

一、平台环境

阿里云服务器两台（时间原因，第一次用服务器，故只用了两台）

操作系统：Ubuntu16.04 64 位

master 节点：公网 ip 39.87.207.188 私网 ip 172.26.145.164

worker1 节点：公网 ip 118.31.16.139 私网 ip 172.26.55.72

二、软件包

Jdk-8u202-linux-x64.tar(java version 1.8.0_202)

hadoop-2.7.7.tar

spark-3.0.1-bin-hadoop2.7

三、搭建步骤

1. 创建用户

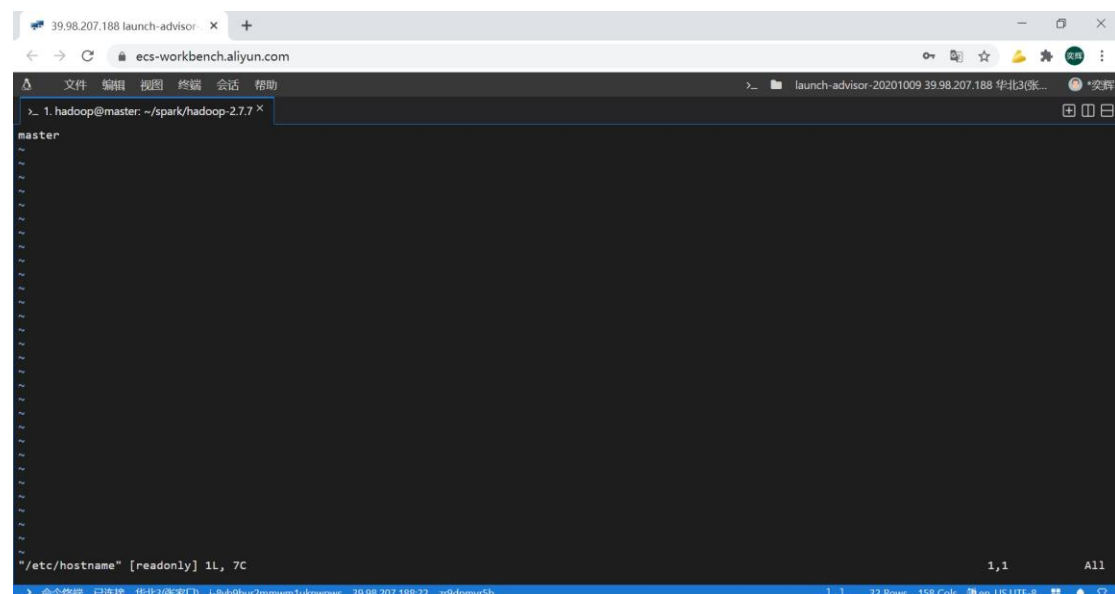
为方便后续操作在两个 linux 系统中均创建新用户 hadoop

2. 修改计算机环境

1) 集群环境：1 台 master, 1 台 worker

2) 修改主机名 hostname，将主机名改为 master 和 worker1（方便后续继续添加）

master 主机上：



worker 主机上：

The screenshot shows a web browser window with the address bar displaying 'ecs-workbench.aliyun.com'. The browser tabs include '39.98.207.188 launch-advisor' and 'ecs-workbench.aliyun.com'. The browser's address bar also shows 'launch-advisor-20201009 39.98.207.188 华北3(张...'. The browser's menu bar includes '文件', '编辑', '视图', '终端', '会话', and '帮助'. The browser's status bar shows '命令终端 已连接 华北3(张家口) i-8v69bur2mmwm1uknwpws 39.98.207.188.22 z9dpmyr5b 1,1 32 Rows 158 Cols en_US.UTF-8 *突辉'. The terminal window shows a prompt '1. hadoop@worker1: ~' and a file listing of '/etc/hostname' [readonly] 1L, 8C. The file listing shows '1,1 All'.

3. 修改自身的 IP 地址, 可通过 ifconfig 查看修改后的 IP 地址

由于阿里云服务器计费的方法问题, 公网 ip 不能直接作为 hosts 来使用, 首先开放 7000-9000 端口和 50070 端口 (懒得一个个加了, 多开一点)

然后修改 hosts 本地文件:

master:

The screenshot shows a web browser window with the address bar displaying 'ecs-workbench.aliyun.com'. The browser tabs include '39.98.207.188 launch-advisor' and 'ecs-workbench.aliyun.com'. The browser's address bar also shows 'launch-advisor-20201009 39.98.207.188 华北3(张...'. The browser's menu bar includes '文件', '编辑', '视图', '终端', '会话', and '帮助'. The browser's status bar shows '命令终端 已连接 华北3(张家口) i-8v69bur2mmwm1uknwpws 39.98.207.188.22 z9dpmyr5b 9,1 32 Rows 158 Cols en_US.UTF-8 *突辉'. The terminal window shows a prompt '1. hadoop@master: ~/spark/hadoop-2.7.7' and the contents of '/etc/hosts' [readonly] 11L, 236C. The contents of '/etc/hosts' are as follows:

```
127.0.0.1 localhost
39.98.207.188 localhost
172.26.145.164 master
118.31.16.139 worker1
# The following lines are desirable for IPv6 capable hosts
::1 localhost ip6-localhost ip6-loopback
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters
```

worker1:


```

spark@i2bp17fphet18llxcwy6pZ:/root$ ping master
PING master (39.98.207.188) 56(84) bytes of data:
64 bytes from master (39.98.207.188): icmp_seq=1 ttl=54 time=32.9 ms
64 bytes from master (39.98.207.188): icmp_seq=2 ttl=54 time=32.9 ms
64 bytes from master (39.98.207.188): icmp_seq=3 ttl=54 time=32.9 ms
^C
--- master ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2001ms
rtt min/avg/max/mdev = 32.933/32.948/32.958/0.010 ms
spark@i2bp17fphet18llxcwy6pZ:/root$ ping worker1
PING worker1 (118.31.16.139) 56(84) bytes of data:
64 bytes from worker1 (118.31.16.139): icmp_seq=1 ttl=64 time=1.33 ms
64 bytes from worker1 (118.31.16.139): icmp_seq=2 ttl=64 time=1.31 ms
^C
--- worker1 ping statistics ---
2 packets transmitted, 2 received, 0% packet loss, time 1000ms
rtt min/avg/max/mdev = 1.315/1.324/1.334/0.037 ms
spark@i2bp17fphet18llxcwy6pZ:/root$ sudo apt-get install ssh
正在读取软件包列表... 完成
正在分析软件包的依赖关系树
正在读取状态信息... 完成
ssh 已经是最新版 (1:7.2p2-4ubuntu2.10)。
升级了 0 个软件包，新安装了 0 个软件包，要卸载 0 个软件包，有 3 个软件包未被升级。
spark@i2bp17fphet18llxcwy6pZ:/root$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/spark/.ssh/id_rsa):
Created directory /home/spark/.ssh/.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/spark/.ssh/id_rsa.
Your public key has been saved in /home/spark/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:bc108KT861e1l1hdSAZNN1ZKJcc2NFDOssAacwq49Es spark@i2bp17fphet18llxcwy6pZ
The key's randomart image is:
+---[RSA 2048]----+
|o+*.ooo          |
|+*B.+o          |
|"Bo"..          |

```

(3) 将 worker1 的 id_rsa.pub 发送给 master

```

> 3.spark@worker1: ~$
root@worker1:~# su spark
spark@worker1:/root$ cd ~
spark@worker1:~$ scp /home/spark/.ssh/id_rsa.pub spark@master:~/.ssh/id_rsa.pub.worker1
The authenticity of host 'master (39.98.207.188)' can't be established.
ECDSA key fingerprint is SHA256:gctP1llgUS20x08f1wq0j5zMgwoiKixjJ/liziyY.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'master,39.98.207.188' (ECDSA) to the list of known hosts.
spark@master:~$
id_rsa.pub
spark@worker1:~$

```

(4) 在 master 上，将公钥加载到用于认证的公钥文件 authorized_key 中

```

39.98.207.188 launch-advisor - x
ecs-workbench.aliyun.com
文件 编辑 视图 终端 会话 帮助
> 1.spark@master: ~/.ssh$
2020-10-09 14:35:10 重新连接成功!
Welcome to Ubuntu 20.04.1 LTS (GNU/Linux 5.4.0-31-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

8 updates can be installed immediately.
8 of these updates are security updates.
To see these additional updates run: apt list --upgradable

Welcome to Alibaba Cloud Elastic Compute Service !

Last login: Fri Oct 9 14:03:53 2020 from 47.96.60.215
root@master:~# su spark
spark@master:/root$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
ys
spark@master:/root$ cd .ssh
bash: cd: .ssh: Permission denied
spark@master:/root$ cd .ssh
bash: cd: .ssh: Permission denied
spark@master:/root$ cd ~/.ssh
spark@master:~/.ssh$ ls
authorized_keys  id_rsa  id_rsa.pub  id_rsa.pub.worker1
spark@master:~/.ssh$

```

(5) 将公钥文件 authorized_key 分发给 woker1 并验证免密登录

```
39.98.207.188 launch-advisor x +
ecs-workbench.aliyun.com
文件 编辑 视图 终端 会话 帮助
launch-advisor-20201009 39.98.207.188 华北3(张...) *突辉

1. spark@master: ~/ssh ^
118.31.243.0/24
spark@master:~/ssh$ scp /home/spark/.ssh/authorized_keys spark@wo
rker1:~/ssh/
The authenticity of host 'worker1 (118.31.16.139)' can't be establ
ished.
ECDSA key fingerprint is SHA256:WkuILr/mOWjU5fcXrL01/oAEW+t2gH/oZS
XMLzR5ecE.
Are you sure you want to continue connecting (yes/no/[fingerprint]
)? yes
Warning: Permanently added 'worker1,118.31.16.139' (ECDSA) to the
list of known hosts.
spark@worker1's password:
authorized_keys 100% 994 31.9KB/s 00:00
spark@master:~/ssh$ ssh worker1
Welcome to Ubuntu 16.04.7 LTS (GNU/Linux 4.4.0-187-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

8 个可升级软件包。
8 个安全更新。

New release '16.04.5 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Welcome to Alibaba Cloud Elastic Compute Service !

spark@worker1:~$ quit
No command 'quit' found, did you mean:
Command 'qgit' from package 'qgit' (universe)
```

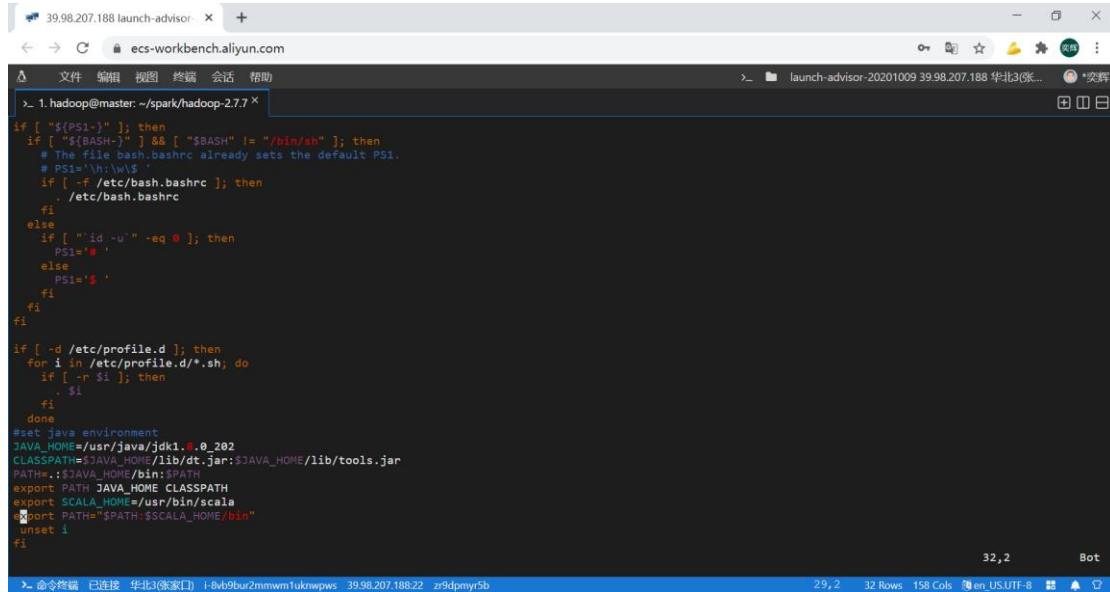
6. 安装 scala 和 java (master 和 worker 均要)

- (1) 用 apt install scala 安装 scala
- (2) 查看 scala 位置

```
39.98.207.188 launch-advisor x +
ecs-workbench.aliyun.com
文件 编辑 视图 终端 会话 帮助
launch-advisor-20201009 39.98.207.188 华北3(张...) *突辉

1. hadoop@master: ~/spark/hadoop-2.7.7 ^
hadoop@master:~/spark/hadoop-2.7.7$ which scala
/usr/bin/scala
hadoop@master:~/spark/hadoop-2.7.7$
```

- (3) 在/usr 文件夹下新建 java 文件夹，将jdk 压缩包移入并解压
- (4) 修改/etc/profile 文件，配置 java 和 scala 环境



```
1. hadoop@master: ~/spark/hadoop-2.7.7
if [ "${PS1:-}" ]; then
if [ "${BASH-}" ] && [ "$BASH" != "/bin/sh" ]; then
# The file bash.bashrc already sets the default PS1.
# PS1='\h:\w\$ '
if [ -f /etc/bash.bashrc ]; then
. /etc/bash.bashrc
fi
else
if [ "$(id -u)" -eq 0 ]; then
PS1='# '
else
PS1='$ '
fi
fi
fi

if [ -d /etc/profile.d ]; then
for i in /etc/profile.d/*.sh; do
if [ -r $i ]; then
. $i
done
done

#set java environment
JAVA_HOME=/usr/java/jdk1.8.0_202
CLASSPATH=$JAVA_HOME/lib/rt.jar:$JAVA_HOME/lib/tools.jar
PATH=$JAVA_HOME/bin:$PATH
export PATH JAVA_HOME CLASSPATH
export SCALA_HOME=/usr/bin/scala
export PATH="$PATH:$SCALA_HOME/bin"
unset i
fi
```

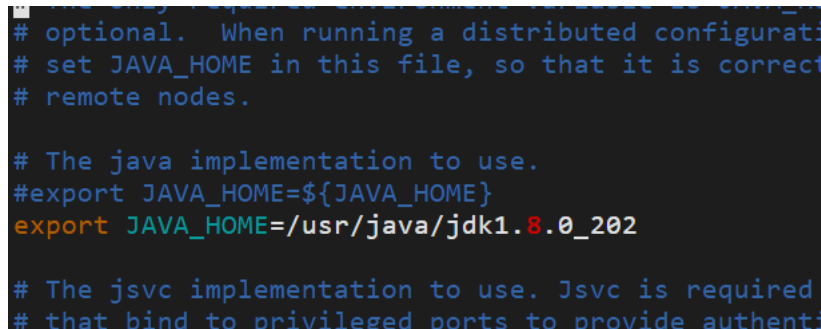
(5) 运行 `source /etc/profile` 命令，使配置生效

7. 安装配置 hadoop

(1) master 上从清华镜像下载 `hadoop-2.7.7.tar.gz` 压缩包，在 `/home` 目录下新建 `spark` 文件夹，将压缩包移入并解压

(2) 进入 `hadoop-2.7.7/etc/hadoop/`

(3) 修改 `hadoop-env.sh`，添加 `JAVA_HOME`

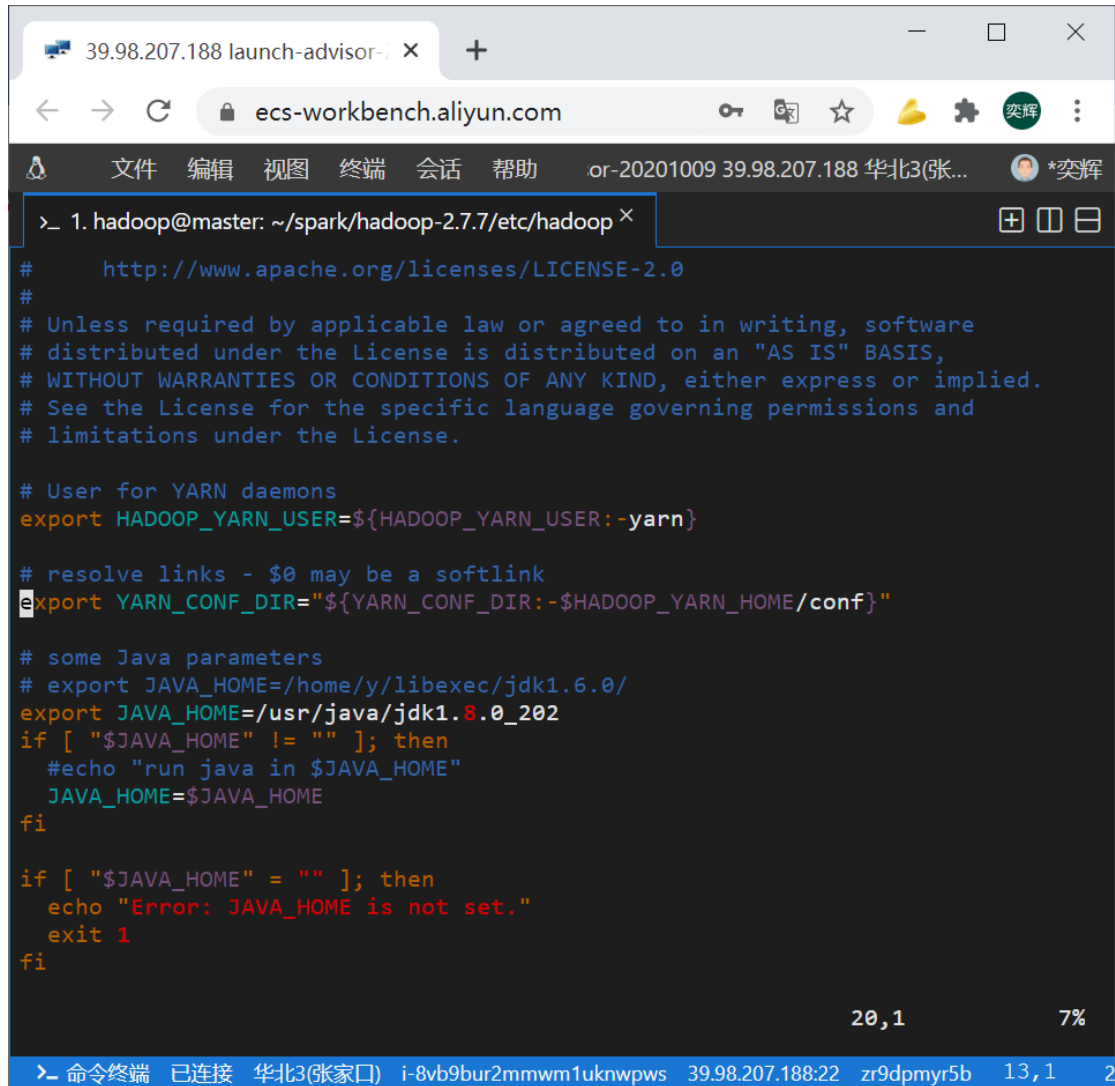


```
# optional. When running a distributed configuration
# set JAVA_HOME in this file, so that it is correct
# remote nodes.

# The java implementation to use.
#export JAVA_HOME=${JAVA_HOME}
export JAVA_HOME=/usr/java/jdk1.8.0_202

# The jsvc implementation to use. Jsvc is required
# that bind to privileged ports to provide authenti
```

(4) 修改 `yarn-env.sh`，添加 `JAVA_HOME`



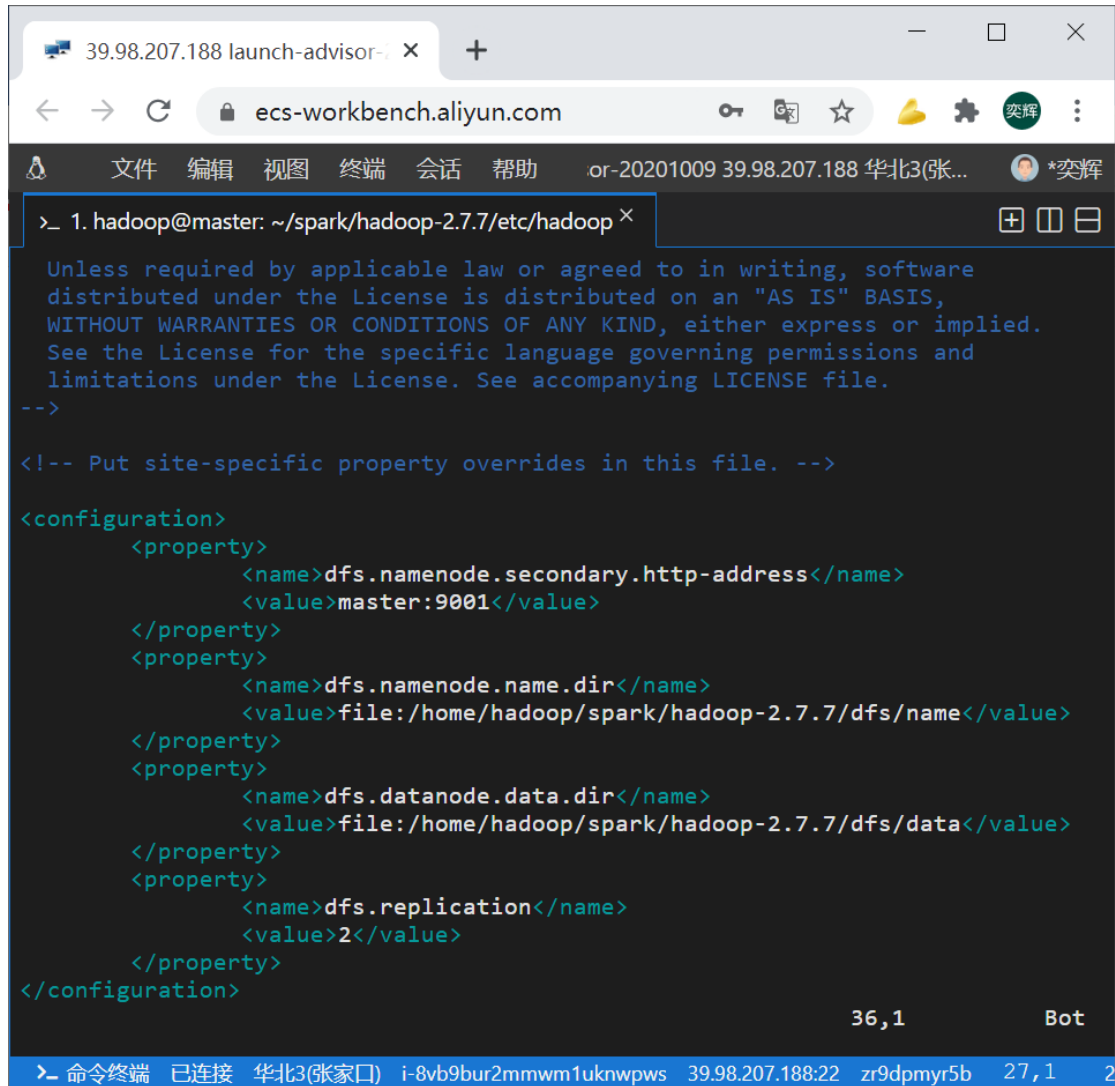
```
>_ 1.hadoop@master: ~/.spark/hadoop-2.7.7/etc/hadoop ×
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
#
# User for YARN daemons
export HADOOP_YARN_USER=${HADOOP_YARN_USER:-yarn}
# resolve links - $0 may be a softlink
export YARN_CONF_DIR="${YARN_CONF_DIR:-$HADOOP_YARN_HOME/conf}"
# some Java parameters
# export JAVA_HOME=/home/y/libexec/jdk1.6.0/
export JAVA_HOME=/usr/java/jdk1.8.0_202
if [ "$JAVA_HOME" != "" ]; then
    #echo "run java in $JAVA_HOME"
    JAVA_HOME=$JAVA_HOME
fi
if [ "$JAVA_HOME" = "" ]; then
    echo "Error: JAVA_HOME is not set."
    exit 1
fi
20,1 7%
```

(5) 修改 core-site.xml, 配置 hdfs 端口



```
<!-- Put site-specific property overrides in this file. -->
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://master:9000/</value>
  </property>
  <property>
    <name>hadoop.tmp.dir</name>
    <value>file:/spark/hadoop-2.7.7/tmp</value>
  </property>
</configuration>
~
~
```

(6) 修改 hdfs-site.xml, 配置 namenode, datanode



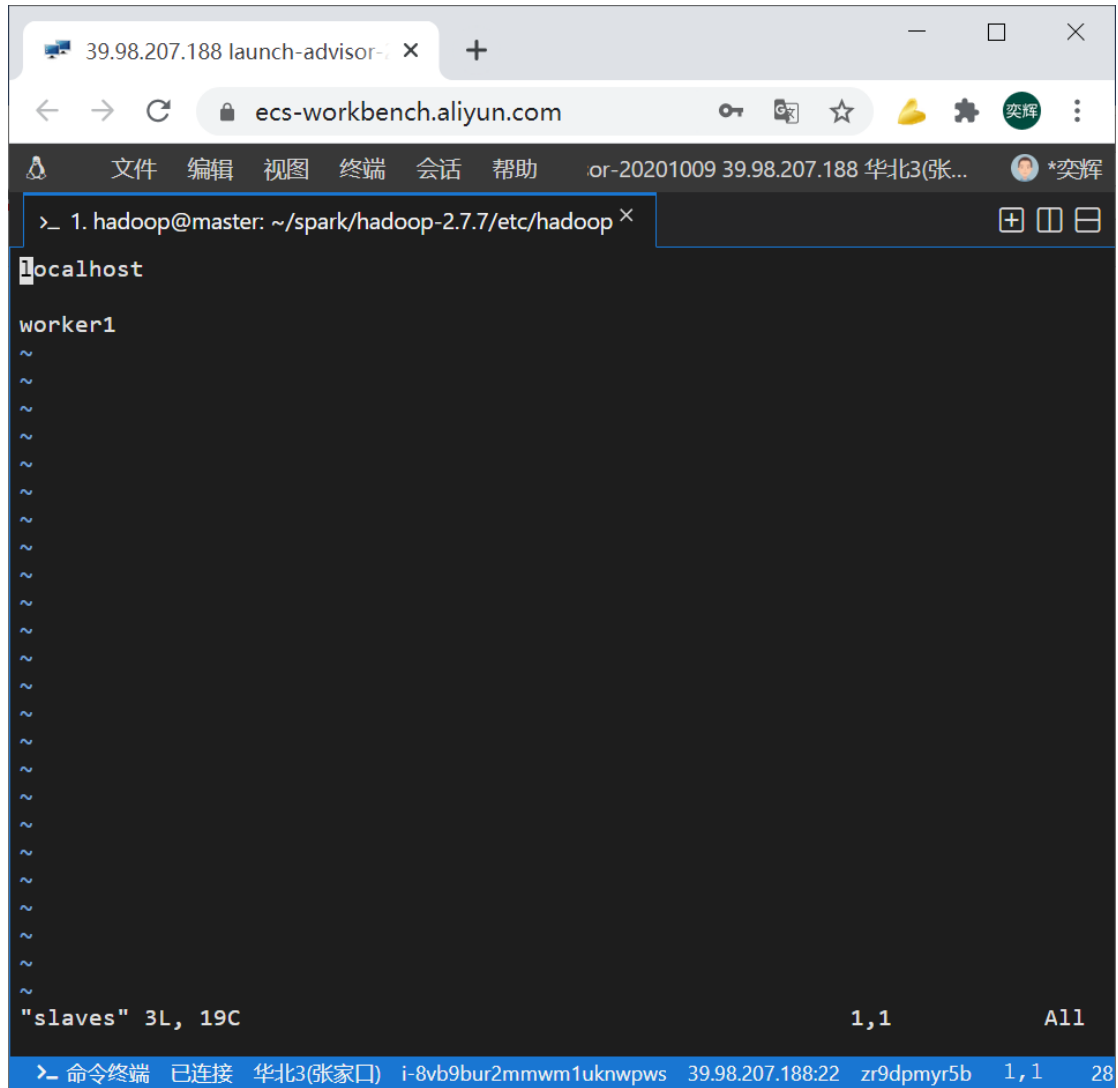
```
Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>master:9001</value>
  </property>
  <property>
    <name>dfs.namenode.name.dir</name>
    <value>file:/home/hadoop/spark/hadoop-2.7.7/dfs/name</value>
  </property>
  <property>
    <name>dfs.datanode.data.dir</name>
    <value>file:/home/hadoop/spark/hadoop-2.7.7/dfs/data</value>
  </property>
  <property>
    <name>dfs.replication</name>
    <value>2</value>
  </property>
</configuration>
```

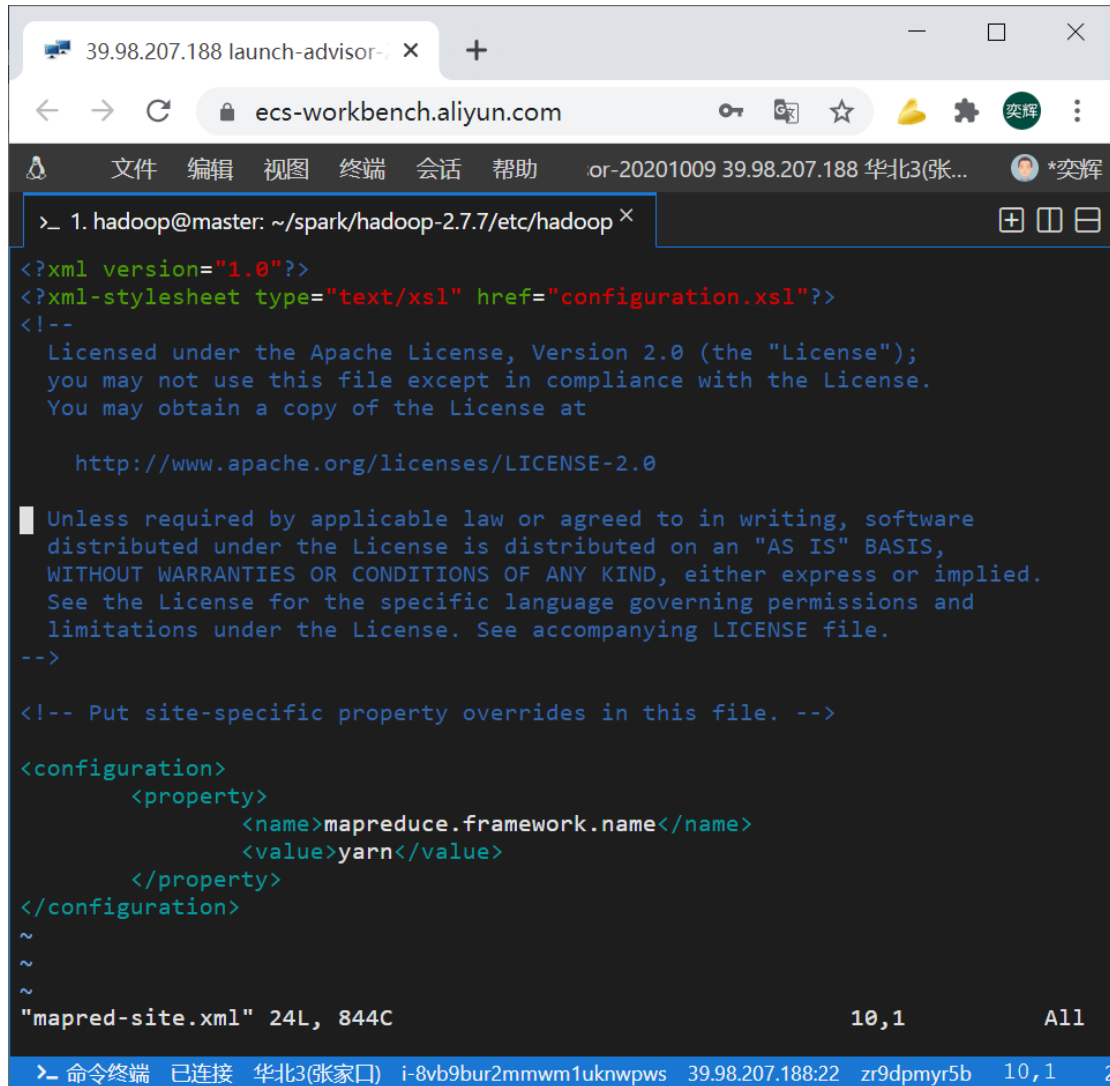
命令终端 已连接 华北3(张家口) i-8vb9bur2mmwm1uknwpws 39.98.207.188:22 zr9dpmyr5b 27,1 2

(7) 修改 slaves 文件，配置 worker 节点的 hostname



(8) 修改 mapred-site.xml

在文件夹中将 mapred-site.xml.template 拷贝成 mapred-site.xml，并修改：



```
> 1. hadoop@master: ~/.spark/hadoop-2.7.7/etc/hadoop x
<?xml version="1.0"?>
<?xml-stylesheet type="text/xsl" href="configuration.xsl"?>
<!--
  Licensed under the Apache License, Version 2.0 (the "License");
  you may not use this file except in compliance with the License.
  You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

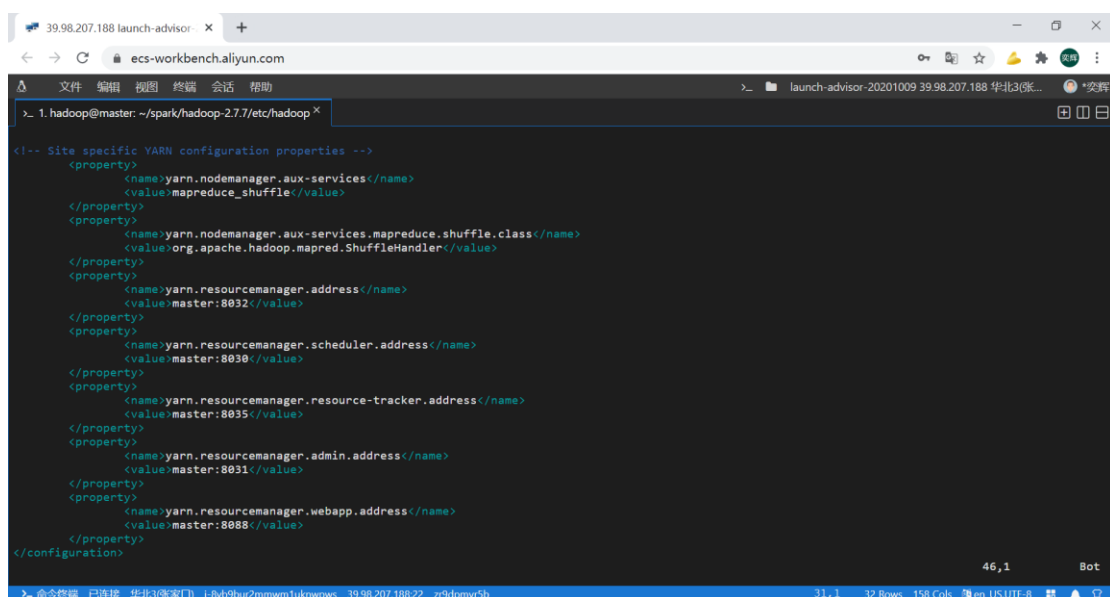
  Unless required by applicable law or agreed to in writing, software
  distributed under the License is distributed on an "AS IS" BASIS,
  WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
  See the License for the specific language governing permissions and
  limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
  <property>
    <name>mapreduce.framework.name</name>
    <value>yarn</value>
  </property>
</configuration>
~
~
~
"mapred-site.xml" 24L, 844C                                10,1      All

> 命令终端 已连接 华北3(张家口) i-8vb9bur2mmwm1uknwpws 39.98.207.188:22 zr9dpmyr5b 10,1 2
```

(9) 配置 yarn-site.xml, 设置端口



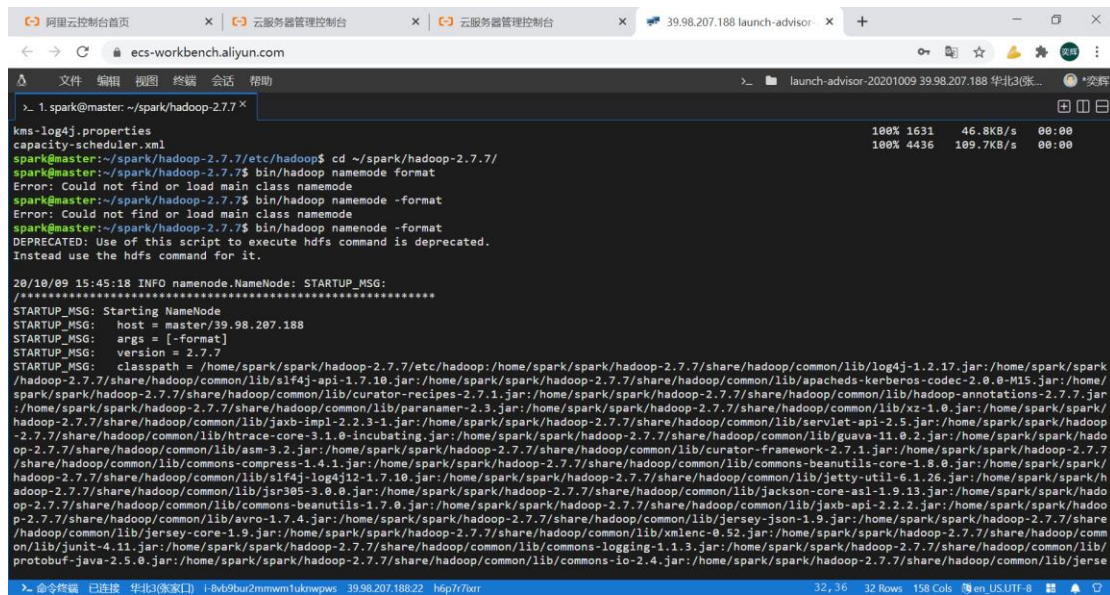
```
> 1. hadoop@master: ~/.spark/hadoop-2.7.7/etc/hadoop x
<!-- Site specific YARN configuration properties -->
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services.mapreduce.shuffle.class</name>
  <value>org.apache.hadoop.mapred.ShuffleHandler</value>
</property>
<property>
  <name>yarn.resourcemanager.address</name>
  <value>master:8032</value>
</property>
<property>
  <name>yarn.resourcemanager.scheduler.address</name>
  <value>master:8030</value>
</property>
<property>
  <name>yarn.resourcemanager.resource-tracker.address</name>
  <value>master:8035</value>
</property>
<property>
  <name>yarn.resourcemanager.admin.address</name>
  <value>master:8031</value>
</property>
<property>
  <name>yarn.resourcemanager.webapp.address</name>
  <value>master:8088</value>
</property>
</configuration>
46,1      Bot

> 命令终端 已连接 华北3(张家口) i-8vb9bur2mmwm1uknwpws 39.98.207.188:22 zr9dpmyr5b 31,1 32 Rows 158 Cols en_US.UTF-8
```

(10) 将配置好的文件分发到各 worker 节点

(11) 格式化 namenode

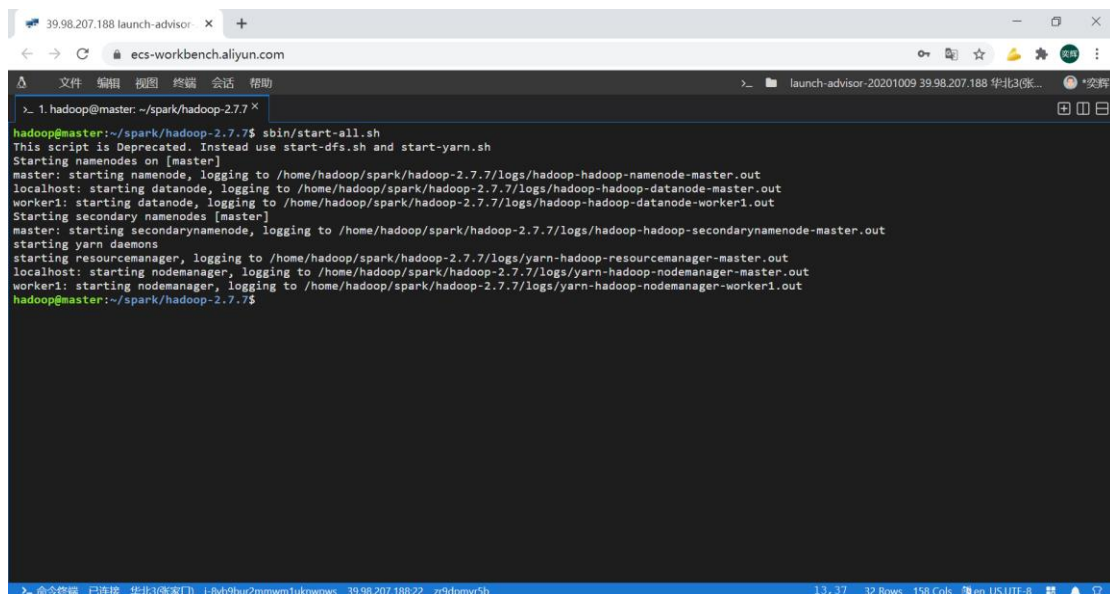
进入 hadoop-2.7.7 文件夹下，运行 bin/hadoop namenode -format



```
kms-log4j.properties
capacity-scheduler.xml
spark@master:~/spark/hadoop-2.7.7/etc/hadoop$ cd ~/spark/hadoop-2.7.7/
spark@master:~/spark/hadoop-2.7.7$ bin/hadoop namenode format
Error: Could not find or load main class namenode
spark@master:~/spark/hadoop-2.7.7$ bin/hadoop namenode -format
Error: Could not find or load main class namenode
spark@master:~/spark/hadoop-2.7.7$ bin/hadoop namenode -format
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

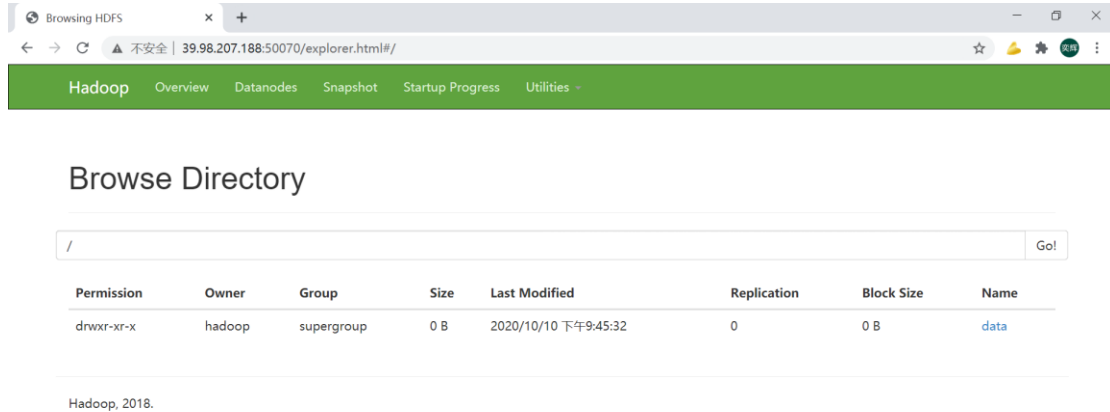
20/10/09 15:45:18 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG: host = master/39.98.207.188
STARTUP_MSG: args = [-format]
STARTUP_MSG: version = 2.7.7
STARTUP_MSG: classpath = /home/spark/spark/hadoop-2.7.7/etc/hadoop:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/log4j-1.2.17.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/slf4j-api-1.7.10.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/apached-kerberos-codec-2.0.0-M15.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/curator-recipes-2.7.1.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/hadoop-annotations-2.7.7.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/paranamer-2.3.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/xz-1.0.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/jaxb-impl-2.2.3-1.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/servlet-api-2.5.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/jtrace-core-3.1.0-incubating.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/guava-11.0.2.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/asm-3.2.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/curator-framework-2.7.1.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/commons-compress-1.4.1.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/commons-beanutils-core-1.8.0.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/jetty-util-6.1.26.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/jsr305-3.0.0.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/commons-beanutils-1.7.0.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/jaxb-api-2.2.2.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/avro-1.7.4.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/jersey-json-1.9.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/jersey-core-1.9.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/xmlenc-0.52.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/junit-4.11.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/commons-logging-1.1.3.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/protobuf-java-2.5.0.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/commons-io-2.4.jar:/home/spark/spark/hadoop-2.7.7/share/hadoop/common/lib/jerse
```

(12) 启动 hadoop 集群



```
hadoop@master:~/spark/hadoop-2.7.7$ sbin/start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [master]
master: starting namenode, logging to /home/hadoop/spark/hadoop-2.7.7/logs/hadoop-hadoop-namenode-master.out
localhost: starting datanode, logging to /home/hadoop/spark/hadoop-2.7.7/logs/hadoop-hadoop-datanode-master.out
worker1: starting datanode, logging to /home/hadoop/spark/hadoop-2.7.7/logs/hadoop-hadoop-datanode-worker1.out
Starting secondary namenodes [master]
master: starting secondarynamenode, logging to /home/hadoop/spark/hadoop-2.7.7/logs/hadoop-hadoop-secondarynamenode-master.out
starting yarn daemons
starting resourcemanager, logging to /home/hadoop/spark/hadoop-2.7.7/logs/yarn-hadoop-resourcemanager-master.out
localhost: starting nodemanager, logging to /home/hadoop/spark/hadoop-2.7.7/logs/yarn-hadoop-nodemanager-master.out
worker1: starting nodemanager, logging to /home/hadoop/spark/hadoop-2.7.7/logs/yarn-hadoop-nodemanager-worker1.out
hadoop@master:~/spark/hadoop-2.7.7$
```

(13) 使用 hdfs 命令上传文件并访问 50070 端口查看运行情况



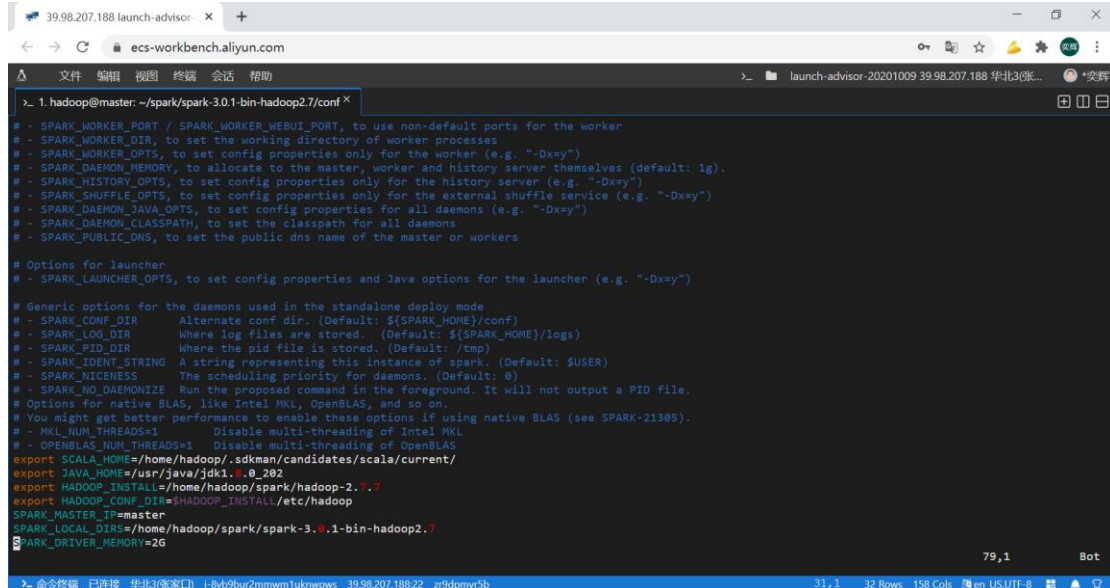
文件上传成功

8. 安装配置 spark

(1) master 上从清华镜像下载 spark-3.0.1-bin-hadoop2.7.tar.gz 压缩包，在/home 目录下 spark 文件夹，将压缩包移入并解压

(2) 进入 spark-3.0.1-bin-hadoop2.7/conf, 复制 spark-env.sh.template 与 slaves.template 文件为 spark-env.sh 与 slaves 文件

(3) 配置 spark-env.sh, 修改运行环境



(4) 配置 slaves, 配置 hostname

Spark Master at spark://mast...

39.98.207.188:8080

Spark Master at spark://master:7077

URL: spark://master:7077
Alive Workers: 2
Cores in use: 3 Total, 0 Used
Memory in use: 3.7 GiB Total, 0.0 B Used
Resources in use:
Applications: 0 Running, 0 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (2)

Worker Id	Address	State	Cores	Memory	Resources
worker-20201010223543-172.26.145.164-34497	172.26.145.164:34497	ALIVE	2 (0 Used)	2.7 GiB (0.0 B Used)	
worker-20201010223545-172.26.55.72-36675	172.26.55.72:36675	ALIVE	1 (0 Used)	1024.0 MiB (0.0 B Used)	

Running Applications (0)

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

Completed Applications (0)

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

可以看到两台服务器都在上面了
(8) 运行官方测试程序

39.98.207.188 launch-advisor

ecs-workbench.aliyun.com

1. hadoop@master: ~/spark/spark-3.0.1-bin-hadoop2.7

```
spark-graphx_2.12-3.0.1.jar
hadoop@master:~/spark/spark-3.0.1-bin-hadoop2.7$ bin/spark-submit --class org.apache.spark.examples.SparkPi --master spark://master:7077 examples/jars/spark-examples_2.12-3.0.1.jar 100
20/10/10 22:43:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
20/10/10 22:43:25 INFO spark.SparkContext: Running Spark version 3.0.1
20/10/10 22:43:25 INFO resource.ResourceUtils: =====
20/10/10 22:43:25 INFO resource.ResourceUtils: Resources for spark.driver:
20/10/10 22:43:25 INFO resource.ResourceUtils: =====
20/10/10 22:43:25 INFO spark.SparkContext: Submitted application: Spark Pi
20/10/10 22:43:25 INFO spark.SecurityManager: Changing view acls to: hadoop
20/10/10 22:43:25 INFO spark.SecurityManager: Changing modify acls to: hadoop
20/10/10 22:43:25 INFO spark.SecurityManager: Changing view acls groups to:
20/10/10 22:43:25 INFO spark.SecurityManager: Changing modify acls groups to:
20/10/10 22:43:25 INFO spark.SecurityManager: SecurityManager: authentication disabled; ui acls disabled; users with view permissions: Set(hadoop); groups with view permissions: Set(); users with modify permissions: Set(hadoop); groups with modify permissions: Set()
20/10/10 22:43:25 INFO util.Utils: Successfully started service 'sparkDriver' on port 41255.
20/10/10 22:43:25 INFO spark.SparkEnv: Registering MapOutputTracker
20/10/10 22:43:25 INFO spark.SparkEnv: Registering BlockManagerMaster
20/10/10 22:43:25 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.storage.DefaultTopologyMapper for getting topology information
20/10/10 22:43:25 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint up
20/10/10 22:43:25 INFO spark.SparkEnv: Registering BlockManagerMasterHeartbeat
20/10/10 22:43:25 INFO storage.DiskBlockManager: Created local directory at /home/hadoop/spark/spark-3.0.1-bin-hadoop2.7/blockmgr-3a52457e-5bc9-4900-bff3-654e3ef50d56
20/10/10 22:43:25 INFO memory.MemoryStore: MemoryStore started with capacity 912.3 MiB
20/10/10 22:43:25 INFO spark.SparkEnv: Registering OutputCommitCoordinator
20/10/10 22:43:25 INFO util.log: Logging initialized @2333ms to org.sparkproject.jetty.util.log.Slf4jLog
20/10/10 22:43:25 INFO server.Server: Jetty-9.4.z-SNAPSHOT; built: 2019-04-29T20:42:08.989Z; git: e1bc35120a6617ee3df052294e433f3a25ce7097; jvm 1.8.0_202-b08
20/10/10 22:43:25 INFO server.Server: Started @2420ms
20/10/10 22:43:25 INFO server.AbstractConnector: Started ServerConnector@24d39a47(HTTP/1.1,[http/1.1]){0.0.0.0:4040}
20/10/10 22:43:25 INFO util.Utils: Successfully started service 'SparkUI' on port 4040.
20/10/10 22:43:25 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@79ab3a71{/jobs,null,AVAILABLE,@Spark}
```

由于速度比较快，截不到运行时的图，只有运行完成截图

Spark Master at spark://mast...

39.98.207.188:8080

Spark Master at spark://master:7077

URL: spark://master:7077

Alive Workers: 2

Cores in use: 3 Total, 0 Used

Memory in use: 3.7 GiB Total, 0.0 B Used

Resources in use:

Applications: 0 Running, 1 Completed

Drivers: 0 Running, 0 Completed

Status: ALIVE

Workers (2)

Worker Id	Address	State	Cores	Memory	Resources
worker-20201010223543-172.26.145.164-34497	172.26.145.164:34497	ALIVE	2 (0 Used)	2.7 GiB (0.0 B Used)	
worker-20201010223545-172.26.55.72-36675	172.26.55.72:36675	ALIVE	1 (0 Used)	1024.0 MiB (0.0 B Used)	

Running Applications (0)

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

Completed Applications (1)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20201010224326-0000	Spark Pi	3	1024.0 MiB		2020/10/10 22:43:26	hadoop	FINISHED	7 s

验证 spark 搭建成功

四，参考资料

- 1.<https://hadoop.apache.org/docs/stable/hadoop-project-dist/Hadoop-common/ClusterSetup.html>
2. <http://spark.apache.org/docs/latest/spark-standalone.html>
3. <https://www.cnblogs.com/tanjiyuan/p/11002988.html>
4. <https://www.cnblogs.com/zhangkaimin/p/10823454.html>