成员信息

- 181250072 李俊杰
- 181250123 师域飞
- 181250127 苏语风
- 181250187 张许妍

1. 集群设计：每台机器的硬件情况和网络配置信息

三台centos7

采用虚拟机：暂定每个系统双核，1GB运行内存（可以调）

network config:

三台虚拟机网络配置信息

```
192.168.21.3 master
192.168.21.4 slave1
192.168.21.5 slave2
```

配置文件信息示例

指令: vim /etc/sysconfig/network-scripts/ifcfg-ens33

```
IPADDR=192.168.21.4  #设置IP地址
NETMASK=255.255.225.0  #设置子网掩码
GATEWAY=192.168.21.2  #设置网关
#DNS1=61.147.  #设置主DNS
BOOTPROTO=static  #启用静态IP地址 ,默认为dhcp
```

2. 集群搭建操作记录

安装java环境

shell脚本删除open-jdk

```
#!/bin/bash
#################################################################
#
#  function: batch uninstall rpm packages
#  setup:
#         1. copy the scripts and save as a file, such as: ex.sh
#         2. switch to root user. su - root
#         3. change the file's permission: chmod +x ex.sh
#         3. running the script with no parameter: ./ex.sh
#  runing:
#        uninstall [rpm package name]
#  author: Topurce Zhou (topurce#at#hotmail.com)
#
#################################################################
if [ "$UID" -ne 0 ]
then
    echo -e 'must be \E[34m\033[1mroot\033[0m to run this script.'
    echo -ne '\E[0m'
    exit 67
fi

if [ ! -f /usr/bin/uninstall ]
then
   echo "building file..."
   scripts="$(cat $0)"
   declare -i index=1
   cat $0 | while read line
   do
       if (( index == 19 ))
       then
          echo 'echo -e "must be \E[34m\033[1mroot\033[0m to run this
script."'>>/usr/bin/uninstall
          echo 'echo -ne "\E[0m"'>>/usr/bin/uninstall
       elif (( index == 23 ))
       then
          echo 'stips="searching packages for \"$1\":"'>>/usr/bin/uninstall
      echo 'usage="usage: $0 \"package name\""'>>/usr/bin/uninstall
       elif (( index != 19 && index != 20 && (index<23 || index>52) ))
       then
          echo $line>>/usr/bin/uninstall
       fi
       index+=1;
   done
   chmod +x /usr/bin/uninstall
   echo "try \"uninstall [package name]\" again."
   exit
fi

stips="searching packages for \"$1\":"
usage="usage: $0 \"rpm package name\""

if [ $# -eq 0 ]
then
    echo "$0: no rpm packages given for uninstall."
    echo $usage
```

```
elif [ $# -gt 1 ]
then
    echo $usage
else
    echo $stips
    rpms="$(rpm -qa | grep $1)"
    declare -i count=0
    for rpmk in $rpms
    do
       count+=1
       echo "package: $rpmk"
    done
    if (( count == 0 ))
    then
       echo "no packages"
       exit
    fi
    echo "packages: $count"
    echo
    read -p "are you sure you want to uninstall all above packages?(y/n)"
    if [[ $REPLY == [Yy] ]]
    then
         echo "starting to uninstall packages..."
         for rpmk in $rpms
     do
         count+=1
         echo "uninstalling package: $rpmk"
             rpm -e --nodeps $rpmk
             if [ $? -eq 0 ]
             then
             echo "done"
             else
                 echo "faild to uninstall $rpmk"
             fi
      done
    fi
fi
```

正常下载解压java

配置环境变量:

vim /etc/profile

source /etc/profile

在开头加入

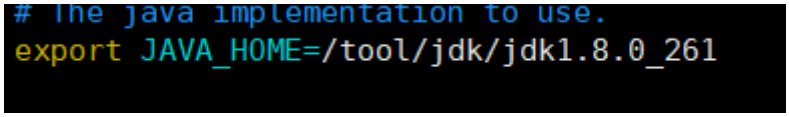
```
export JAVA_HOME=/tool/jdk/jdk1.8.0_261
export JRE_HOME=${JAVA_HOME}/jre
export CLASSPATH=.:${JAVA_HOME}/lib:${JRE_HOME}/lib
export  PATH=${JAVA_HOME}/bin:$PATH
```

scala 配置,配置环境变量即可（同java）

```
export PATH=/tool/scala/scala-2.11.12/bin:$PATH
```

hadoop配置

通过etc/hadoop/hadoop env.sh 配置运行环境参数，至少需要配置一下



配置两个configuration文件

（如果要配置yarn的话，命令在下面）

vim /tool/hadoop/hadoop-2.7.7/etc/hadoop/core-site.xml

vim /tool/hadoop/hadoop-2.7.7/etc/hadoop/hdfs-site.xml

vim /tool/hadoop/hadoop-2.7.7/etc/hadoop/yarn-site.xml

vim /tool/hadoop/hadoop-2.7.7/etc/hadoop/mapred-site.xml

自己提前建好文件夹

```
<configuration>

<!--hdfs临时路径-->

<property>

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

  <value>/data/hdfs</value>

</property>

<!--hdfs 的默认地址、端口 访问地址-->

<property>

  <name>fs.defaultFS</name>

  <value>hdfs://master:9000</value>

</property>

</configuration>



<configuration>

<!--hdfs web的地址 -->

<property>

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

  <value>master:50070</value>
```

```
</property>
<!-- 副本数-->
<property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>
<!-- 是否启用hdfs权限检查 false 关闭 -->
    <property>
    <name>dfs.permissions.enabled</name>
    <value>false</value>
  </property>
<!-- 块大小,默认字节, 可使用 k m g t p e-->
    <property>
    <name>dfs.blocksize</name>
  <!--128m-->
     <value>134217728</value>
  </property>
</configuration>
```

初始化HDFS

$HADOOP_HOME/bin/hdfs namenode -format

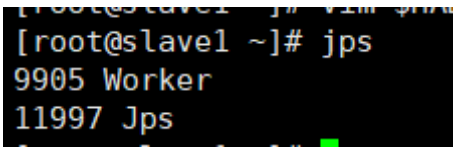
启动

在所有NameNode 上启动：

$HADOOP_HOME/bin/hdfs namenode

在所有DataNode 上启动：

$HADOOP_HOME/bin/hdfs datanode



```
export HADOOP_HOME=/tool/hadoop/hadoop-2.7.7
export PATH=$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$PATH
```

spark 配置

master上：/tool/spark2/spark-2.4.7-bin-hadoop2.7/sbin/start-master.sh

（老版的貌似要配置spark slave配置文件（所有机器都配）

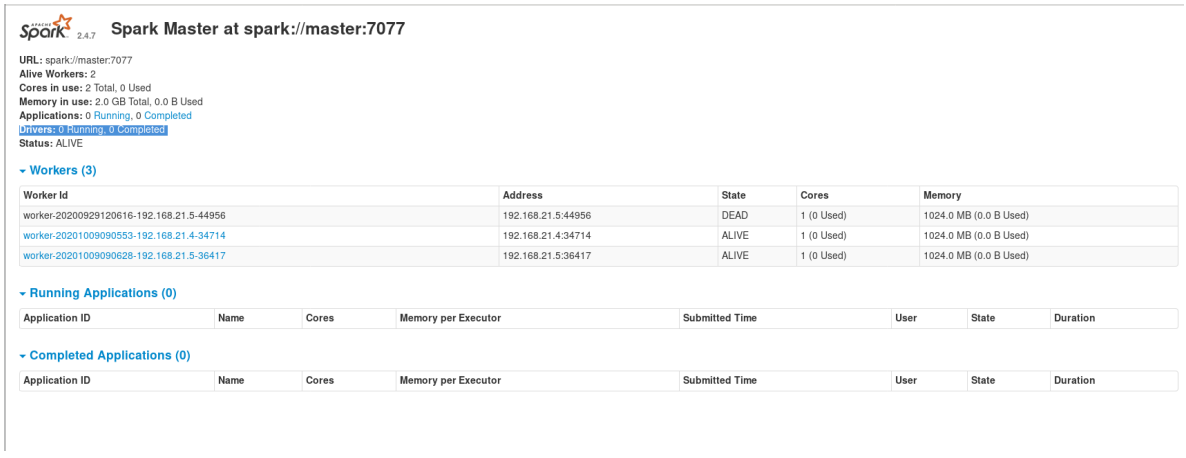
配置/tool/spark2/spark-2.4.7-bin-hadoop2.7/conf/slaves）

slave上：/tool/spark2/spark-2.4.7-bin-hadoop2.7/sbin/start-slave.sh 192.168.21.3:7077

```
export HADOOP_HOME=/tool/hadoop/hadoop-2.7.7
export PATH=$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$PATH
export SPARK_HOME=/tool/spark2/spark-2.4.7-bin-hadoop2.7
export PATH=$PATH:$SPARK_HOME/bin
export PATH=/tool/scala/scala-2.11.12/bin:$PATH
```

！注意关闭防火墙

systemctl stop firewalld.service
systemctl status firewalld.service



Spark Master at spark://master:7077

URL: spark://master:7077
Alive Workers: 2
Cores in use: 2 Total, 0 Used
Memory in use: 2.0 GB Total, 0.0 B Used
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (3)

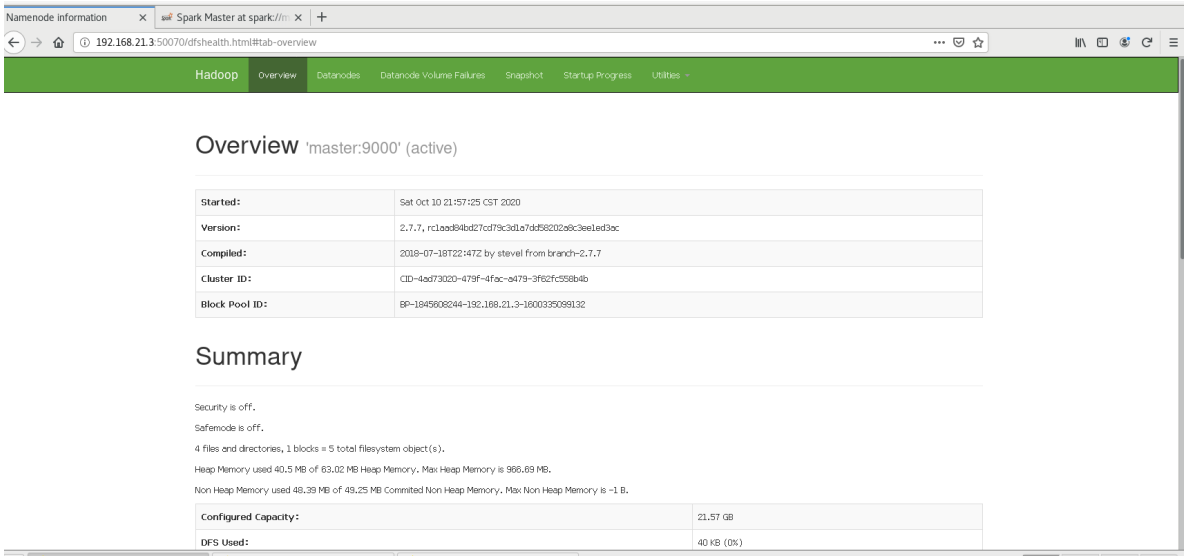
Worker Id	Address	State	Cores	Memory
worker-20200929120616-192.168.21.5-44956	192.168.21.5:44956	DEAD	1 (0 Used)	1024.0 MB (0.0 B Used)
worker-20201009090553-192.168.21.4-34714	192.168.21.4:34714	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)
worker-20201009090628-192.168.21.5-36417	192.168.21.5:36417	ALIVE	1 (0 Used)	1024.0 MB (0.0 B Used)

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration

Completed Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration


Namenode information
Spark Master at spark://m
192.168.21.3:50070/dfshealth.html#tab-overview
Hadoop
Overview
Datanodes
Datanode Volume Failures
Snapshot
Startup Progress
Utilities
Overview 'master:9000' (active)
Started: Sat Oct 10 21:57:25 CST 2020
Version: 2.7.7, rc1aad84bd27cd79c3d1a7dd58202a8c3ee1ed3ac
Compiled: 2018-07-18T22:47Z by stevel from branch-2.7.7
Cluster ID: CID-4ad73020-479f-4fac-a479-3f62fc558b4b
Block Pool ID: BP-1845608244-192.168.21.3-1600335099132
Summary
Security is off.
Safemode is off.
4 files and directories, 1 blocks = 5 total filesystem object(s).
Heap Memory used 40.5 MB of 63.02 MB Heap Memory. Max Heap Memory is 966.69 MB.
Non Heap Memory used 48.39 MB of 49.25 MB Commited Non Heap Memory. Max Non Heap Memory is -1 B.
Configured Capacity: 21.57 GB
DFS Used: 40 KB (0%)