Java环境安装

1. 新建文件夹用于安装java环境
2. 上传压缩包

Rz java压缩包

1. 解压压缩包

Tar –zxvf java

1. 配置环境变量
2. Vim /etc/profile

Export JAVA\_HOME=/home/xqiang/java/jdk1.8.0

Export CLASS\_PATH=.:$JAVA\_HOME/lib/dt.jar;$JAVA\_HOME/lib/tools.jar

PATH=.:$PATH:$JAVA\_HOME/bin

1. source /etc/profile 使配置生效

Hadoop环境搭建

1. 新建文件夹用于安装haddop环境
2. 上传haddop压缩包,一般是2.7.4版本的hadoop压缩包

Rz 指令

1. 解压压缩包

Tar –zxvf Hadoop

1. 配置环境变量
2. Vim /etc/profile

Export HADDOP\_HOME=/home/xqiang/haddop/hadoop2.7-4

Export HADOOP\_COMMAND\_NATIVE\_CONF=$HADOOP\_HOME/lib/native

Export HADOOP\_CONF\_DIR=$HADOOP\_HOME/etc/hadoop

Export .:$PATH:$HADOOP\_HOME/bin:$HADOOP\_HOME/sbin

1. Source /etc/profile 是文件生效
2. 修改主机的名字
3. Hostname server01 或者

Vim /etc/hostname 进行修改

1. 修改host文件
2. 在文件的末尾加上ip地址

192.168.117.128 hostname ==与hostname对应

五、创建账号

1) 创建hadoop账号，设置密码

Useradd Hadoop

Passwd Hadoop ==>设置密码

1. 授予hadoop roo用户权限

编辑 /etc/sudoers文件，添加权限

Hadoop ALL=(ALL) ALL

1. 更改文件的属主

Chown –R Hadoop:Hadoop /home/Hadoop/hadoop2.7-4

1. 授予权限

Chmod 777 \*

六、创建SSH无密码登录

1) 生成秘钥

Ssh-keygen –r rsa ==>默认的位置为/root/.ssh/

1. 将生成的公钥copy到的授权的key中

cat ~/.ssh/id\_rsa.pub >> ~/.ssh/authorized\_keys

将其他的从机的公钥也copy到key文件中

1. 测试连接其他的主机

七、创建一些文件，用于存储分布式数据

 mkdir /root/hadoop

 mkdir /root/hadoop/tmp

 mkdir /root/hadoop/var

 mkdir /root/hadoop/dfs

 mkdir /root/hadoop/dfs/name

 mkdir /root/hadoop/dfs/data

1、配置core.site.xml文件

1. <property>
2. <name>hadoop.tmp.dir</name> -->临时文件的目录
3. <value>/root/hadoop/tmp</value>
4. <description>Abase for other temporary directories.</description>
5. </property>
6. <property>
7. <name>fs.default.name</name> -->第三方连接时的地址
8. <value>hdfs://hserver1:9000</value>
9. </property>
10. <property>

<name>io.file.buffer.size</name> -->io文件大小  
<value>131072</value>  
</property>

1. </configuration>

2、修改hadoop-env.sh文件，将其中的JAVA\_HOME 路径进行修改

 将export   JAVA\_HOME=${JAVA\_HOME}

 修改为：

 export   JAVA\_HOME=/opt/java/jdk1.8.0\_121

3、配置hdfs.site.xml文件

<property>

   <name>dfs.name.dir</name> --->namenode 节点

   <value>/root/hadoop/dfs/name</value>

   <description>Path on the local filesystem where the NameNode stores the namespace and transactions logs persistently.</description>

</property>

<property>

   <name>dfs.data.dir</name> -->data node 节点

   <value>/root/hadoop/dfs/data</value>

   <description>Comma separated list of paths on the local filesystem of a DataNode where it should store its blocks.</description>

</property>

<property>

   <name>dfs.replication</name> -->文件备份的块数

   <value>1</value>

</property>

<property>

      <name>dfs.permissions</name>

      <value>false</value>

      <description>need not permissions</description>

</property>

==========更改默认的配置=========

<property>

<name>dfs.datanode.fsdataset.volume.choosing.policy</name>

<value>org.apache.hadoop.hdfs.server.datanode.fsdataset.AvailableSpaceVolumeChoosingPolicy</value>

</property>

<property>

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

<value>127.0.0.1:8305</value>

</property>

<property>

<name>dfs.namenode.secondary.http-address</name>

<value>127.0.0.1:8310</value>

</property>

4、配置mapred-site.xml文件

 <property>

   <name>mapred.job.tracker</name> --> 作业跟踪地址

   <value>hserver1:49001</value>

</property>

<property> -->map-reduce本地目录

      <name>mapred.local.dir</name>

       <value>/root/hadoop/var</value>

</property>

<property> -->mapreduce 框架选用

       <name>mapreduce.framework.name</name>

       <value>yarn</value>

</property>

<property>

<name>yarn.app.mapreduce.am.staging-dir</name>

<value>/home/work/data/hadoop/tmp/hadoop-yarn/staging</value>

</property>

<property>

<name>mapreduce.jobhistory.address</name>

<value>127.0.0.1:8330</value>

</property>

<property>

<name>mapreduce.jobhistory.webapp.address</name>

<value>127.0.0.1:8331</value>

</property>

<property>

<name>mapreduce.jobhistory.done-dir</name>

<value>${yarn.app.mapreduce.am.staging-dir}/history/done</value>

</property>

<property>

<name>mapreduce.jobhistory.intermediate-done-dir</name>

<value>${yarn.app.mapreduce.am.staging-dir}/history/done\_intermediate</value>

</property>

<property>

<name>mapreduce.jobhistory.joblist.cache.size</name>

<value>1000</value>

</property>

<property>

<name>mapreduce.tasktracker.map.tasks.maximum</name>

<value>8</value>

</property>

<property>

<name>mapreduce.tasktracker.reduce.tasks.maximum</name>

<value>8</value>

</property>

<property>

<name>mapreduce.jobtracker.maxtasks.perjob</name>

<value>5</value>

<description>The maximum number of tasks for a single job.

A value of -1 indicates that there is no maximum.

</description>

</property>

5、配置yarn.xml文件

<configuration>

<property>

<name>yarn.nodemanager.aux-services</name>

<value>mapreduce\_shuffle</value>

</property>

<property>

<name>yarn.resourcemanager.hostname</name>

<value>127.0.0.1</value>

</property>

<property>

<name>yarn.resourcemanager.webapp.address</name>

<value>127.0.0.1:8320</value>

</property>

<property>

<name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>

<value>org.apache.hadoop.mapred.ShuffleHandler</value>

</property>

<property>

<name>yarn.log-aggregation-enable</name>

<value>true</value>

</property>

<property>

<name>yarn.log-aggregation.retain-seconds</name>

<value>864000</value>

</property>

<property>

<name>yarn.log-aggregation.retain-check-interval-seconds</name>

<value>86400</value>

</property>

<property>

<name>yarn.nodemanager.remote-app-log-dir</name>

<value>/YarnApp/Logs</value>

</property>

<property>

<name>yarn.log.server.url</name>

<value>http://127.0.0.1:8325/jobhistory/logs/</value>

</property>

<property>

<name>yarn.nodemanager.local-dirs</name>

<value>/home/work/data/hadoop/dfs/tmp/</value>

</property>

<property>

<name>yarn.scheduler.maximum-allocation-mb</name>

<value>5000</value>

</property>

<property>

<name>yarn.scheduler.minimum-allocation-mb</name>

<value>1024</value>

</property>

<property>

<name>yarn.nodemanager.vmem-pmem-ratio</name>

<value>4.1</value>

</property>

<property>

<name>yarn.nodemanager.vmem-check-enabled</name>

<value>false</value>

</property>

<name>yarn.resourcemanager.address</name>

   <value>master:18040</value>

 </property>

 <property>

 <name>yarn.resourcemanager.scheduler.address</name>

   <value>master:18030</value>

 </property>

 <property>

   <name>yarn.resourcemanager.webapp.address</name>

   <value>master:18088</value>

 </property>

 <property>

   <name>yarn.resourcemanager.resource-tracker.address</name>

   <value>master:18025</value>

 </property>

 <property>

   <name>yarn.resourcemanager.admin.address</name>

   <value>master:18141</value>

 </property>

</configuration>

6、执行初始化脚本

./Hadoop namenode –format

八、启动文件系统

1) ./start-dfs.sh

2) ./start-yarn.sh

九、haddop指令操作

连接地址 <http://www.cnblogs.com/LHWorldBlog/p/8514994.html>

<https://blog.csdn.net/m0_38003171/article/details/79086780>

1) 创建文件

Hdfs dfs –mkdir /user

2)删除文件夹

Hdfs dfs –rmr /file

3)查看指定目录下的内容

Hdfs dfs –ls [文件目录]

4)打开某个已存在的文件

hdfs dfs –cat [file\_path]

5)将本地文件存储至hadoop目录

hdfs dfs –put [本地地址] [hadoop目录]

6) 将本地目录存储至hadoop目录

hdfs dfs –put [本地目录] [hadoop目录]

7) 删除指定目录的文件

**hdfs dfs –rm /user/t/ok.txt**

**8) 将hadoop上某个文件down至本地目录下**

hadoop dfs -get [文件目录] [本地目录]  
  **hadoop dfs –get /user/t/ok.txt /home/t**

9) 在hadoop指定目录下新建一个空文件

**hdfs dfs  -touchz  /user/new.txt**

**10)在hadoop上某个文件重命名**

**hdfs dfs –mv  /user/test.txt  /user/ok.txt**  （将test.txt重命名为ok.txt）

11)将hadoop指定目录下所有保存为一个文件，并down至本地

**hdfs dfs –getmerge /user /home/t**

12)将正在运行的hadoop作业kill

**hadoop job –kill  [job-id]**

13)安全模式

退出安全模式

**hadoop dfsadmin -safemode leave**

**进入安全模式**

**hadoop dfsadmin -safemode enter**

**负载均衡**

**bin/start-balancer.sh**