanode.data.dir**</name>**
3. **<value>**file:/hadoop/hdfs/dfs/data/datanode1**</value>**
4. **</property>**

修改各个结点的/etc/hosts文件，记录ip地址映射

vim /etc/hosts

1. 192.168.104.138  namenode
2. 192.168.104.144  rscmng
3. 192.168.104.142  datanode1
4. 192.168.104.143  datanode2
5. 启动

namenode切换到hdfs用户，格式化hdfs

1. $HADOOP\_PREFIX/bin/hdfs namenode -format mycluster

启动hdfs

1. $HADOOP\_PREFIX/sbin/start-dfs.sh

rscmng切换到yarn用户，启动yarn

1. $HADOOP\_PREFIX/sbin/start-yarn.sh

创建/mr-history/done并修改权限

1. mkdir /mr-history
2. mkdir /mr-history/done
3. chown -R hdfs:hadoop /mr-history/
4. chmod -R 775 /mr-history/

启动historyserver

1. $HADOOP\_PREFIX/sbin/mr-jobhistory-daemon.sh --config $HADOOP\_CONF\_DIR start historyserver

打开网页输入网址检验是否配置成功

<http://namenode:50070>

<http://rscmng:8088>

<http://rscmng:19888>