1 hadoop 单机配置

/usr/local/hadoop/etc/hadoop/hadoop-env.sh

export JAVA\_HOME="/usr/lib/jvm/java-1.8.0-openjdk-1.8.0.131-11.b12.el7.x86\_64/jre"

export HADOOP\_CONF\_DIR="/usr/local/hadoop/etc/hadoop"

分析单词出现的次数

./bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.6.jar wordcount oo xx

#-----------------------------------------------------#

ALL 表示所有主机，

NODE 表示 node1,node2,node3

NN1: 表示 namenode

#-----------------------------------------------------#

完全分布式集群搭建 -- HDFS

192.168.1.10 nn01 namenode,secondarynamenode

192.168.1.11 node1 datanode

192.168.1.12 node2 datanode

192.168.1.13 node3 datanode

ALL: 配置 /etc/hosts

ALL: 安装 java-1.8.0-openjdk-devel

core-site.xml 配置

<configuration>

<property>

<name>fs.defaultFS</name>

<value>hdfs://nn01:9000</value>

</property>

<property>

<name>hadoop.tmp.dir</name>

<value>/var/hadoop</value>

</property>

</configuration>

ALL: 创建 /var/hadoop

hdfs-site.xml 配置

<configuration>

<property>

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

<value>nn01:50070</value>

</property>

<property>

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

<value>nn01:50090</value>

</property>

<property>

<name>dfs.replication</name>

<value>2</value>

</property>

</configuration>

ALL: 同步配置到所有主机

NN01: 格式化 namenode

./bin/hdfs namenode -format

NN01: 启动集群

./sbin/start-dfs.sh

停止集群可以使用 ./sbin/stop-dfs.sh

ALL: 验证角色 jps

NN01: 验证集群是否组建成功

./bin/hdfs dfsadmin -report

服务启动日志路径 /usr/local/hadoop/logs

mapred-site.xml 配置

<configuration>

<property>

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

<value>yarn</value>

</property>

</configuration>

yarn-site.xml 配置

<configuration>

<!-- Site specific YARN configuration properties -->

<property>

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

<value>nn01</value>

</property>

<property>

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

<value>mapreduce\_shuffle</value>

</property>

</configuration>

ALL: 同步配置到主机

NN1: 启动服务 ./sbin/start-yarn.sh

ALL: 验证角色 jps

NN1: 验证节点状态 ./bin/yarn node -list

增加修复节点

按照单机方法安装一台机器，部署运行的 java 环境

拷贝 namenode 的文件到本机

启动 datanode

./sbin/hadoop-daemons.sh start datanode

设置同步带宽

./bin/hdfs dfsadmin -setBalancerBandwidth 60000000

./sbin/start-balancer.sh

删除节点

<property>

<name>dfs.hosts.exclude</name>

<value>/usr/local/hadoop/etc/hadoop/exclude</value>

</property>

开始导出数据

./bin/hdfs dfsadmin -refreshNodes

查看状态

Normal 正常状态

Decommissioned in Program 数据正在迁移

Decommissioned 数据迁移完成

yarn 增加 nodemanager

./sbin/yarn-daemon.sh start nodemanager

yarn 停止 nodemanager

./sbin/yarn-daemon.sh stop nodemanager

yarn 查看节点状态

./bin/yarn node -list

NFS 网关

1 配置 /etc/hosts (NFSGW)

192.168.1.10 nn01

192.168.1.11 node1

192.168.1.12 node2

192.168.1.13 node3

192.168.1.15 nfsgw

2 添加用户(nfsgw, nn01)

groupadd -g 200 nsd1803

useradd -u 200 -g 200 nsd1803

NN01: 3 停止集群

./sbin/stop-all.sh

NN01: 4 增加配置 core-site.xml

<property>

<name>hadoop.proxyuser.nsd1803.groups</name>

<value>\*</value>

</property>

<property>

<name>hadoop.proxyuser.nsd1803.hosts</name>

<value>\*</value>

</property>

NN01: 5 同步配置到 node1 node2 node3

NN01: 6 启动集群 ./sbin/start-dfs.sh

NN01: 7 查看状态

./bin/hdfs dfsadmin -report

NFSGW: 安装 java-1.8.0-openjdk-devel

NFSGW: 同步 nn01 的 /usr/local/hadoop 到NFSGW的相同目录下

NFSGW: hdfs-site.xml 增加配置

<property>

<name>nfs.exports.allowed.hosts</name>

<value>\* rw</value>

</property>

<property>

<name>nfs.dump.dir</name>

<value>/var/nfstmp</value>

</property>

NFSGW: 创建转储目录，并给用户 nsd1803 赋权

mkdir /var/nfstmp

chown nsd1803:nsd1803 /var/nfstmp

NFSGW: 给 /usr/local/hadoop/logs 赋权

setfacl -m u:nsd1803:rwx

创建数据根目录 /var/hadoop

mkdir /var/hadoop