

云主机实现大数据实验



华为技术有限公司

目录

1 准备华为云环境	3
1.1 实验介绍	3
1.1.1 关于本实验	3
1.1.2 实验目的	3
1.1.3 准备实验环境	5
1.2 购买华为云 ECS	5
1.3 购买 OBS	9
2 搭建 Hadoop 集群	13
2.1 实验介绍	13
2.1.1 关于本实验	13
2.1.2 实验目的	13
2.2 Hadoop 集群搭建	13
2.2.1 配置 ECS	13
2.2.2 安装 OpenJDK	18
2.3 搭建 Hadoop 集群	20
2.3.1 搭建 Hadoop 集群	20
2.3.2 测试与 OBS 互联	30
3 MapReduce 程序实验	32
3.1 实验介绍	32
3.1.1 关于本实验	32
3.2 实验目的	32
3.3 MapReduce 存算分离	32
3.3.1 测试 Hadoop 集群功能	32
4 释放云服务资源	34
4.1 释放弹性云服务器 ECS	34
4.2 删除 OBS 桶	35
5 附录 Linux 常用命令	36

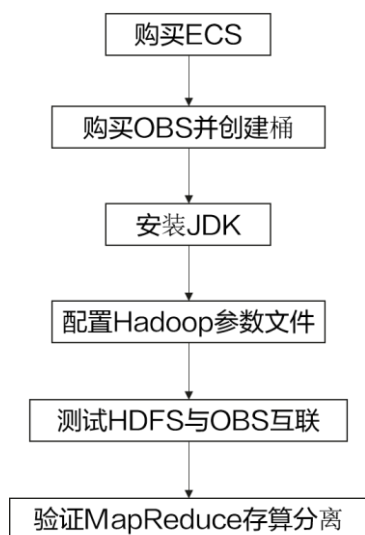
5.1 ls 命令	36
5.2 cd 命令	36
5.3 pwd 命令	36
5.4 mkdir 命令	37
5.5 rm 命令	37
5.6 rmdir 命令	37
5.7 mv 命令	38
5.8 cp 命令	38
5.9 cat 命令	39
5.10 more 命令	39
5.11 less 命令	40
5.12 head 命令	41
5.13 tail 命令	41
5.14 vim 命令	42
5.14.1 启动 vim	42
5.14.2 vim 的模式	42
5.14.3 插入命令	42
5.14.4 查找命令	42
5.14.5 撤销和重做	43
5.14.6 删除命令	43
5.14.7 拷贝和粘贴	43
5.14.8 退出命令	44
5.14.9 帮助命令	44

1 准备华为云环境

1.1 实验介绍

1.1.1 关于本实验

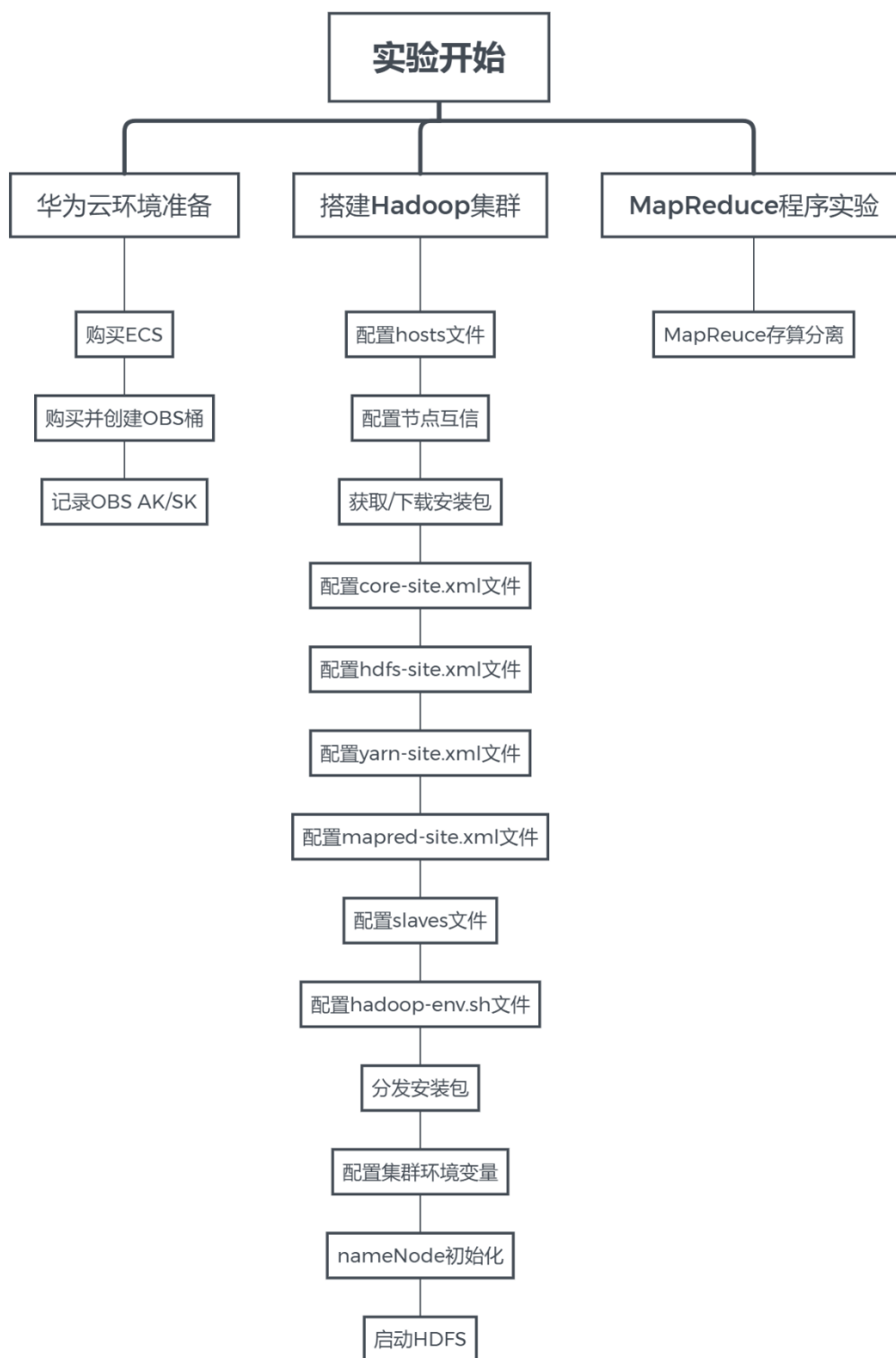
本实验通过购买华为云 ECS 和在华为云上选取对象存储服务 OBS 服务，为后续实验提供搭建环境基础。本实验的基本步骤包含：购买并配置 ECS；购买 OBS 并获取 AK、SK 信息；搭建 Hadoop 集群；搭建 MapReduce 集群并验证存算分离。



1.1.2 实验目的

- 掌握华为云上购买 ECS 步骤。
- 掌握华为云上选取对象存储服务 OBS。
- 掌握 Hadoop 集群搭建。
- 掌握 MapReduce 实验实现存算分离。

1.1.3 实验流程图



1.1.4 准备实验环境

为了满足基于 ECS 实验需要，建议每位学员提前注册华为云账号并充值，确保通过 PUTTY 访问外网畅通。

检查设备

实验开始之前请每位学员检查自己的实验设备是否齐全，实验清单如下。

设备名称	数量	备注
弹性云服务器 ECS	4 个	CPU 架构为 “鲲鹏计算”， “鲲鹏通用计算增强型”， 2vCPUs 4GB 选择 “公共镜像”， CentOS7.6，系统盘建议配置 40GB
对象存储服务 OBS	1 个	选择 “华北-北京四”， “标准存储”
弹性公网 IP	4 个	需购买弹性公网 IP

1.2 购买华为云 ECS

步骤 1 浏览器登录华为云

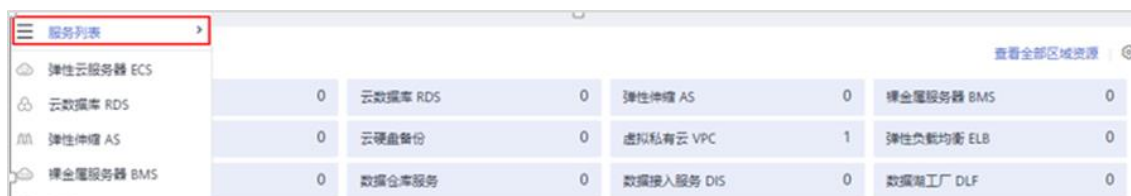
打开华为云地址：<https://www.huaweicloud.com/>，点击“登录”，输入用户名、密码，如下图：



步骤 2 点击“控制台”，如下图：



点击“服务列表”，如下图：



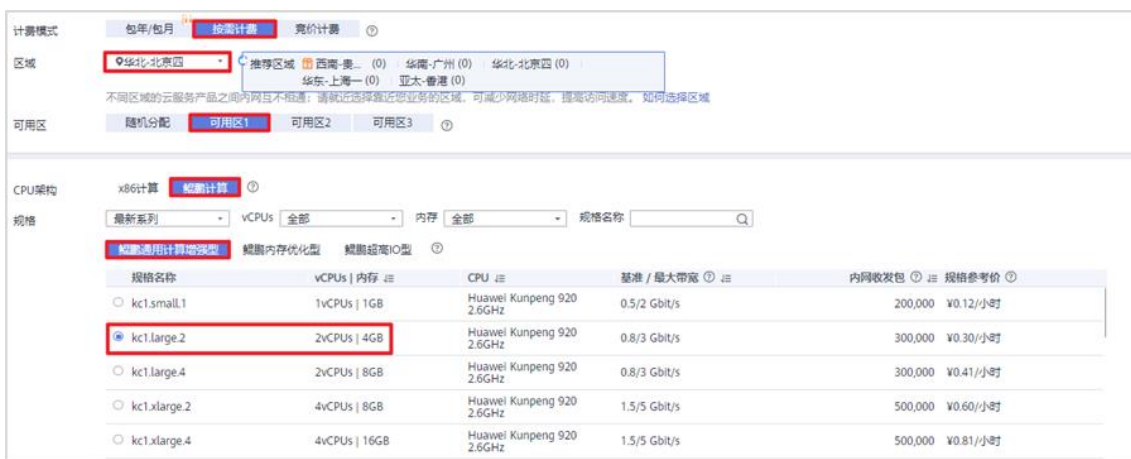
选择“弹性云服务器 ECS”，如下图：



步骤 3 选择“买弹性云服务器 ECS”，如下图：



选择“按需计费”，“可用区 2”，CPU 架构“鲲鹏计算”，选择“鲲鹏通用计算增强型”，2vCPUs|4GB，如下图：



步骤 4 配置操作系统和磁盘

选择“公共镜像”，CentOS7.6，系统盘建议配置 40GB，购买数量 4 台，点击“网络配置”，如下图：



步骤 5 配置网络，网络选择“vpc-default”，安全组选择“Sys-default”，“现在购买”，选择“全动态 BGP”，“按流量计费”，“5M”，点击“高级配置”，如下图：



步骤 6 配置密码，自定义云服务器名称，自行设置 root 登录密码，云备份选择“暂不购买”，点击“确认配置”，点中“我已经阅读并同意”，点击“立即购买”，如下图：



注：本次需购买 4 个 ECS，每个 ECS 规格相同，其中 1 个主节点、3 个从节点。

步骤 7 点击“我已经阅读并同意”，可以点击“返回云服务器列表”，创建过程需要等待几分钟。



1.3 购买 OBS

步骤 1 进入控制台，选择“对象存储服务”，如下图：



步骤 2 选择“创建桶”，如下图：



步骤 3 自定义桶名称，选择“标准存储”，点击“立即创建”，如下图：





步骤 7 获取 AK/SK，点击“我的凭证”，选择“访问秘钥”，如下图：



点击新增访问秘钥，根据提示进行操作。



步骤 8 通过手机号码，接收短信验证码，点击“确定”，如下图：



操作完成后，得到文件，打开即可得到 AK/SK。



A	B	C
User Name	Access Key Id	Secret Access Key
hvv...	8MIA...	Ttsx...

2 搭建 Hadoop 集群

2.1 实验介绍

2.1.1 关于本实验

本实验通过搭建开源 hadoop 集群帮助学员掌握 hadoop 搭建方法。并且使开源 hadoop 与华为云 OBS 服务互联，使 hadoop 集群可读取 OBS 数据。

2.1.2 实验目的

- 掌握 hadoop 集群搭建方法
- 掌握 hadoop 集群与华为云 OBS 互联方法

2.2 Hadoop 集群搭建

2.2.1 配置 ECS

步骤 1 使用 putty 登录 ECS 下载 Putty 工具访问网址

<https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html>，选择 putty.exe

Alternative binary files

The installer packages above will provide versions of all of these (except PuTTYtel),
(Not sure whether you want the 32-bit or the 64-bit version? Read the [FAQ entry](#).)

putty.exe (the SSH and Telnet client itself)

32-bit:	putty.exe	(or by FTP)	(signature)
64-bit:	putty.exe	(or by FTP)	(signature)

pscp.exe (an SCP client, i.e. command-line secure file copy)

32-bit:	pscp.exe	(or by FTP)	(signature)
64-bit:	pscp.exe	(or by FTP)	(signature)

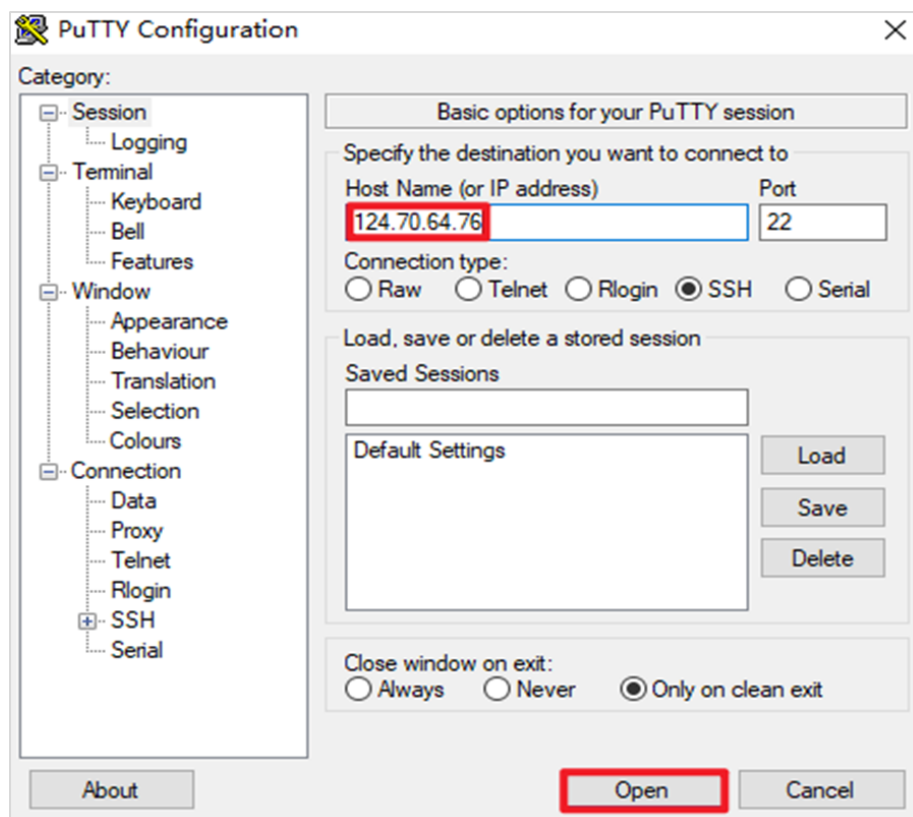
psftp.exe (an SFTP client, i.e. general file transfer sessions much like FTP)

32-bit:	psftp.exe	(or by FTP)	(signature)
64-bit:	psftp.exe	(or by FTP)	(signature)

puttytel.exe (a Telnet-only client)

32-bit:	puttytel.exe	(or by FTP)	(signature)
64-bit:	puttytel.exe	(or by FTP)	(signature)

网址根据实际 node 节点 IP 地址进行填写。

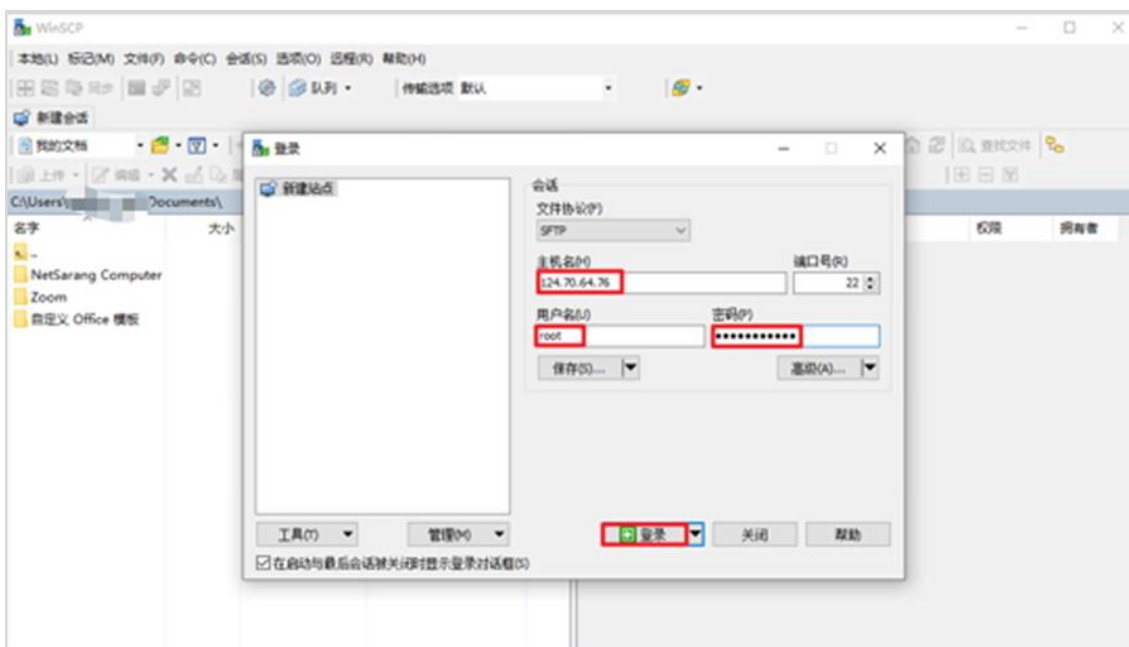


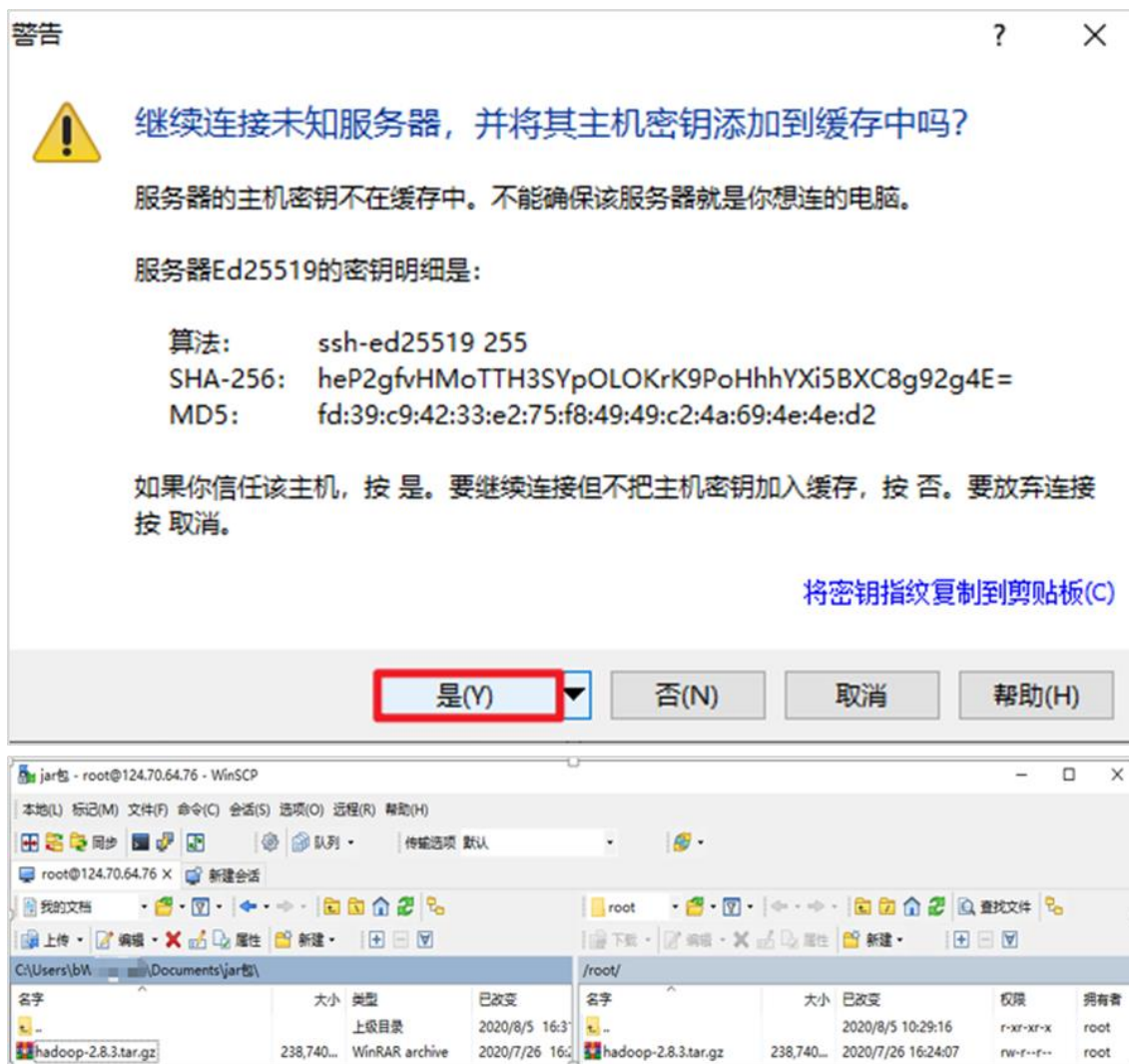


步骤 2 node1 节点登录成功后执行命令在线下载安装包

wget https://archive.apache.org/dist/hadoop/common/hadoop-2.8.3/hadoop-2.8.3.tar.gz

或通过 WinSCP 上传至服务器，访问 <https://winscp.net/eng/docs/lang:chs> 下载安装





步骤 3 下载 OBSFileSystem

登录

<https://github.com/huaweicloud/obsa-hdfs/tree/master/release>

下载 hadoop-huaweicloud-2.8.3-hw-39.jar

shuangyan2 update release jar files			062df25 on 18 Jun	History
..				
doc	update release doc and jar files	2 months ago		
hadoop-huaweicloud-2.8.3-hw-35.jar	update release doc and jar files	2 months ago		
hadoop-huaweicloud-2.8.3-hw-36.jar	update release doc and jar files	2 months ago		
hadoop-huaweicloud-2.8.3-hw-37.jar	update release doc and jar files	2 months ago		
hadoop-huaweicloud-2.8.3-hw-39.jar	update release jar files	last month		
hadoop-huaweicloud-3.1.1-hw-35.jar	update release doc and jar files	2 months ago		
hadoop-huaweicloud-3.1.1-hw-36.jar	update release doc and jar files	2 months ago		
hadoop-huaweicloud-3.1.1-hw-37.jar	update release doc and jar files	2 months ago		
hadoop-huaweicloud-3.1.1-hw-39.jar	update release jar files	last month		

点击下载 jar 包



步骤 4 下载 OpenJDK

执行命令

```
wget https://github.com/AdoptOpenJDK/openjdk8-binaries/releases/download/jdk8u191-b12/OpenJDK8U-jdk_aarch64_linux_hotspot_8u191b12.tar.gz
```

步骤 5 关闭防火墙

```
systemctl stop firewalld
systemctl disable firewalld
```

```
[root@ecs-1-0001 ~]# systemctl stop firewalld
[root@ecs-1-0001 ~]# systemctl disable firewalld
```

步骤 6 配置节点互信

node1~node4 节点分别执行如下命令：

```
#ssh-keygen -t rsa
```

提问框按默认连续回车即可，生成/root/.ssh/id_rsa.pub 文件。

```
[root@ecs-1-0001 ~]# ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/root/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /root/.ssh/id_rsa.
Your public key has been saved in /root/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:/YBIe/64ah+J8k3g3bmSoh5fKXp4qH2lPwRoREEIfHc root@ecs-1-0001
The key's randomart image is:
+---[RSA 2048]---+
|o. ++.          |
| ..... E        |
```

node1~node4 节点分别执行命令 cat /root/.ssh/id_rsa.pub 命令。

```
cat /root/.ssh/id_rsa.pub
```

```
[root@ecs-0001 ~]# cat /root/.ssh/id_rsa.pub
ssh-rsa AAAAB3NzaClyc2EAAAADAQABAAQCCayM/Qqk8QP8suWTZ5sUEO3AUhvd4
Mph+A2eiDBzdzcTY7+HVf+oUNpUUt3aUdUQGDjSNFvQxa7af7csHz3fju2kxwNvIxVRI
gcPG0yqGVWlqr5srt6FV9YjG3Xx6W5RiHToAEmYJ/w9GXlrMS+tT7epgM6uhRylL30rOr
yzyIdOU4T6M+qozK7WXYCzIbQ3gc0FnVtW8ponT9NgkB5lPHM0LuuusVdXqwjBG/gQuYR
fuQSRNlxBSnOQRpFi8nKmtGqhh6XHXjN8cIZHKAAe3LJ33RsViLluIVXoV7t root@ecs-
```

将 4 个节点的内容拷贝汇总到一个文本中，再将该文本内容拷贝到 node1、node2、node3、node4 的/root/.ssh/authorized_keys 中。

```
vim /root/.ssh/authorized_keys
```

node1~node4 节点分别执行命令 vim /etc/hosts，加入 node1~node4 对应 IP 及 node 节点名。

```
vim /etc/hosts
```

```
root@ecs-0001 ~# vim /etc/hosts
#::1 localhost localhost.localdomain localhost6 localhost6.localdomain6
#127.0.0.1 localhost localhost.localdomain localhost4 localhost4.localdomain4
#127.0.0.1 localhost localhost
#192.168.0.173 ecs-0001-0001
#192.168.0.232 ecs-0002-0002
#192.168.0.88 ecs-0003-0003
#192.168.0.195 ecs-0004-0004
```

node1~node4 节点分别执行命令 ssh node1~node4，选择 yes 后，确保能够无密码跳转到目的节点。node1 节点无密码跳转到 node2 节点如下图，其余同理。

```
[root@ecs-0001 ~]# ssh ecs-0002-0002
The authenticity of host 'ecs-0002-0002 (192.168.0.232)' can't be established.
ECDSA key fingerprint is SHA256:L5aXUqIBHa3517WJJS1evtKvH12FwtMerW9s/AZLxp8.
ECDSA key fingerprint is MD5:cl:ec:cf:53:85:46:85:7a:ea:a5:17:58:f4:dc:df:d3.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'ecs-bianxinwen-0002,192.168.0.232' (ECDSA) to the list of known hosts.
Last login: Thu Aug 6 11:04:05 2020 from 119.3.119.18

Welcome to Huawei Cloud Service
```

2.2.2 安装 OpenJDK

步骤 1 创建目录

node1~node4 四个节点分别执行下列命令。

```
mkdir -p /home/modules/data/buf/
```

```
mkdir -p /home/test_tools/
```

```
mkdir -p /home/nm/localdir
```

```
[root@ecs-0001 ~]# mkdir -p /home/modules/data/buf/
[root@ecs-0001 ~]# mkdir -p /home/test_tools/
[root@ecs-0001 ~]# mkdir -p /home/nm/localdir
[root@ecs-0002 ~]# mkdir -p /home/modules/data/buf/
[root@ecs-0002 ~]# mkdir -p /home/test_tools/
[root@ecs-0002 ~]# mkdir -p /home/nm/localdir
[root@ecs-0003 ~]# mkdir -p /home/modules/data/buf/
[root@ecs-0003 ~]# mkdir -p /home/test_tools/
[root@ecs-0003 ~]# mkdir -p /home/nm/localdir
[root@ecs-0004 ~]# mkdir -p /home/modules/data/buf/
[root@ecs-0004 ~]# mkdir -p /home/test_tools/
[root@ecs-0004 ~]# mkdir -p /home/nm/localdir
```

步骤 2 安装 OpenJDK

在 node1 节点，执行如下命令，将 jdk 安装包拷贝到/usr/lib/jvm 目录下。

```
cp OpenJDK8U-jdk_aarch64_linux_hotspot_8u191b12.tar.gz /usr/lib/jvm/
```

在 node1 节点，执行如下命令，将 jdk 安装包拷贝到其他几个节点，其中 ecs-000 为自己节点前面名称，根据实际情况填写，i 为节点数。

```
for i in {2..4};do scp /usr/lib/jvm/OpenJDK8U-jdk_aarch64_linux_hotspot_8u191b12.tar.gz root@node-000${i}:/usr/lib/jvm;/done
```

```
[root@ecs-0001 ~]# for i in {2..4};do scp /usr/lib/jvm/OpenJDK8U-jdk_aarch64_linux_hotspot_8u191b12.tar.gz root@ecs-000${i}:/usr/lib/jvm;/done
OpenJDK8U-jdk_aarch64_linux_hotspot_8u191b12. 100% 72MB 122.6MB/s 00:00
OpenJDK8U-jdk_aarch64_linux_hotspot_8u191b12. 100% 72MB 122.5MB/s 00:00
OpenJDK8U-jdk_aarch64_linux_hotspot_8u191b12. 100% 72MB 117.3MB/s 00:00
```

步骤 3 在 node1~node4 四个节点分别执行命令

```
cd /usr/lib/jvm/
```

```
tar zxvf OpenJDK8U-jdk_aarch64_linux_hotspot_8u191b12.tar.gz
```

```
[root@ecs-0001 ~]# cd /usr/lib/jvm/
[root@ecs-0001 jvm]# tar zxvf OpenJDK8U-jdk_aarch64_linux_hotspot_8u191b12.tar.gz
./jdk8u191-b12/
./jdk8u191-b12/bin/
./jdk8u191-b12/bin/pack200
./jdk8u191-b12/bin/tnameserv
./jdk8u191-b12/bin/jstatd
./jdk8u191-b12/bin/rmiregistry
```

步骤 4 在 node1~node4 四个节点上编辑/etc/profile 增加如下的配置

```
vim /etc/profile
```

```
[root@ecs-0001 jvm]# vim /etc/profile
```

添加

```
export JAVA_HOME=/usr/lib/jvm/jdk8u191-b12
```

```
for i in /etc/profile.d/*.sh; do
    if [ -r "$i" ]; then
        if [ "${-#*i}" != "$-" ]; then
            . "$i"
        else
            . "$i" >/dev/null
        fi
    fi
done

unset i
unset -f pathmunge

export JAVA_HOME=/usr/lib/jvm/jdk8u191-b12
```

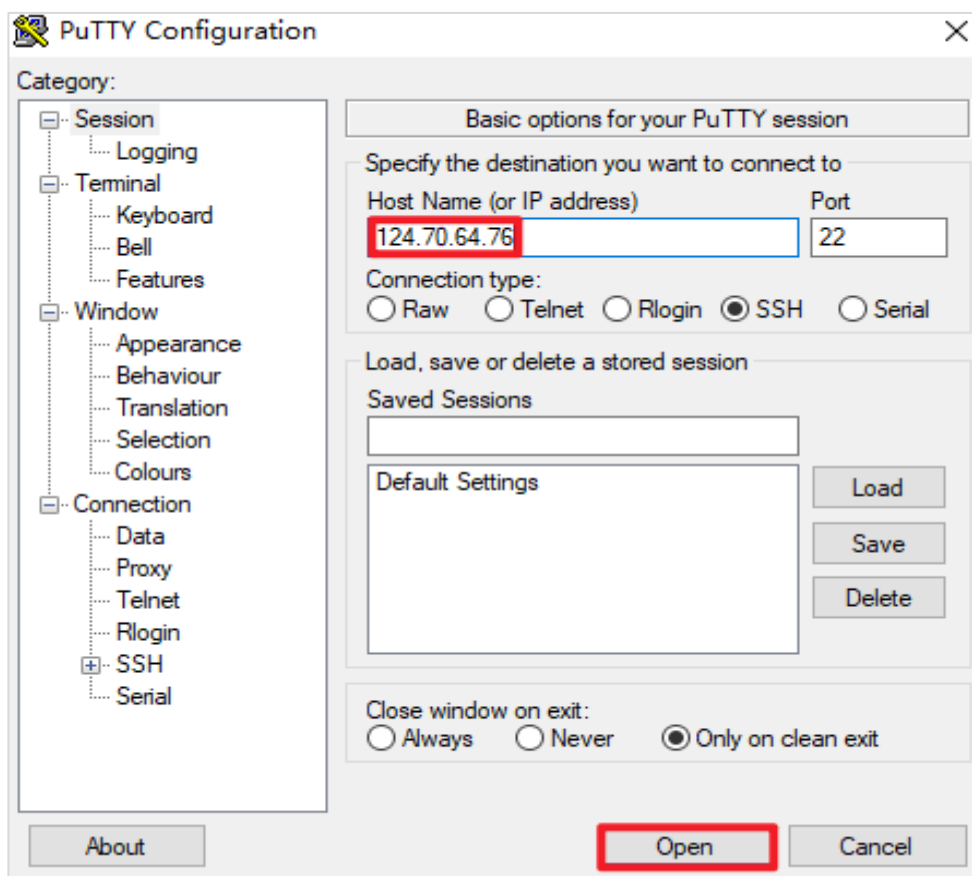
步骤 5 确认 java 版本

```
source /etc/profile
java -version
```

2.3 搭建 Hadoop 集群

2.3.1 搭建 Hadoop 集群

步骤 1 登录 node1 节点，解压 hadoop 安装包



```
cp hadoop-2.8.3.tar.gz /home/modules/
cd /home/modules/
tar zxvf hadoop-2.8.3.tar.gz
```

步骤 2 配置 hadoop 环境变量

```
vim /home/modules/hadoop-2.8.3/etc/hadoop/hadoop-env.sh
编辑加入
export JAVA_HOME=/usr/lib/jvm/jdk8u191-b12
```

步骤 3 配置 hadoop core-site.xml 配置文件

执行命令：

```
vim /home/modules/hadoop-2.8.3/etc/hadoop/core-site.xml
```

参数配置如下：

```
<configuration>
<property>
  <name>fs.obs.readahead.inputstream.enabled</name>
  <value>true</value>
</property>
<property>
  <name>fs.obs.buffer.max.range</name>
  <value>6291456</value>
```

```
</property>
<property>
  <name>fs.obs.buffer.part.size</name>
  <value>2097152</value>
</property>
<property>
  <name>fs.obs.threads.read.core</name>
  <value>500</value>
</property>
<property>
  <name>fs.obs.threads.read.max</name>
  <value>1000</value>
</property>
<property>
  <name>fs.obs.write.buffer.size</name>
  <value>8192</value>
</property>
<property>
  <name>fs.obs.read.buffer.size</name>
  <value>8192</value>
</property>
<property>
  <name>fs.obs.connection.maximum</name>
  <value>1000</value>
</property>
<property>
  <name>fs.defaultFS</name>
  <value>hdfs://node1:8020</value>
</property>
<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/modules/hadoop-2.8.3/tmp</value>
</property>
<property>
  <name>fs.obs.access.key</name>
  <value>U6YSHXXWCOHJGCBCMOQG</value>
</property>
<property>
  <name>fs.obs.secret.key</name>
  <value>kjHdB4s1ldylm1sv0CwAC73UK075LCJ7E7qaBuAF</value>
</property>
<property>
  <name>fs.obs.endpoint</name>
  <value>obs.cn-north-4.myhuaweicloud.com:5080</value>
</property>
<property>
  <name>fs.obs.buffer.dir</name>
  <value>/home/modules/data/buf</value>
```

```

</property>
<property>
  <name>fs.obs.impl</name>
  <value>org.apache.hadoop.fs.obs.OBSFileSystem</value>
</property>
<property>
  <name>fs.obs.connection.ssl.enabled</name>
  <value>>false</value>
</property>
<property>
  <name>fs.obs.fast.upload</name>
  <value>>true</value>
</property>
<property>
  <name>fs.obs.socket.send.buffer</name>
  <value>65536</value>
</property>
<property>
  <name>fs.obs.socket.recv.buffer</name>
  <value>65536</value>
</property>
<property>
  <name>fs.obs.max.total.tasks</name>
  <value>20</value>
</property>
<property>
  <name>fs.obs.threads.max</name>
  <value>20</value>
</property>
</configuration>

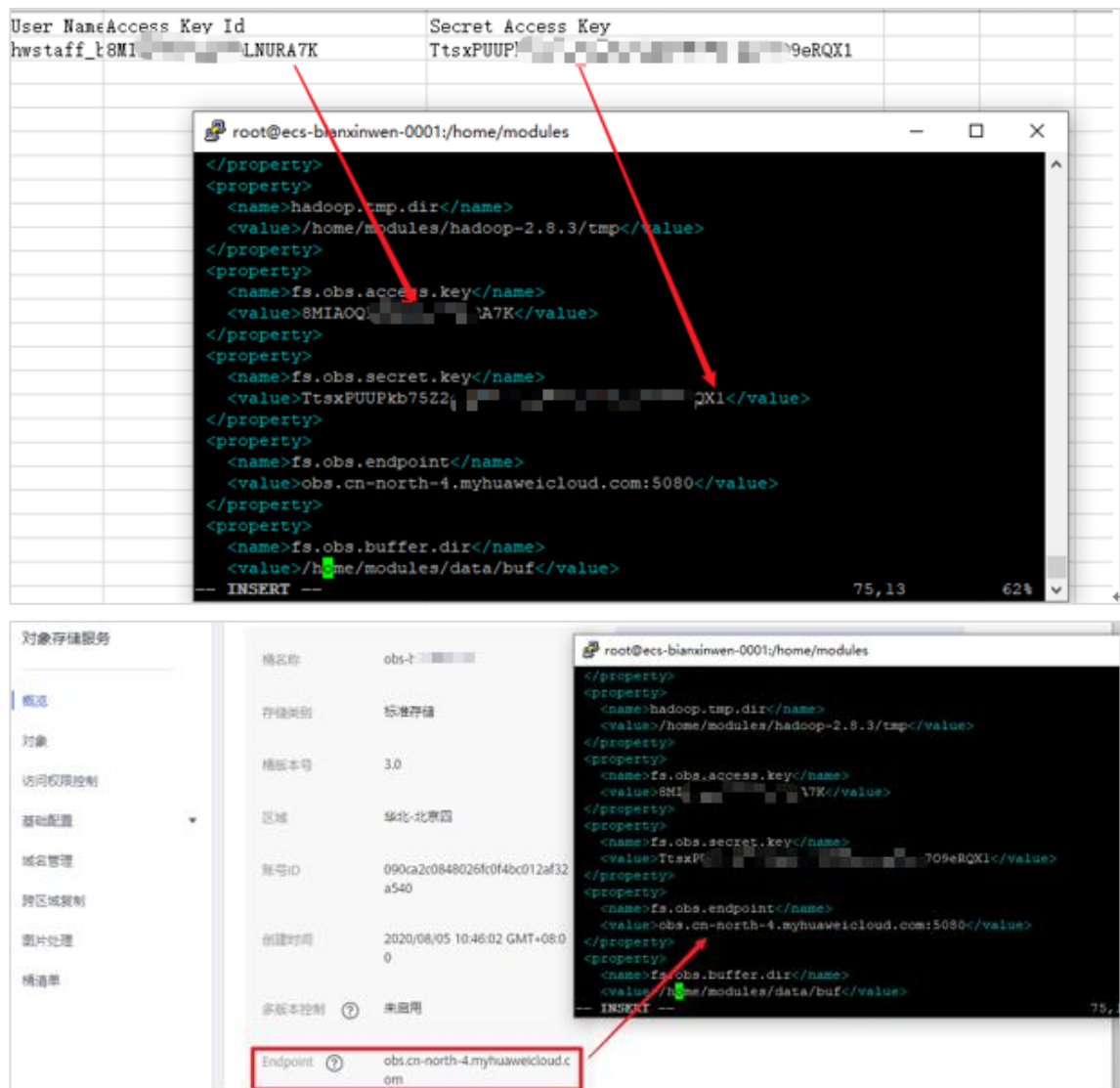
```

注：fs.obs.access.key、fs.obs.secret.key、fs.obs.endpoint 需根据实际情况修改。

```

<property>
  <name>fs.defaultFS</name>
  <value>hdfs://ecs--0001:8020</value>
</property>
<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/modules/hadoop-2.8.3/tmp</value>
</property>

```

The screenshot displays the configuration of HDFS-site.xml in a terminal window. The configuration includes the following properties:

```

</property>
<property>
  <name>hadoop.tmp.dir</name>
  <value>/home/modules/hadoop-2.8.3/tmp</value>
</property>
<property>
  <name>fs.obs.access.key</name>
  <value>8MIAOQ...A7K</value>
</property>
<property>
  <name>fs.obs.secret.key</name>
  <value>TtsxPUUPkb75Z2...9eRQX1</value>
</property>
<property>
  <name>fs.obs.endpoint</name>
  <value>obs.cn-north-4.myhuaweicloud.com:5080</value>
</property>
<property>
  <name>fs.obs.buffer.dir</name>
  <value>/home/modules/data/buf</value>
</property>
-- INSERT --
  
```

Red arrows indicate the mapping of values from the OBS console to the configuration file. The OBS console shows the following details:

- 桶名称: obs-...
- 存储类别: 标准存储
- 桶版本号: 3.0
- 区域: 华北-北京四
- 桶ID: 090ca2c0848026fc0f4bc012af32a540
- 创建时间: 2020/08/05 10:46:02 GMT+08:00
- Endpoint: obs.cn-north-4.myhuaweicloud.com

步骤 4 配置 hdfs-site.xml

```
#vim /home/modules/hadoop-2.8.3/etc/hadoop/hdfs-site.xml
```

参数配置如下:

```

<configuration>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>node1:50090</value>
  </property>
  <property>
    <name>dfs.namenode.secondary.https-address</name>
    <value>node1:50091</value>
  </property>
</configuration>
  
```

```
</configuration>
```

```
<configuration>
  <property>
    <name>dfs.replication</name>
    <value>3</value>
  </property>
  <property>
    <name>dfs.namenode.secondary.http-address</name>
    <value>ecs-1-██████████-0001:50090</value>
  </property>
  <property>
    <name>dfs.namenode.secondary.https-address</name>
    <value>ecs-1-██████████-0001:50091</value>
  </property>
</configuration>
```

步骤 5 配置 yarn-site.xml

```
vim /home/modules/hadoop-2.8.3/etc/hadoop/yarn-site.xml
```

参数配置如下(node1 为自己实际节点名):

```
<configuration>
<property>
  <name>yarn.nodemanager.local-dirs</name>
<value>/home/nm/localdir</value>
</property>
<property>
  <name>yarn.nodemanager.resource.memory-mb</name>
  <value>28672</value>
</property>
<property>
  <name>yarn.scheduler.minimum-allocation-mb</name>
  <value>3072</value>
</property>
<property>
  <name>yarn.scheduler.maximum-allocation-mb</name>
  <value>28672</value>
</property>
<property>
  <name>yarn.nodemanager.resource.cpu-vcores</name>
  <value>38</value>
</property>
<property>
  <name>yarn.scheduler.maximum-allocation-vcores</name>
```

```
<value>38</value>
</property>
<property>
  <name>yarn.nodemanager.aux-services</name>
  <value>mapreduce_shuffle</value>
</property>
<property>
  <name>yarn.resourcemanager.hostname</name>
  <value>node1</value>
</property>
<property>
  <name>yarn.log-aggregation-enable</name>
  <value>true</value>
</property>
<property>
  <name>yarn.log-aggregation.retain-seconds</name>
  <value>106800</value>
</property>
<property>
  <name>yarn.nodemanager.vmem-check-enabled</name>
  <value>false</value>
  <description>Whether virtual memory limits will be enforced for containers</description>
</property>
<property>
  <name>yarn.nodemanager.vmem-pmem-ratio</name>
  <value>4</value>
  <description>Ratio between virtual memory to physical memory when setting memory limits for
containers</description>
</property>
<property>
  <name>yarn.resourcemanager.scheduler.class</name>
  <value>org.apache.hadoop.yarn.server.resourcemanager.scheduler.fair.FairScheduler</value>
</property>
<property>
  <name>yarn.log.server.url</name>
  <value>http://node1:19888/jobhistory/logs</value>
</property>
</configuration>
```

步骤 6 配置 mapred-site.xml

执行下列命令

```
cd /home/modules/hadoop-2.8.3/etc/hadoop/
mv mapred-site.xml.template mapred-site.xml
vim /home/modules/hadoop-2.8.3/etc/hadoop/mapred-site.xml
```

参数配置如下(node1 为自己实际节点名):

```
<configuration>
```

```
<property>
  <name>mapreduce.framework.name</name>
  <value>yarn</value>
</property>
<property>
  <name>mapreduce.jobhistory.address</name>
  <value>node1:10020</value>
</property>
<property>
  <name>mapreduce.jobhistory.webapp.address</name>
  <value>node1:19888</value>
</property>
<property>
  <name>mapred.task.timeout</name>
  <value>1800000</value>
</property>
</configuration>
```

步骤 7 配置 slaves

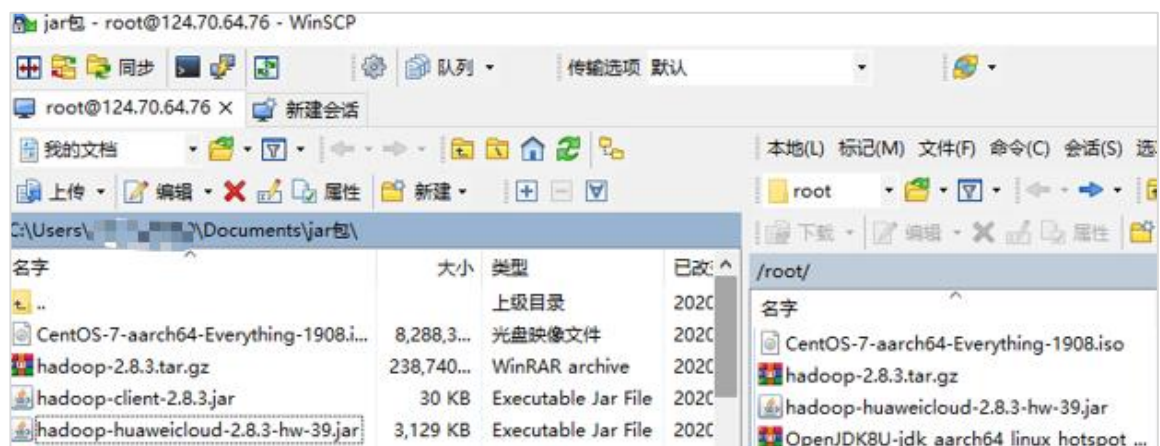
```
vim /home/modules/hadoop-2.8.3/etc/hadoop/slaves
```

编辑内容如下：

```
node2
node3
node4
```

步骤 8 配置 jar 包

通过 WinSCP 上传至服务器



```
cp hadoop-huaweicloud-2.8.3-hw-39.jar /home/modules/hadoop-2.8.3/share/hadoop/common/lib/
cp hadoop-huaweicloud-2.8.3-hw-39.jar /home/modules/hadoop-2.8.3/share/hadoop/tools/lib/
cp hadoop-huaweicloud-2.8.3-hw-39.jar /home/modules/hadoop-
2.8.3/share/hadoop/httpfs/tomcat/webapps/webhdfs/WEB-INF/lib/
cp hadoop-huaweicloud-2.8.3-hw-39.jar /home/modules/hadoop-2.8.3/share/hadoop/hdfs/lib/
```

```
[root@ecs-...-0001 ~]# cp hadoop-huaweicloud-2.8.3-hw-39.jar /home/module
s/hadoop-2.8.3/share/hadoop/common/lib/
[root@ecs-...-0001 ~]# cp hadoop-huaweicloud-2.8.3-hw-39.jar /home/module
s/hadoop-2.8.3/share/hadoop/tools/lib
[root@ecs-...-0001 ~]# cp hadoop-huaweicloud-2.8.3-hw-39.jar /home/module
s/hadoop-2.8.3/share/hadoop/httpfs/tomcat/webapps/webhdfs/WEB-INF/lib/
[root@ecs-...-0001 ~]# cp hadoop-huaweicloud-2.8.3-hw-39.jar /home/module
s/hadoop-2.8.3/share/hadoop/hdfs/lib/
```

步骤 9 分发 hadoop 包到各节点

node1~node4 四个节点下执行下列命令(node-000 为自己节点前面名称, i 为节点数, 根据自己实际节点名进行编写)

```
for i in {2..4};do scp -r /home/modules/hadoop-2.8.3 root@node-000${i}:/home/modules/;done
```

步骤 10 配置环境变量

node1~node4 四个节点下执行下列命令:

```
vim /etc/profile
```

添加如下:

```
export HADOOP_HOME=/home/modules/hadoop-2.8.3
export PATH=$JAVA_HOME/bin:$PATH
export PATH=$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$PATH
export HADOOP_CLASSPATH=/home/modules/hadoop-
2.8.3/share/hadoop/tools/lib/*:$HADOOP_CLASSPATH
```

node1 节点添加如图所示, 其余节点同理。



```
fi
for i in /etc/profile.d/*.sh /etc/profile.d/sh.local ; do
    if [ -r "$i" ]; then
        if [ "${-#*i}" != "$-" ]; then
            . "$i"
        else
            . "$i" >/dev/null
        fi
    fi
done
unset i
unset -f pathmunge

export JAVA_HOME=/usr/lib/jvm/jdk8u191-b12

export HADOOP_HOME=/home/modules/hadoop-2.8.3
export PATH=$JAVA_HOME/bin:$PATH
export PATH=$HADOOP_HOME/bin:$HADOOP_HOME/sbin:$PATH
export HADOOP_CLASSPATH=/home/modules/hadoop-2.8.3/share/hadoop/tools/lib/*:$HAD
OOP_CLASSPATH
```

node1~node4 四个节点下执行下列命令:

```
source /etc/profile
```

```
[root@ecs-1-0001 ~]# vim /etc/profile
[root@ecs-1-0001 ~]# source /etc/profile

[root@ecs-1-0002 ~]# vim /etc/profile
[root@ecs-1-0002 ~]# source /etc/profile

[root@ecs-1-0003 ~]# vim /etc/profile
[root@ecs-1-0003 ~]# source /etc/profile

[root@ecs-1-0004 ~]# vim /etc/profile
[root@ecs-1-0004 ~]# source /etc/profile
```

步骤 11 namenode 初始化

node1 节点执行 namenode 初始化

执行下列命令：

hdfs namenode -format

初始化成功后，启动 hdfs 如图所示：

```
[root@ecs-1-0001 ~]# hdfs namenode -format
20/08/07 14:46:44 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   user = root
STARTUP_MSG:   host = localhost/127.0.0.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.8.3
STARTUP_MSG:   classpath = /home/modules/hadoop-2.8.3/etc/hadoop:/home/modules/h
```

步骤 12 启动 HDFS

node1 节点执行命令：

start-dfs.sh

```
[root@ecs-1-0001 hadoop]# hdfs namenode -format
20/08/07 11:02:46 INFO namenode.NameNode: STARTUP_MSG:
/*****
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   user = root
STARTUP_MSG:   host = localhost/127.0.0.1
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.8.3
STARTUP_MSG:   classpath = /home/modules/hadoop-2.8.3/etc/hadoop:/home/modules/h
adoop-2.8.3/share/hadoop/common/lib/htrace-core4-4.0.1-incubating.jar:/home/modu
```

步骤 13 执行 hdfs 命令

hdfs dfs -mkdir /bigdata

hdfs dfs -ls /


```
[root@ecs-4-0001 ~]# hdfs dfs -mkdir /bigdata
2020/08/07 15:13:05 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
[root@ecs-4-0001 ~]# hdfs dfs -ls /
2020/08/07 15:14:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x  - root supergroup          0 2020-08-07 15:13 /bigdata
drwxr-xr-x  - root supergroup          0 2020-08-07 15:08 /11
```

2.3.2 测试与 OBS 互联

步骤 1 在 OBS 上传文件 选择“对象存储服务 OBS”, 如下图:



选择“桶名称”, 如下图:





华为云 控制台

对象存储服务

对象 / obs-bianxin

对象 已删除对象 碎片

对象是数据的基本单位，在OBS中文件和文件夹都是对象。您可以上传任何类型（文本、图片、视频等）的文件，并在桶中对这些文件进行管理。 [了解更多](#)

上传对象 新建文件夹 恢复 删除 修改存储类别

输入对象名前缀搜索

名称	存储类别	大小	加密状态	恢复状态	最后修改时间	操作
<input type="checkbox"/> palyerinfo.txt	标准存储	118 byte	未加密	--	2020/08/07 15:29:28...	下载 分享 更多

OBS 文件上传成功。

步骤 2 执行 hdfs 命令查看 OBS 文件

obs-bxw 为实际自己创建的桶名称

```
# hdfs dfs -ls obs://obs-bxw/
```

Hadoop 集群与 OBS 互联成功。

```
[root@ecs-...-0001 ~]# hdfs dfs -ls obs://obs-bxw/
20/08/07 16:31:49 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
20/08/07 16:31:49 INFO services.ObsClient: Storage|1|HTTP+XML|ObsClient|2020-08-07 16:31:49|2020-08-07 16:31:49|0|
20/08/07 16:31:49 WARN services.ObsClient: [OBS SDK Version=3.20.2.1];[Endpoint=http://obs.cn-north-4.myhuaweicloud.com:5080/];[Access Mode=Virtual Hosting]
20/08/07 16:31:49 INFO internal.RestStorageService: OkHttp cost 84 ms to apply http request
```


3 MapReduce 程序实验

3.1 实验介绍

3.1.1 关于本实验

本实验使用 MapReduce 验证存算分离，数据存放在 OBS 上，计算在 ECS 执行。通过本实验可以掌握大数据存算分离操作。

3.2 实验目的

- 掌握 MapReduce 程序执行方法
- 掌握存算分离执行方法

3.3 MapReduce 存算分离

3.3.1 测试 Hadoop 集群功能

步骤 1 node1 节点启动 YARN 执行命令

start-yarn.sh

```
root@ecs-...-0001 ~]# start-yarn.sh
Starting yarn daemons
Starting resourcemanager, logging to /home/modules/hadoop-2.8.3/logs/yarn-root-resourcemanager-ecs-bianxinwen-0001.out
ecs-...-0002: starting nodemanager, logging to /home/modules/hadoop-2.8.3/logs/yarn-root-nodemanager-ecs-bianxinwen-0002.out
ecs-...-0003: starting nodemanager, logging to /home/modules/hadoop-2.8.3/logs/yarn-root-nodemanager-ecs-bianxinwen-0003.out
ecs-...-0004: starting nodemanager, logging to /home/modules/hadoop-2.8.3/logs/yarn-root-nodemanager-ecs-bianxinwen-0004.out
```

步骤 2 测试文件

palyerinfo.txt 内容如下：

Alex James Lax Genu
Kerry Mary Olivia William
Hale Edith Vera Robert
Mary Olivia James Lax
Edith Vera Robertm Genu

执行 hadoop wordcount

obs-bxw 为实际自己创建的桶名

hadoop jar /home/modules/hadoop-2.8.3/share/hadoop/mapreduce/hadoop-mapreduce-examples-2.8.3.jar wordcount obs://obs-bxw/ /output

```
[root@ecs-4c24g2cnq-0001 ~]# hadoop jar /home/modules/hadoop-2.8.3/share/hadoop
/mapreduce/hadoop-mapreduce-examples-2.8.3.jar wordcount obs://obs-bxw/ /
output
20/08/07 17:06:44 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
20/08/07 17:06:44 INFO services.ObsClient: Storage|1|HTTP+XML|ObsClient|2020-
08-07 17:06:44|2020-08-07 17:06:44|101
```

查看结果

hdfs dfs -cat /output/part-r-00000

```
[root@ecs-4c24g2cnq-0001 ~]# hdfs dfs -cat /output/part-r-00000
```

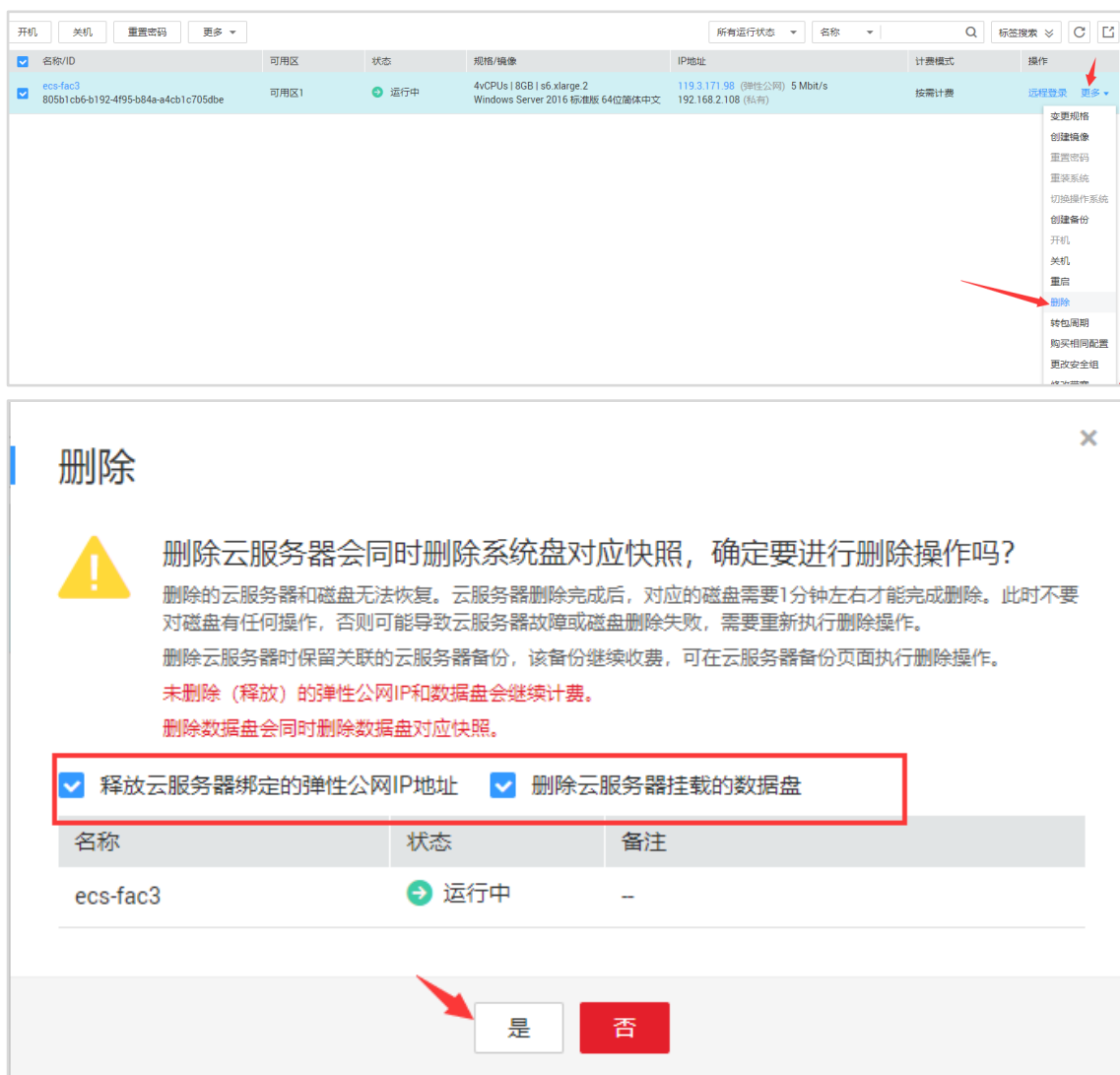
Alex	1
Edith	2
Genu	2
Hale	1
James	2
Kerry	1
Lax	2
Mary	2
Olivia	2
Robert	1
Robertm	1
Vera	2
William	1

至此应用 MapReduce 实现存算分离计算成功!

4 释放云服务资源

4.1 释放弹性云服务器 ECS

进入弹性云服务管理页面，删除弹性云服务服务器 ECS 和弹性公网 IP



删除云服务器会同时删除系统盘对应快照，确定要进行删除操作吗？

删除的云服务器和磁盘无法恢复。云服务器删除完成后，对应的磁盘需要1分钟左右才能完成删除。此时不要对磁盘有任何操作，否则可能导致云服务器故障或磁盘删除失败，需要重新执行删除操作。

删除云服务器时保留关联的云服务器备份，该备份继续收费，可在云服务器备份页面执行删除操作。

未删除（释放）的弹性公网IP和数据盘会继续计费。

删除数据盘会同时删除数据盘对应快照。

☒ 释放云服务器绑定的弹性公网IP地址 ☒ 删除云服务器挂载的数据盘

名称	状态	备注
ecs-fac3	运行中	--

是 否

4.2 删除 OBS 桶

进入 OBS 服务，点击“对象存储”，在右下方点击“删除”，删除桶。



华为云 控制台

对象存储服务 对象存储服务 开源声明 使用流程 使用指南

诚邀您参加对象存储服务使用体验调研，您宝贵的意见和建议是我们持续提升产品体验的源动力，感谢您的参与！

OBS Browser+ 下载 图形化管理工具，支持批量上传大文件，文件夹。 了解更多

obsutil 下载 命令行管理工具，支持桶和对象的基本操作。 了解更多

OBS FTP 下载 文件传输工具，可替代FTP Server管理OBS对象。 了解更多

obsfs 下载 并行文件系统挂载工具，可实现本地文件系统操作对象。 了解更多

在控制台上您还可以创建97个桶。

桶名称	存...	区域...	存...	Data+	对...	创建时间	操作
obs-0880	标准存储	华北-北京...	0 byte		0	2020/10/24 19...	修改存储类别 删除
cxlyyz	标准存储	华东-上海...	824 byte	该区域暂...	2	2020/10/24 17...	修改存储类别 删除

选择“是”删除桶



删除桶

确定要对以下桶进行删除操作吗？
桶删除后无法恢复，请谨慎操作。

桶名称	存储类别	区域
obs-0880	标准存储	华北-北京四

是 否

5 附录 Linux 常用命令

5.1 ls 命令

ls 命令不仅可以查看 linux 文件夹包含的文件而且可以查看文件权限(包括目录、文件夹、文件权限)查看目录信息等等。

命令格式: ls [选项][目录名]

常用参数

- l : 列出长数据串, 包含文件的属性与权限数据等
- a : 列出全部的文件, 连同隐藏文件 (开头为 . 的文件) 一起列出来 (常用)
- d : 仅列出目录本身, 而不是列出目录的文件数据
- h : 将文件容量以较易读的方式 (GB, kB 等) 列出来
- R : 连同子目录的内容一起列出 (递归列出), 等于该目录下的所有文件都会显示出来

5.2 cd 命令

最基本的命令语句, 其他的命令语句要进行操作, 都是建立在使用 cd 命令上的。用于切换当前目录至 dirName。

命令格式: cd [目录名]

5.3 pwd 命令

查看"当前工作目录"的完整路径。

命令格式: pwd [选项]

常用参数:

- P : 显示实际物理路径, 而非使用连接 (link) 路径
- L : 当目录为连接路径时, 显示连接路径

5.4 mkdir 命令

用来创建指定的名称的目录，要求创建目录的用户在当前目录中具有写权限，并且指定的目录名不能是当前目录中已有的目录。

命令格式：mkdir [选项] 目录

常用参数

-m, --mode=模式，设定权限<模式> (类似 chmod)，而不是 rwxrwxrwx 减 umask

-p, --parents 可以是一个路径名称。此时若路径中的某些目录尚不存在,加上此选项后,系统将自动建立好那些尚不存在的目录,即一次可以建立多个目录;

-v, --verbose 每次创建新目录都显示信息

--help 显示此帮助信息并退出

--version 输出版本信息并退出

5.5 rm 命令

删除一个目录中的一个或多个文件或目录，如果没有使用 -r 选项，则 rm 不会删除目录。如果使用 rm 来删除文件，通常仍可以将该文件恢复原状。

命令格式：rm [选项] 文件

常用参数

-f, --force 忽略不存在的文件，从不给出提示。

-i, --interactive 进行交互式删除

-r, -R, --recursive 指示 rm 将参数中列出的全部目录和子目录均递归地删除。

-v, --verbose 详细显示进行的步骤

--help 显示此帮助信息并退出

--version 输出版本信息并退出

5.6 rmdir 命令

该命令从一个目录中删除一个或多个子目录项，删除某目录时也必须具有对父目录的写权限。

命令格式：rmdir [选项] 目录

常用参数

-p 递归删除目录 dirname，当子目录删除后其父目录为空时，也一同被删除。如果整个路径被删除或者由于某种原因保留部分路径，则系统在标准输出上显示相应的信息。

-v, --verbose 显示指令执行过程

5.7 mv 命令

可以用来移动文件或者将文件改名（move (rename) files）。当第二个参数类型是文件时，mv 命令完成文件重命名。当第二个参数是已存在的目录名称时，源文件或目录参数可以有多个，mv 命令将各参数指定的源文件均移至目标目录中。

命令格式：mv [选项] 源文件或目录 目标文件或目录

常用参数

-b：若需覆盖文件，则覆盖前先行备份

-f：force 强制的意思，如果目标文件已经存在，不会询问而直接覆盖

-i：若目标文件 (destination) 已经存在时，就会询问是否覆盖

-u：若目标文件已经存在，且 source 比较新，才会更新(update)

-t：--target-directory=DIRECTORY move all SOURCE arguments into DIRECTORY，即指定 mv 的目标目录，该选项适用于移动多个源文件到一个目录的情况，此时目标目录在前，源文件在后

5.8 cp 命令

将源文件复制至目标文件，或将多个源文件复制至目标目录。

命令格式：cp [选项] 源文件 目录 或 cp [选项] -t 目录 源文件

常用参数

-t --target-directory 指定目标目录

-i --interactive 覆盖前询问（使前面的 -n 选项失效）

-n --no-clobber 不要覆盖已存在的文件（使前面的 -i 选项失效）

-f --force 强行复制文件或目录，不论目的文件或目录是否已经存在

-u --update 使用这项参数之后，只会在源文件的修改时间较目的文件更新时，或是对应的目的文件并不存在，才复制文件

5.9 cat 命令

用来显示文件内容，或者将几个文件连接起来显示，或者从标准输入读取内容并显示，它常与重定向符号配合使用。

命令格式：cat [选项] [文件]

常用参数

-A, --show-all 等价于 -vET

-b, --number-nonblank 对非空输出行编号

-e 等价于 -vE

-E, --show-ends 在每行结束处显示 \$

-n, --number 对输出的所有行编号,由 1 开始对所有输出的行数编号

-s, --squeeze-blank 有连续两行以上的空白行，就代换为一行的空白行

-t 与 -vT 等价

-T, --show-tabs 将跳格字符显示为 ^I

-v, --show-nonprinting 使用 ^ 和 M- 引用，除了 LFD 和 TAB 之外

5.10 more 命令

more 命令和 cat 的功能一样都是查看文件里的内容，但有所不同的是 more 可以按页来查看文件的内容，还支持直接跳转行等功能。

命令格式：more [-dlfpcsu] [-num] [+/- pattern] [+ linenum] [file ...]

常用参数

+n 从第 n 行开始显示

-n 定义屏幕大小为 n 行

+/pattern 在每个档案显示前搜寻该字串 (pattern)，然后从该字串前两行之后开始显示

-c 从顶部清屏，然后显示

-d 提示“Press space to continue, 'q' to quit (按空格键继续，按 q 键退出)”，禁用响铃功能

-l 忽略 Ctrl+I (换页) 字符

-p 通过清除窗口而不是滚屏来对文件进行换页，与 -c 选项相似

-s 把连续的多个空行显示为一行

-u 把文件内容中的下画线去掉

操作指令

Enter: 向下 n 行, 需要定义。默认为 1 行

Ctrl+F: 向下滚动一屏

空格键: 向下滚动一屏

Ctrl+B: 返回上一屏

= : 输出当前行的行号

: f : 输出文件名和当前行的行号

V : 调用 vi 编辑器

!命令 : 调用 Shell, 并执行命令

q : 退出 more

5.11 less 命令

less 与 more 类似, 但使用 less 可以随意浏览文件, 而 more 仅能向前移动, 却不能向后移动, 而且 less 在查看之前不会加载整个文件。

命令格式: less [参数] 文件

常用参数

-b <缓冲区大小> 设置缓冲区的大小

-e 当文件显示结束后, 自动离开

-f 强迫打开特殊文件, 例如外围设备代号、目录和二进制文件

-g 只标志最后搜索的关键词

-i 忽略搜索时的大小写

-m 显示类似 more 命令的百分比

-N 显示每行的行号

-o <文件名> 将 less 输出的内容在指定文件中保存起来

-Q 不使用警告音

-s 显示连续空行为一行

-S 行过长时间将超出部分舍弃

-x <数字> 将“tab”键显示为规定的数字空格

操作命令

/字符串: 向下搜索“字符串”的功能

?字符串：向上搜索“字符串”的功能

n：重复前一个搜索（与 / 或 ? 有关）

N：反向重复前一个搜索（与 / 或 ? 有关）

b 向后翻一页

d 向后翻半页

h 显示帮助界面

Q 退出 less 命令

u 向前滚动半页

y 向前滚动一行

空格键 滚动一行

回车键 滚动一页

[pagedown]：向下翻动一页

[pageup]：向上翻动一页

5.12 head 命令

head 用来显示档案的开头至标准输出中，默认 head 命令打印其相应文件的开头 10 行。

命令格式：head [参数] [文件]

常用参数

-q 隐藏文件名

-v 显示文件名

-c<字节> 显示字节数

-n<行数> 显示的行数

5.13 tail 命令

显示指定文件末尾内容，不指定文件时，作为输入信息进行处理。常用查看日志文件。

命令格式：tail [必要参数] [选择参数] [文件]

常用参数

-f 循环读取

-q 不显示处理信息

- v 显示详细的处理信息
- c<数目> 显示的字节数
- n<行数> 显示行数
- pid=PID 与-f 合用,表示在进程 ID,PID 死掉之后结束.
- q, --quiet, --silent 从不输出给出文件名的首部
- s, --sleep-interval=S 与-f 合用,表示在每次反复的间隔休眠 S 秒

5.14 vim 命令

vim 编辑器是所有 Unix 及 Linux 系统下标准的编辑器, 它的强大不逊色于任何最新的文本编辑器;vi 也是 Linux 中最基本的文本编辑器,vim 就是 vi 的升级版。

5.14.1 启动 vim

在命令行窗口中输入以下命令即可

```
vim
```

直接启动 vim

vim filename 打开 vim 并创建名为 filename 的文件

5.14.2 vim 的模式

正常模式 (按 Esc 或 Ctrl+[进入) 左下角显示文件名或为空

插入模式 (按 i 键进入) 左下角显示--INSERT--

可视模式 (不知道如何进入) 左下角显示--VISUAL--

导航命令

% 括号匹配

5.14.3 插入命令

i 在当前位置生前插入

I 在当前行首插入

a 在当前位置后插入

5.14.4 查找命令

/text 查找 text, 按 n 键查找下一个, 按 N 键查找前一个。

?text 查找 text, 反向查找, 按 n 键查找下一个, 按 N 键查找前一个。

:set hlsearch 高亮搜索结果, 所有结果都高亮显示, 而不是只显示一个匹配。

:set nohlsearch 关闭高亮搜索显示

:nohlsearch 关闭当前的高亮显示, 如果再次搜索或者按下 n 或 N 键, 则会再次高亮。

:set incsearch 逐步搜索模式, 对当前键入的字符进行搜索而不必等待键入完成。

:set wrapscan 重新搜索, 在搜索到文件头或尾时, 返回继续搜索, 默认开启。

:set nu 显示行号

5.14.5 撤销和重做

u 撤销 (Undo)

U 撤销对整行的操作

Ctrl + r 重做 (Redo) , 即撤销的撤销。

5.14.6 删除命令

x 删除当前字符

3x 删除当前光标开始向后三个字符

X 删除当前字符的前一个字符。X=dh

dl 删除当前字符, dl=x

dh 删除前一个字符

dd 删除当前行

10d 删除当前行开始的 10 行。

D 删除当前字符至行尾。D=d\$

d\$ 删除当前字符之后的所有字符 (本行)

5.14.7 拷贝和粘贴

yy 拷贝当前行

nyy 拷贝当前行开始的 n 行, 比如 2yy 拷贝当前行及其下一行。

p 在当前光标后粘贴, 如果之前使用了 yy 命令来复制一行, 那么就在当前行的下一行粘贴。

shift+p 在当前行前粘贴

5.14.8 退出命令

:wq 保存并退出

ZZ 保存并退出

:q! 强制退出并忽略所有更改

:e! 放弃所有修改，并打开原来文件。

5.14.9 帮助命令

:help or F1 显示整个帮助

:help xxx 显示 xxx 的帮助，比如 :help i, :help CTRL-[(即 Ctrl+[的帮助)。

:help 'number' Vim 选项的帮助用单引号括起

:help 特殊键的帮助用<>扩起

:help -t Vim 启动参数的帮助用-

: help i_ 插入模式下 Esc 的帮助，某个模式下的帮助用模式_主题的模式

帮助文件中位于||之间的内容是超链接，可以用 Ctrl+]进入链接，Ctrl+o (Ctrl + t) 返回

其他非编辑命令