

优于别人，并不高贵，真正的高贵应该是优于过去的自己

Hadoop完全分布式环境搭建

编写分发文件脚本

应用场景如下：比如有三台主机master1,slave1,slave2

如果简历完全分布式的集群就需要将文件从master1拷贝到slave从机上

那么可以使用rsync命令分发单个文件，也可以使用如下脚本分发文件夹或者文件

```
#!/bin/bash

#1 获取输入参数个数，如果没有参数，直接退出
# $#代表获得命令行参数个数
pcount=$#
if((pcount==0)); then
echo no args;
exit;
fi

#2 获取文件名称
# $1代表获得命令行第一个参数
p1=$1
fname=`basename $p1`
echo fname=$fname

#3 获取上级目录到绝对路径
pdir=`cd -P $(dirname $p1); pwd`
echo pdir=$pdir

#4 获取当前用户名称
user=`whoami`

#5 rsync命令可以分发文件到指定主机
# 这里slave就是从机的机名
for((host=1; host<3; host++))
do
    echo ----- slave$host -----
    rsync -rvl $pdir/$fname $user@slave$host:$pdir
done
```

集群规划

	hadoop102	hadoop103	hadoop104
HDFS	NameNodeDataNode	DataNode	SecondaryNameNode DataNode
YARN	NodeManager	ResourceManager NodeManager	NodeManager

配置集群

配置文件都在hadoop2.7.2/etc/hadoop下

配置core-site.xml

```
<!-- 指定HDFS中NameNode的地址 -->
<property>
    <name>fs.defaultFS</name>
    <value>hdfs://hadoop102:9000</value>
</property>

<!-- 指定Hadoop运行时产生文件的存储目录 -->
<property>
    <name>hadoop.tmp.dir</name>
    <value>/opt/module/hadoop-2.7.2/data/tmp</value>
</property>
```

配置hadoop-env.sh

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

配置hdfs-site.xml

```
<!-- 指定副本数量 -->
<property>
    <name>dfs.replication</name>
    <value>3</value>
</property>

<!-- 指定SecondaryNamenode主机配置 -->
<property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>hadoop104:50090</value>
</property>
```

配置yarn-env.sh

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

配置yarn-site.xml

```
<!-- reducer获取数据的方式 -->
<property>
    <name>yarn.nodemanager.aux-services</name>
    <value>mapreduce_shuffle</value>
</property>

<!-- 指定YARN的ResourceManager的地址 -->
<property>
    <name>yarn.resourcemanager.hostname</name>
    <value>hadoop103</value>
</property>
```

配置mapred-env.sh

```
export JAVA_HOME=/opt/module/jdk1.8.0_144
```

配置mapred-site.xml

【需要拷贝mapred-site.xml.template 然后重命名即可】

```
<!-- 指定mr运行在yarn上 -->
<property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
</property>
```

分发配置到集群中

```
# xsync就是刚刚编写的分发脚本
xsync /opt/module/hadoop-2.7.2/
```

集群格式化

```
# 这部分必须没有报错才行，不然就重来一篇，必须在hadoop根目录下执行
hadoop namenode -format
```

集群启动测试

在主机hadoop102在sbin目录下，运行start-dfs.sh启动HDFS
在主机hadoop103在sbin目录下，运行start-yarn.sh启动yarn
然后使用jps查看进程即可

```
[zhutian@hadoop102 sbin]$ j
----- hadoop102 -----
3409 JobHistoryServer
3505 Jps
3018 NameNode
3164 DataNode
----- hadoop103 -----
2904 DataNode
3020 Jps
----- hadoop104 -----
3104 Jps
2905 DataNode
3020 SecondaryNameNode
```