華東師範大學

实 验 报 告

（2020 /2021学年 第一学期）

课程名称 大数据系统

实验名称 Hbase

指导教师 姚俊杰

学生姓名	龙旷飞 李泽浩 张炎	班级学号	10172100255 10185102142 10171900239

一．实验环境

阿里云平台

二．实验内容

HBase

一个构建在HDFS上的分布式列存储系统；

基于Google BigTable模型开发的，典型的key/value系统；

Apache Hadoop生态系统中的重要一员，主要用于海量结构化数据存储；

从逻辑上讲，HBase将数据按照表、行和列进行存储。与hadoop一样，Hbase目标主要依靠横向扩展，通过不断增加廉价的商用服务器，来增加计算和存储能力。

三．基本原理

分布式Hbase协调工作说明

zookeeper

1.保证任何时候，集群中只有一个master

2.存贮所有Region的寻址入口

3.实时监控Region Server的状态，将Region server的上线和下线信息实时通知给Master

4.存储Hbase的schema，包括有哪些table，每个table有哪些column family

master

1.为Region server分配region

2.负责region server的负载均衡

3.发现失效的region server并重新分配其上的region

4.GFS上的垃圾回收

5.处理schema更新请求

Region server

1.Region server 维护Master分配给它的region，处理对这些region的IO请求

2.Region server 负责切分在运行过程中变得过大的region

可以看到，client访问hbase上数据的过程并不需要master参与（寻址访问zookeeper和region server，数据读写访问region server），master仅仅维护着table和region的元数据信息，负载很低

四．配置总结

文字叙述

一、环境的准备与安装

通过订购阿里云轻量应用服务器，在端口口进行环境配置

二、Hadoop的安装与伪分布式配置

1.首先创建一个hadoop用户，并且设置密码，为其增加sudo权限
使用hadoop用户登录，之后首先更新一下apt
2.集群与单节点模式都需要用到SSH登录，所以安装一下SSH server
3.登录本机
4.登录需要输入密码。为了方便，将SSH配置成
无密码
5.登录利用ssh-keygen生成密钥，并将密钥加
入
6.编辑完成之后让环境变量生效，之后对其进行检验
7.检验hadoop是否可用
8.检验成功，显示当前hadoop版本信息
9.使用jps命令来判断是否启动成功
10.可以看到，出现了NameNode，DataNode和SecondaryNameNode三个
进程，说明启动成功在50070端口还可以查看节点信息
11.关闭Hadoop的命令
12.将Hadoop加入环境变量，这样在任意目录中都可以使用hdfs命令

三、HBase的安装与伪分布式配置

1.下载HBase
2.添加环境变量并使其生效
3.添加HBase权限并查看HBase版本以检验是否安装成功
4.可以看到，显示HBase的版本
5.安装成功后进行伪分布式的
配置
6.修改指定HBase数据与存储路径
7.登录ssh，启动Hadoop
8.Hadoop启动成功。再启动HBase
9.HBase配置完成

五．配置过程

图表叙述

配置用户和用户组

搭建hadoop集群环境要求所有主机的用户和用户组要完全一致。配置过程如下：

1.新建用户，建议用adduser命令。

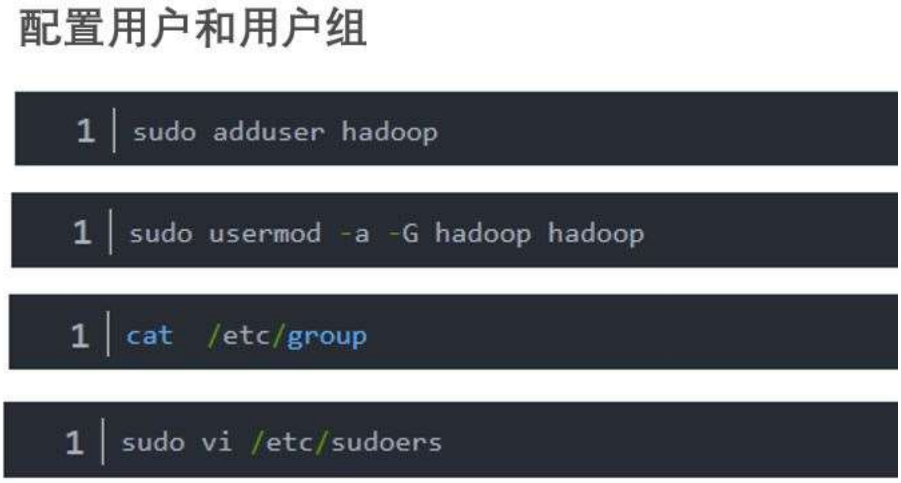
2.在创建hadoop用户的同时也创建了hadoop用户组，下面把hadoop用户加入到hadoop用户组。

3.前面一个hadoop是组名，后面一个hadoop是用户名。完成后输入一下命令查询结果。

4.然后再把hadoop用户赋予root权限，让他可以使用sudo命令。

# User privilege specification root ALL=(ALL:ALL) ALL hadoop ALL=(ALL:ALL) ALL *# 添加这一行配置*

保存退出，hadoop用户就拥有了root权限。



配置用户和用户组

```
sudo adduser hadoop
```

```
sudo usermod -a -G hadoop hadoop
```

```
cat  /etc/group
```

```
sudo vi /etc/sudoers
```

配置主机名

我们一开始就给出了3台主机的主机名（master、slave1、slave2），把虚拟机中三台主机名修改为对应名字。

1.编辑 /etc/hostname ，把里面的内容更改为你所需要的设定的主机名。

2.更改hosts文件，修改（或增加）127.0.1.1后面的名称改为设定值。

3.重启服务器，更改生效

4.验证。重新登录后，会发现主机名已经更改为设定值了。

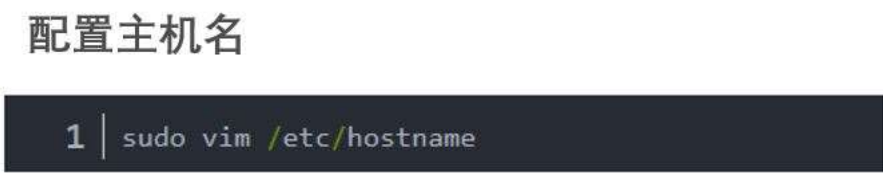
5.修改host文件，配置域名。编辑hosts文件：

添加下面内容：

192.168.0.166 master

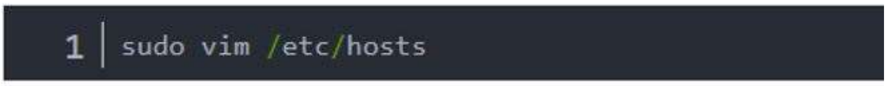
192.168.0.167 slave1

192.168.0.168 slave2

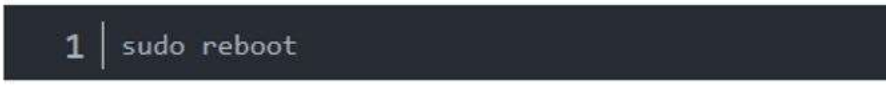


配置主机名

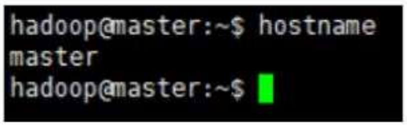
```
sudo vim /etc/hostname
```



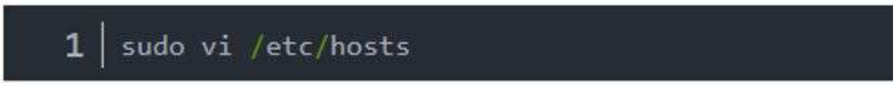
```
sudo vim /etc/hosts
```



```
sudo reboot
```



```
hadoop@master:~$ hostname
master
hadoop@master:~$ 
```



```
sudo vi /etc/hosts
```

安装SSH

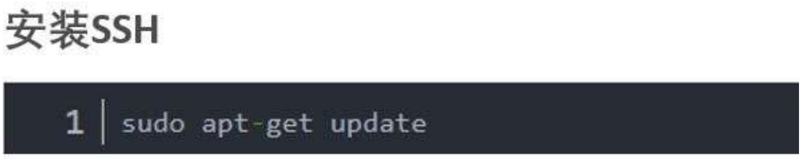
在Hadoop运行过程中，主从机之间是通过SSH进行通信的，所以需要对所有主机进行SSH的安装和配置工作。

1.先更新一下apt。

2.接下来，安装SSH。

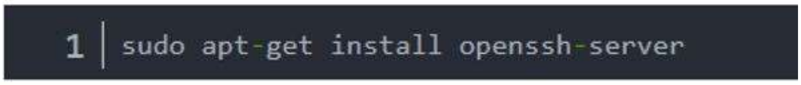
3.安装完成之后，使用下面的命令来查看SSH是否安装成功。

安装好SSH之后，就可以使用SSH进行远程操作了。

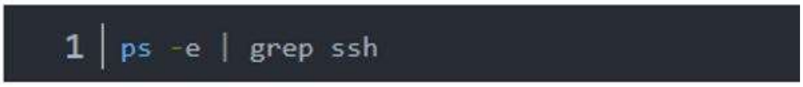


安装SSH

```
sudo apt-get update
```



```
sudo apt-get install openssh-server
```



```
ps -e | grep ssh
```

配置免密登录SSH

主机间免密登录才能实现主机间的顺畅通信，因此该环节非常重要！

1.在master主机上生成密钥对。

2.将公钥（~/.ssh/id_rsa.pub中的内容）复制到文件authorized_keys中去。

3.分别在从机slave1,slave2上都进行同样的操作，同时将两个从机的公玥都复制奥

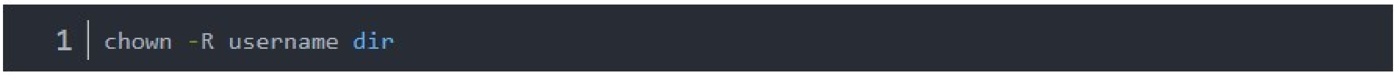
主机master的~/.ssh/authorized_keys中去。

4.将master上的~/.ssh/authorized_keys通过scp命令复制到从机slave1,slave2中去。

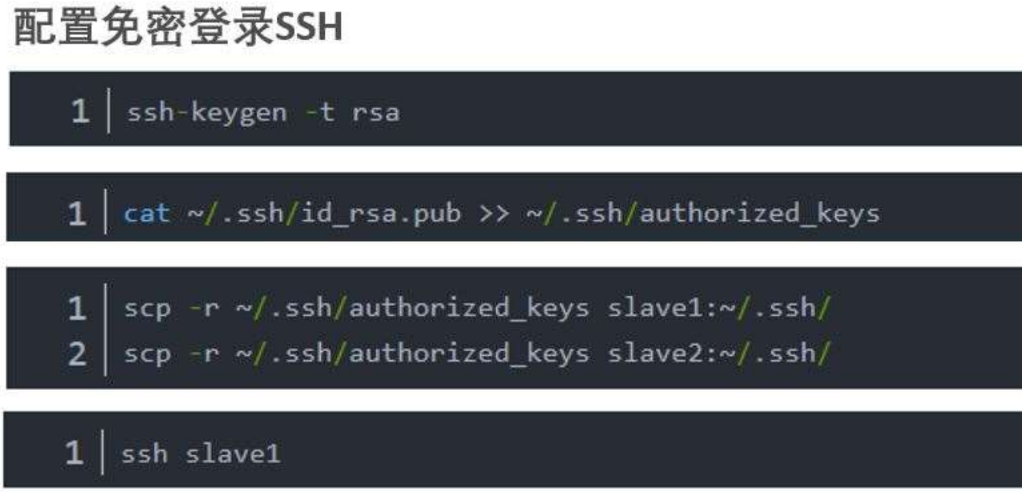
5.验证一下免密登录是否成功。在master上登录slave1、slave2（其他主机上验证方法也是一样的）。

注意：最好将JDK，Hadoop，Zookeeper，HBase所在目录设置为hadoop用户所有，否者后续启动相关应用时会报错。

设置目录及目录下所有文件所属用户：



```
chown -R username dir
```



配置免密登录SSH

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

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```
scp -r ~/.ssh/authorized_keys slave1:~/.ssh/
scp -r ~/.ssh/authorized_keys slave2:~/.ssh/
```

```
ssh slave1
```

关闭防火墙

集群需要开放很多端口，因此，为了避免出现端口未开放的问题，关闭了防火墙

Zookeeper

1.下载Zookeeper。

2.将Zookeeper下载到本地后，通过scp命令将安装包发送到master主机上。下面以master为例介绍安装配置，其他的主机安装和配置是完全一样的。

3.进入服务器根目录下，将zookeeper-3.5.6-bin.tar.gz解压至/usr/local/cluster中。

4.进入 /usr/local/cluster 中，为了方便日后版本的更新，这里使用软链接的方法。

5.设置环境变量，在 ~/.bashrc 添加内容。

6.使环境变量立刻生效。

7.配置zookeeper

** • **建立数据和日志文件

** • **进入conf目录创建并修改zoo.cfg文件

这里还需要在数据目录 /usr/local/cluster/zookeeper/data下新建名为myid的文件，各个主机对应的内容是不同的，master的内容是0，slave1的内容是1，slave2的内容是2，分别对应server.x中的x。server.A=B:C:D,其中A是一个数字， 表示这是第几号server。B是该server所在的IP地址。C配置该server和集群中的leader交换消息所使用的端口。D配置选举leader时所使用的端口。

8.使用scp命令，将配置好的Zookeeper发送到其他从节点上去

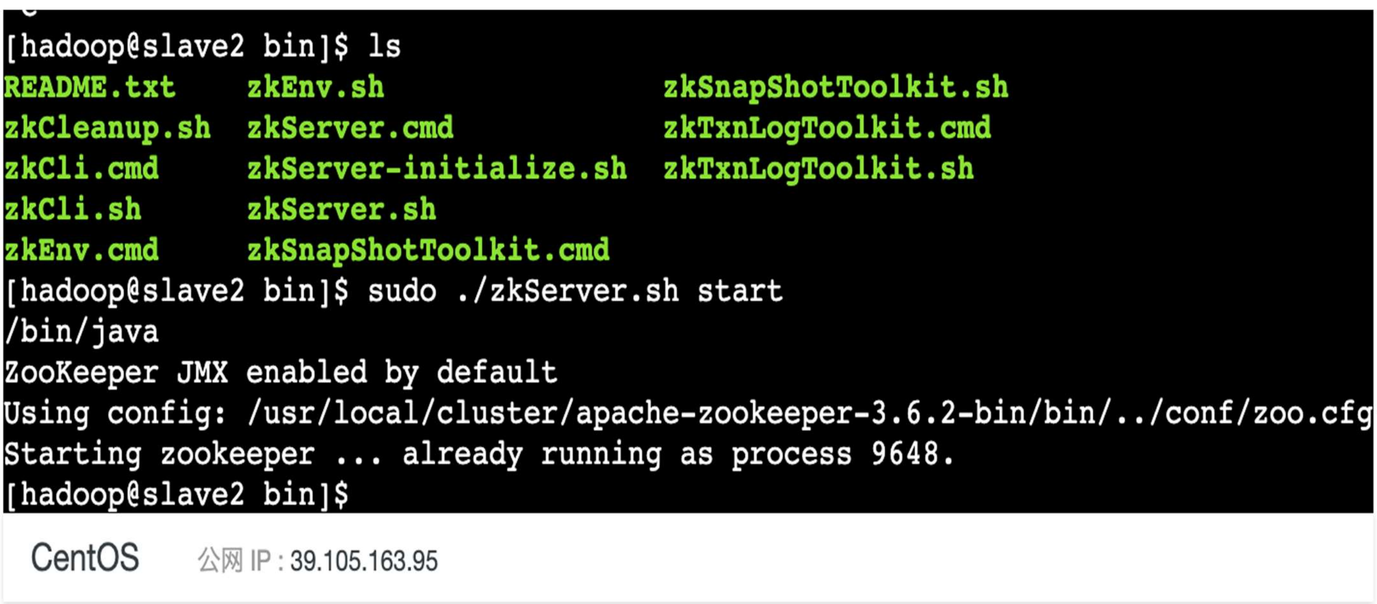
9.启动zookeeper。

在各个节点执行命令：

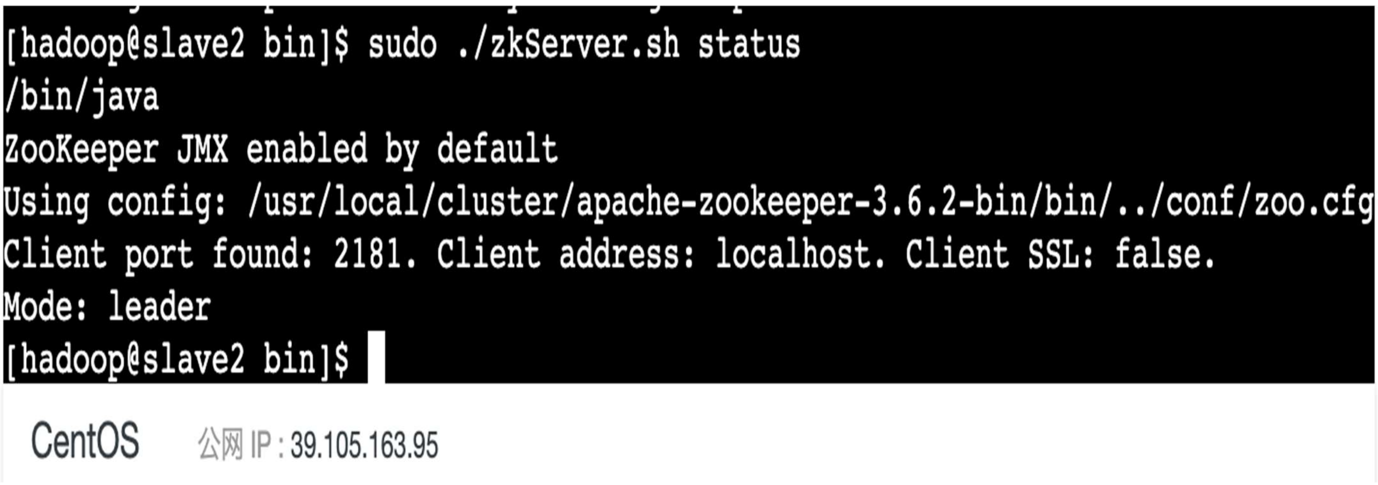
正常启动：

退出保存，启动zookeeper即可。

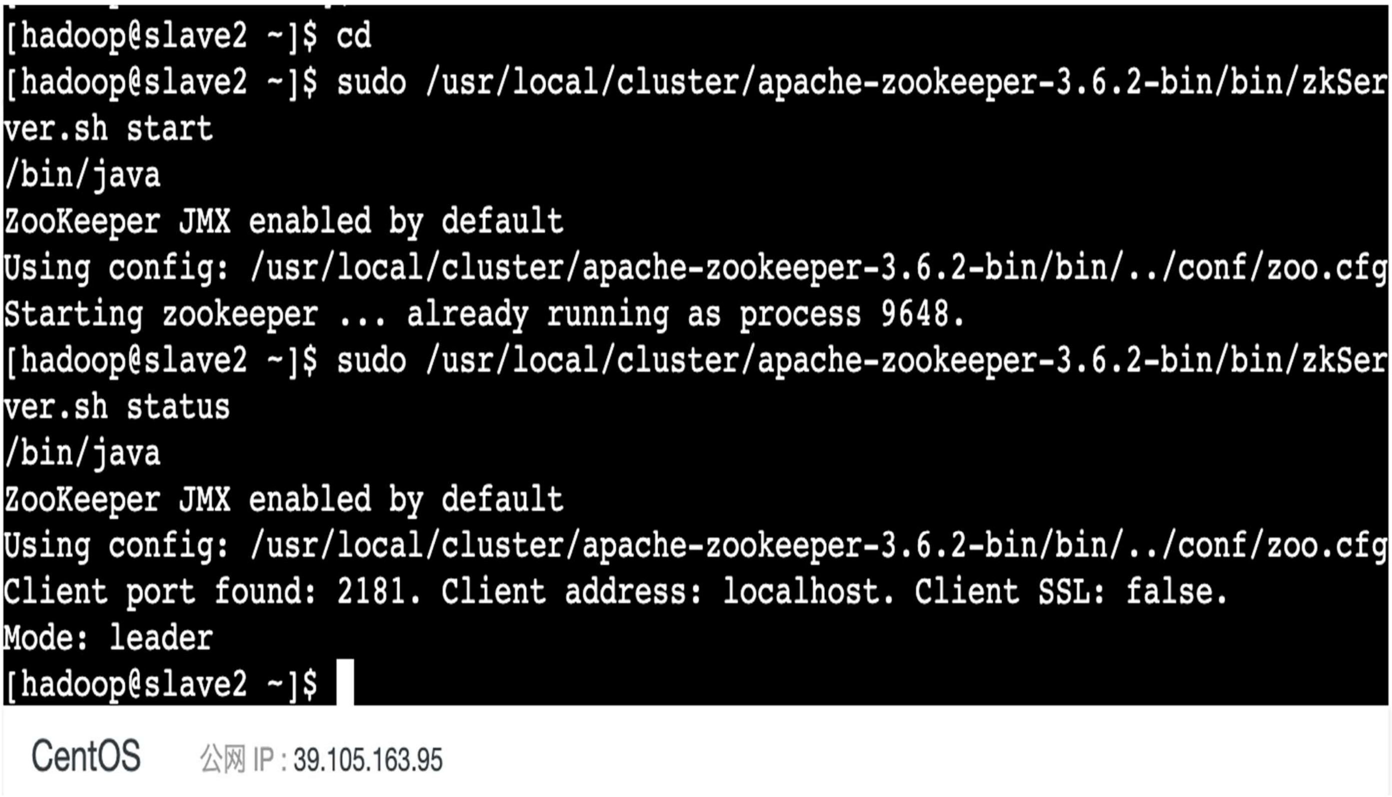
10.查看各个主机的状态。



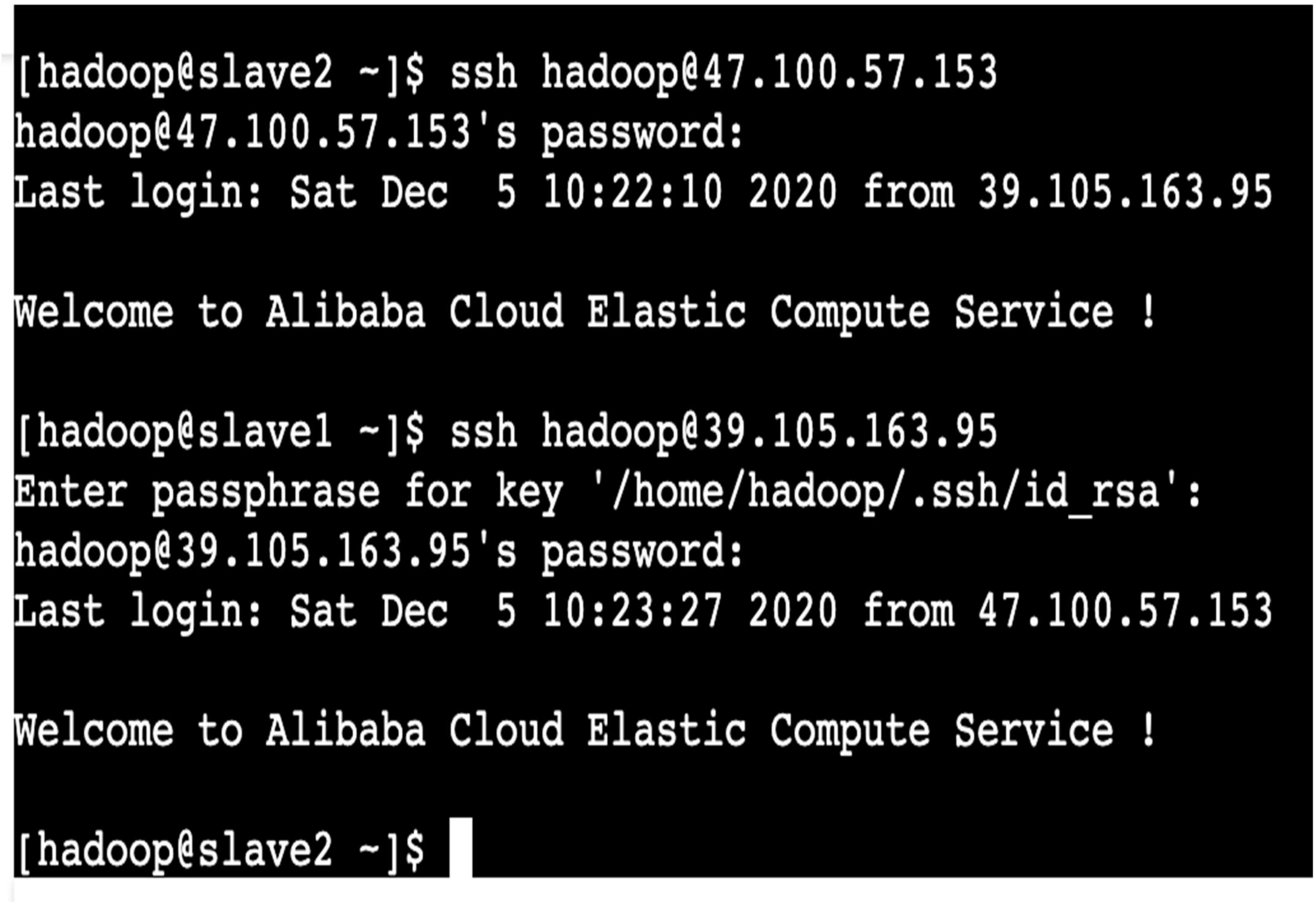
```
[hadoop@slave2 bin]$ ls
README.txt    zkEnv.sh                 zkSnapShotToolkit.sh
zkCleanup.sh  zkServer.cmd             zkTxnLogToolkit.cmd
zkCli.cmd     zkServer-initialize.sh   zkTxnLogToolkit.sh
zkCli.sh      zkServer.sh
zkEnv.cmd     zkSnapShotToolkit.cmd
[hadoop@slave2 bin]$ sudo ./zkServer.sh start
/bin/java
ZooKeeper JMX enabled by default
Using config: /usr/local/cluster/apache-zookeeper-3.6.2-bin/bin/../conf/zoo.cfg
Starting zookeeper ... already running as process 9648.
[hadoop@slave2 bin]$
```



```
[hadoop@slave2 bin]$ sudo ./zkServer.sh status
/bin/java
ZooKeeper JMX enabled by default
Using config: /usr/local/cluster/apache-zookeeper-3.6.2-bin/bin/../conf/zoo.cfg
Client port found: 2181. Client address: localhost. Client SSL: false.
Mode: leader
[hadoop@slave2 bin]$
```



```
[hadoop@slave2 ~]$ cd
[hadoop@slave2 ~]$ sudo /usr/local/cluster/apache-zookeeper-3.6.2-bin/bin/zkSer
ver.sh start
/bin/java
ZooKeeper JMX enabled by default
Using config: /usr/local/cluster/apache-zookeeper-3.6.2-bin/bin/../conf/zoo.cfg
Starting zookeeper ... already running as process 9648.
[hadoop@slave2 ~]$ sudo /usr/local/cluster/apache-zookeeper-3.6.2-bin/bin/zkSer
ver.sh status
/bin/java
ZooKeeper JMX enabled by default
Using config: /usr/local/cluster/apache-zookeeper-3.6.2-bin/bin/../conf/zoo.cfg
Client port found: 2181. Client address: localhost. Client SSL: false.
Mode: leader
[hadoop@slave2 ~]$
```



```
[hadoop@slave2 ~]$ ssh hadoop@47.100.57.153
hadoop@47.100.57.153's password:
Last login: Sat Dec  5 10:22:10 2020 from 39.105.163.95

Welcome to Alibaba Cloud Elastic Compute Service !

[hadoop@slave1 ~]$ ssh hadoop@39.105.163.95
Enter passphrase for key '/home/hadoop/.ssh/id_rsa':
hadoop@39.105.163.95's password:
Last login: Sat Dec  5 10:23:27 2020 from 47.100.57.153

Welcome to Alibaba Cloud Elastic Compute Service !

[hadoop@slave2 ~]$
```

Hadoop

1.下载Hadoop。

2.将Hadoop下载到本地后，通过scp命令将安装包发送到master上。

3.进入服务器根目录下，将hadoop-2.10.0.tar.gz解压至 /usr/local/cluster中。

4.进入 /usr/local/cluster 中，为了方便日后版本的更新，这里使用软链接的方法。

5.设置环境变量，在 ~/.bashrc 添加如下内容。

# hadoop export HADOOP_HOME=/usr/local/cluster/hadoop export PATH=$PATH:$HADOOP_HOME/bin:$HADOOP_HOME/sbin

6.使环境变量立刻生效。

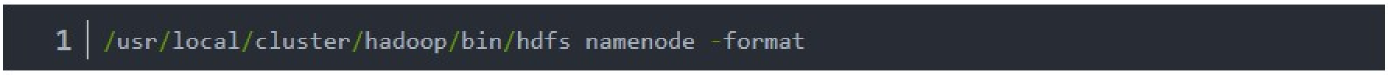
7.验证Hadoop是否安装成功。如果输出hadoop版本信息即安装成功。

8.配置hadoop。

** • **进入hadoop的配置目录。

** • **新建几个文件夹，配置文件中需要用到。

配置文件修改完以后，将master下hadoop文件夹复制到slave1和slave2中。

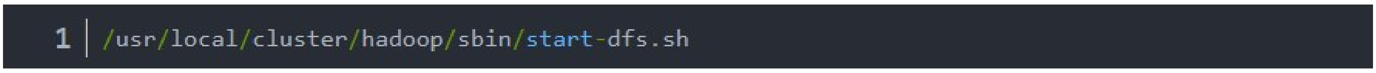


```
/usr/local/cluster/hadoop/bin/hdfs namenode -format
```

9.运行Hadoop。注意：启动hadoop的命令都只在master上执行。

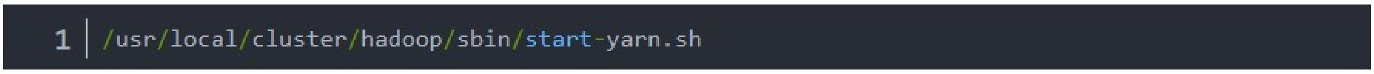
** • **启动namenode，如果是第一次启动namenode，需要对namenode进行格式化。命令如下：

启动hdfs：



```
/usr/local/cluster/hadoop/sbin/start-dfs.sh
```

验证hdfs是否启动成功。

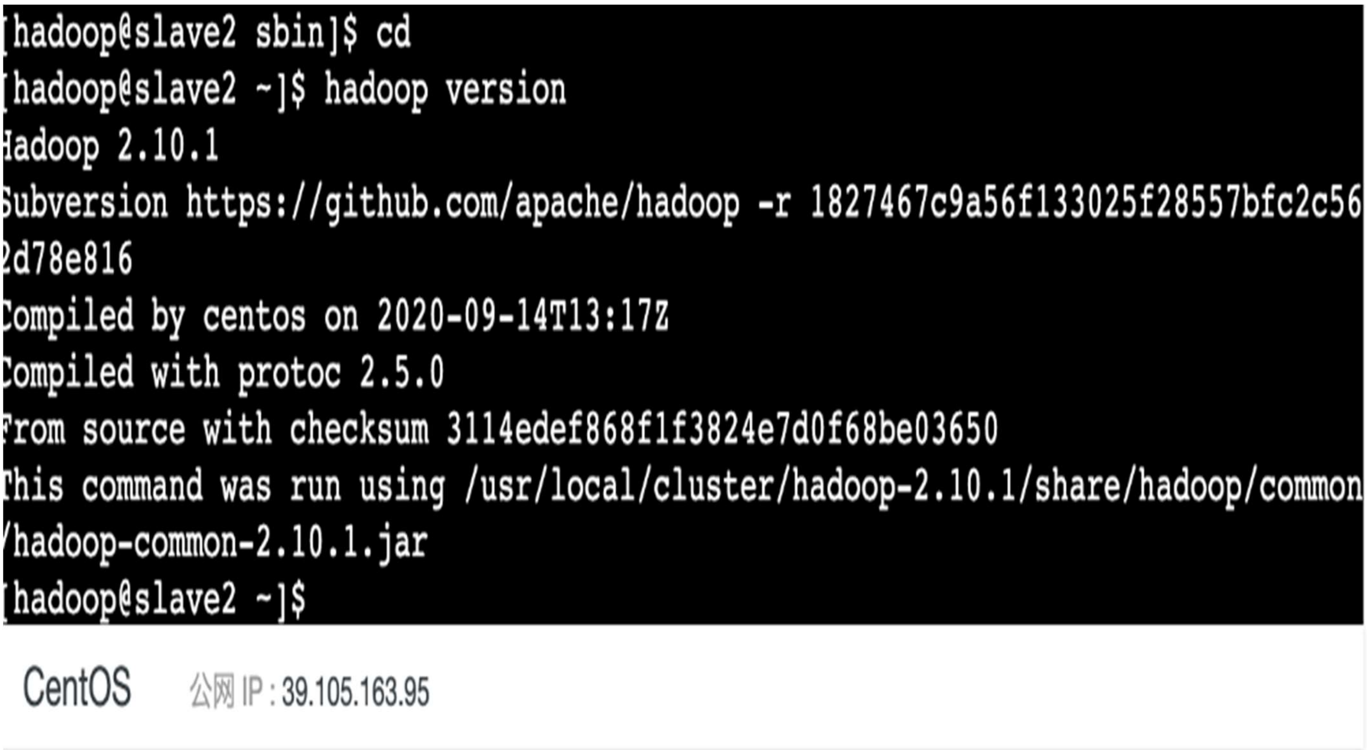


```
/usr/local/cluster/hadoop/sbin/start-yarn.sh
```

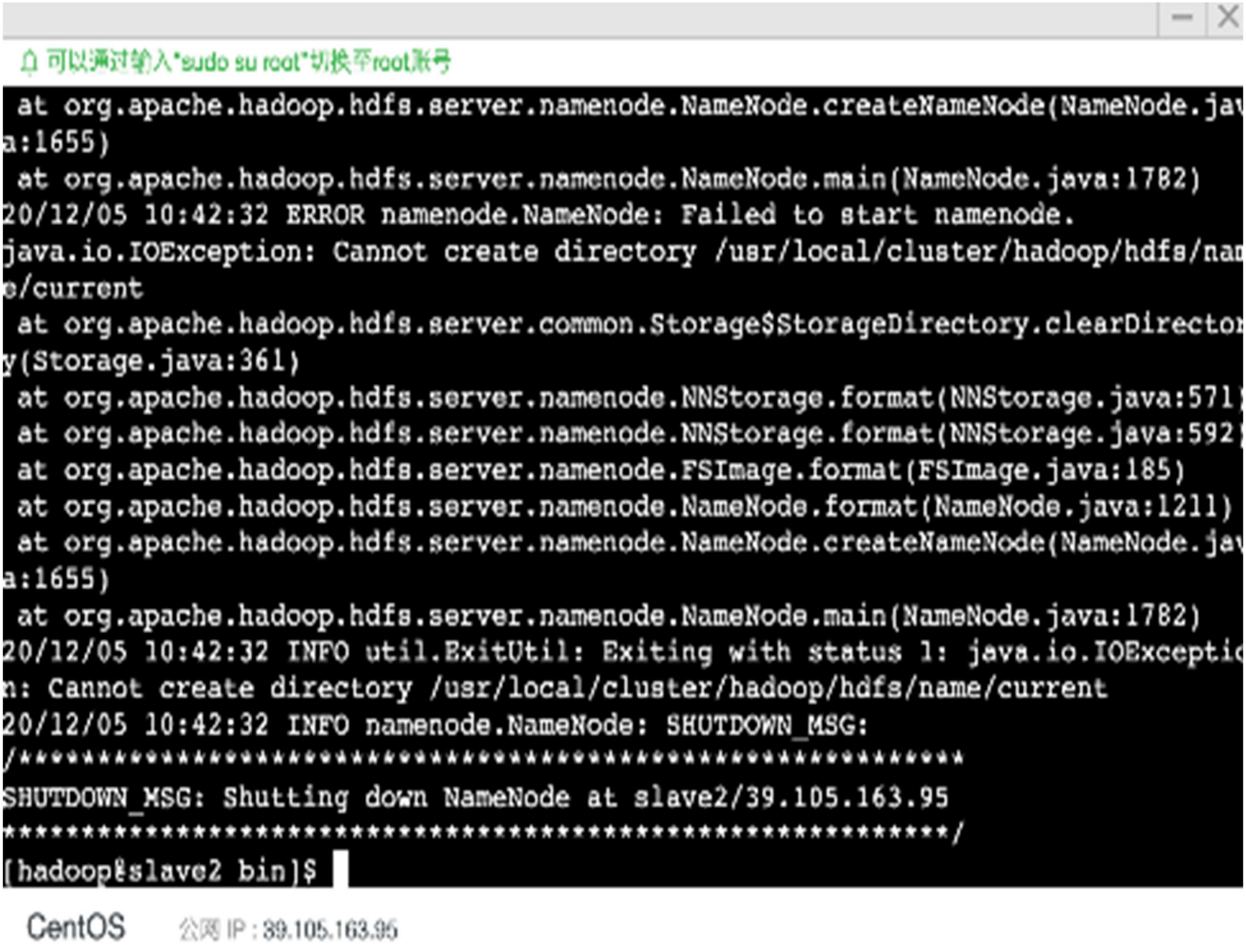
访问：

http://xxx.xxx.xxx.xxx:50070/dfshealth.html#tab-overview

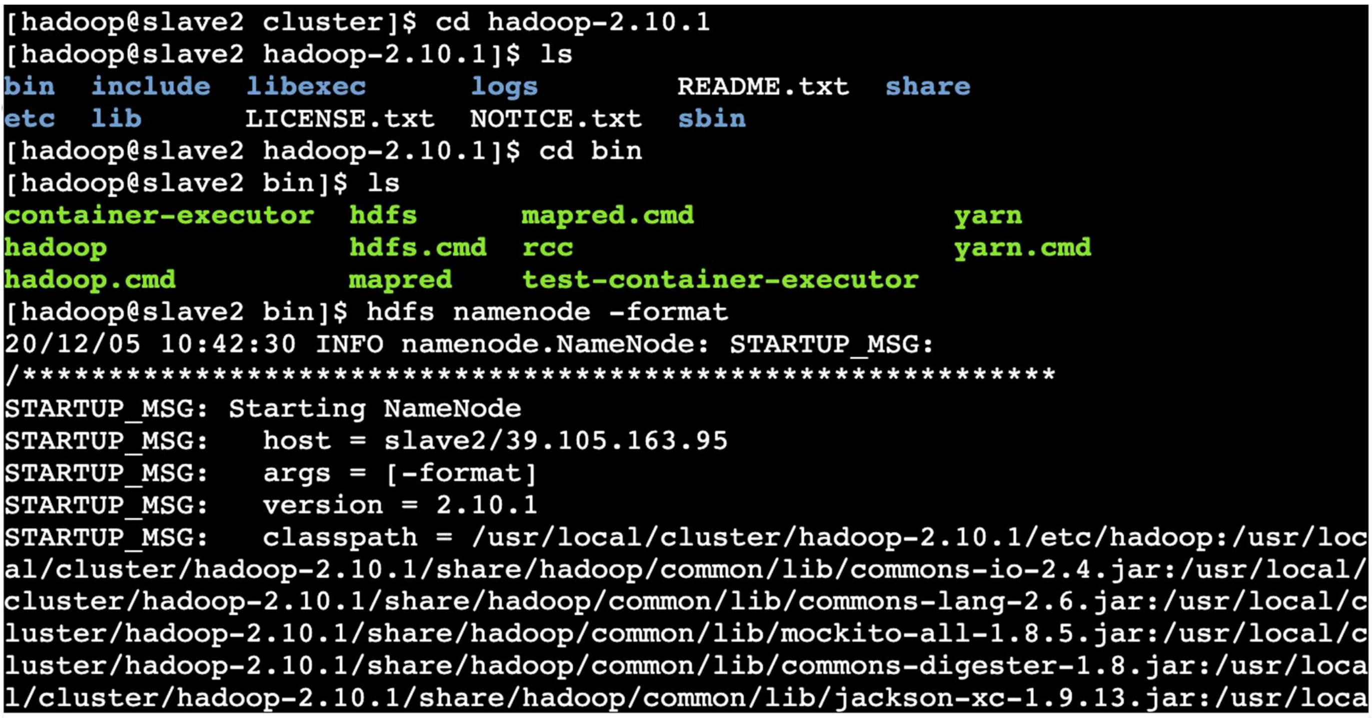
** • **启动yarn



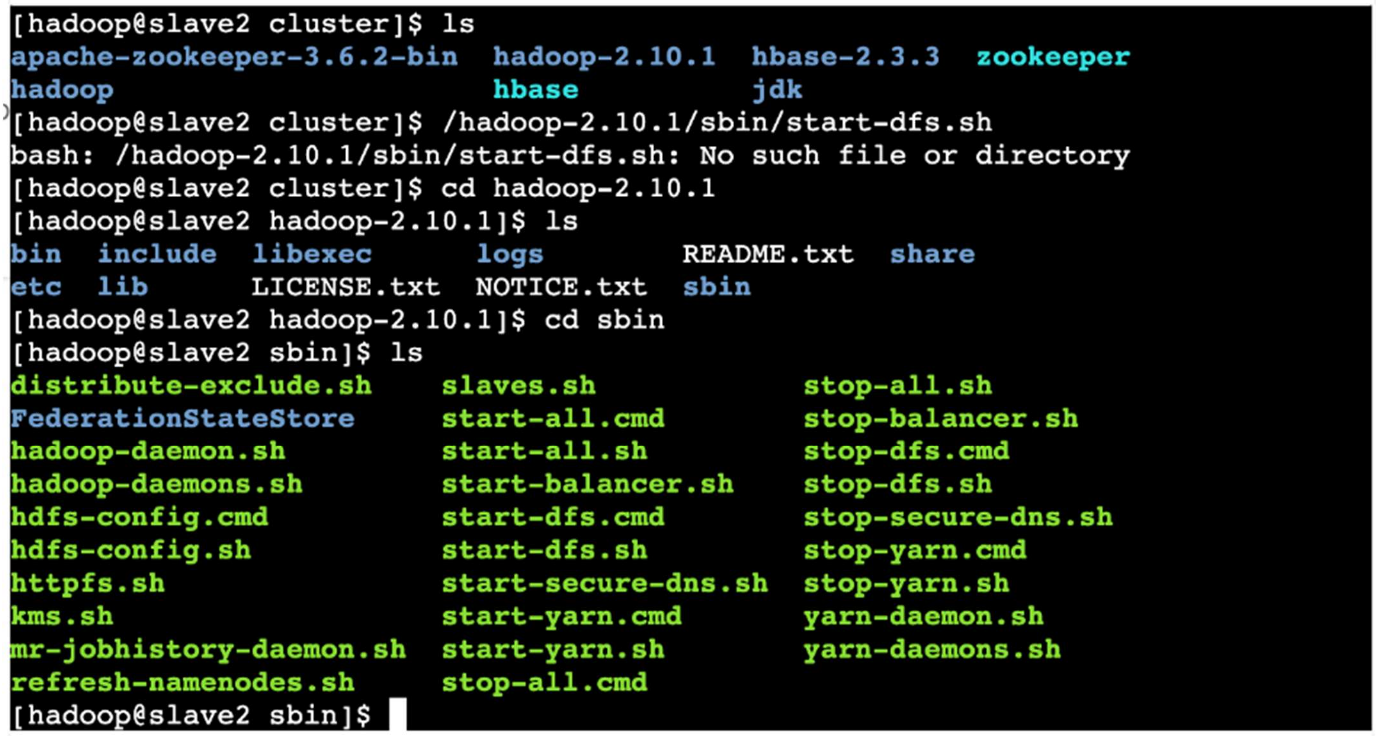
```
[hadoop@slave2 sbin]$ cd
[hadoop@slave2 ~]$ hadoop version
Hadoop 2.10.1
Subversion https://github.com/apache/hadoop -r 1827467c9a56f133025f28557bfc2c56
2d78e816
Compiled by centos on 2020-09-14T13:17Z
Compiled with protoc 2.5.0
From source with checksum 3114edef868f1f3824e7d0f68be03650
This command was run using /usr/local/cluster/hadoop-2.10.1/share/hadoop/common
/hadoop-common-2.10.1.jar
[hadoop@slave2 ~]$
```



```
 at org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.jav
a:1655)
 at org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1782)
20/12/05 10:42:32 ERROR namenode.NameNode: Failed to start namenode.
java.io.IOException: Cannot create directory /usr/local/cluster/hadoop/hdfs/nam
e/current
 at org.apache.hadoop.hdfs.server.common.Storage$StorageDirectory.clearDirector
y(Storage.java:361)
 at org.apache.hadoop.hdfs.server.namenode.NNStorage.format(NNStorage.java:571)
 at org.apache.hadoop.hdfs.server.namenode.NNStorage.format(NNStorage.java:592)
 at org.apache.hadoop.hdfs.server.namenode.FSImage.format(FSImage.java:185)
 at org.apache.hadoop.hdfs.server.namenode.NameNode.format(NameNode.java:1211)
 at org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.jav
a:1655)
 at org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1782)
20/12/05 10:42:32 INFO util.ExitUtil: Exiting with status 1: java.io.IOExceptio
n: Cannot create directory /usr/local/cluster/hadoop/hdfs/name/current
20/12/05 10:42:32 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at slave2/39.105.163.95
************************************************************/
[hadoop@slave2 bin]$
```



```
[hadoop@slave2 cluster]$ cd hadoop-2.10.1
[hadoop@slave2 hadoop-2.10.1]$ ls
bin  include  libexec      logs        README.txt  share
etc  lib      LICENSE.txt  NOTICE.txt  sbin
[hadoop@slave2 hadoop-2.10.1]$ cd bin
[hadoop@slave2 bin]$ ls
container-executor  hdfs      mapred.cmd                yarn
hadoop              hdfs.cmd  rcc                       yarn.cmd
hadoop.cmd          mapred    test-container-executor
[hadoop@slave2 bin]$ hdfs namenode -format
20/12/05 10:42:30 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
STARTUP_MSG: Starting NameNode
STARTUP_MSG:   host = slave2/39.105.163.95
STARTUP_MSG:   args = [-format]
STARTUP_MSG:   version = 2.10.1
STARTUP_MSG:   classpath = /usr/local/cluster/hadoop-2.10.1/etc/hadoop:/usr/loc
al/cluster/hadoop-2.10.1/share/hadoop/common/lib/commons-io-2.4.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/common/lib/commons-lang-2.6.jar:/usr/local/c
luster/hadoop-2.10.1/share/hadoop/common/lib/mockito-all-1.8.5.jar:/usr/local/c
luster/hadoop-2.10.1/share/hadoop/common/lib/commons-digester-1.8.jar:/usr/loca
l/cluster/hadoop-2.10.1/share/hadoop/common/lib/jackson-xc-1.9.13.jar:/usr/loca
```



```
[hadoop@slave2 cluster]$ ls
apache-zookeeper-3.6.2-bin  hadoop-2.10.1  hbase-2.3.3  zookeeper
hadoop                      hbase          jdk
[hadoop@slave2 cluster]$ /hadoop-2.10.1/sbin/start-dfs.sh
bash: /hadoop-2.10.1/sbin/start-dfs.sh: No such file or directory
[hadoop@slave2 cluster]$ cd hadoop-2.10.1
[hadoop@slave2 hadoop-2.10.1]$ ls
bin  include  libexec      logs        README.txt  share
etc  lib      LICENSE.txt  NOTICE.txt  sbin
[hadoop@slave2 hadoop-2.10.1]$ cd sbin
[hadoop@slave2 sbin]$ ls
distribute-exclude.sh    slaves.sh             stop-all.sh
FederationStateStore     start-all.cmd         stop-balancer.sh
hadoop-daemon.sh         start-all.sh          stop-dfs.cmd
hadoop-daemons.sh        start-balancer.sh     stop-dfs.sh
hdfs-config.cmd          start-dfs.cmd         stop-secure-dns.sh
hdfs-config.sh           start-dfs.sh          stop-yarn.cmd
httpfs.sh                start-secure-dns.sh   stop-yarn.sh
kms.sh                   start-yarn.cmd        yarn-daemon.sh
mr-jobhistory-daemon.sh  start-yarn.sh         yarn-daemons.sh
refresh-namenodes.sh     stop-all.cmd
[hadoop@slave2 sbin]$ 
```

HBase

1. 下载HBase。

2. 将HBase下载到本地后，通过scp命令将安装包发送到master主机上。

3. 进入服务器根目录下，将hbase-1.3.6-bin.tar.gz解压至 /usr/local/cluster中。

4. 进入 /usr/local/cluster 中，为了方便日后版本的更新，这里使用软链接的方法。

5.设置环境变量，在 ~/.bashrc 添加如下内容。

# hbase export HBASE_HOME=/usr/local/cluster/hbase export PATH=$PATH:$HBASE_HOME/bin

6.使环境变量立刻生效。

7.验证Hbase是否安装成功。

如果输出hadoop版本信息即安装成功。

8.配置HBase

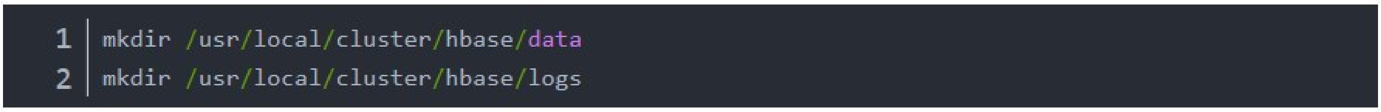
主要修改conf目录下的三个文件：hbase-env.sh、hbase-site.xml、regionservers。

** • **hbase-env.sh

export JAVA_HOME=/usr/local/cluster/java

export HBASE_CLASSPATH=/usr/local/cluster/hbase/lib

export HBASE_PID_DIR=/usr/local/cluster/hbase/data



```
mkdir /usr/local/cluster/hbase/data
mkdir /usr/local/cluster/hbase/logs
```

export HBASE_LOG_DIR=/usr/local/cluster/hbase/logs

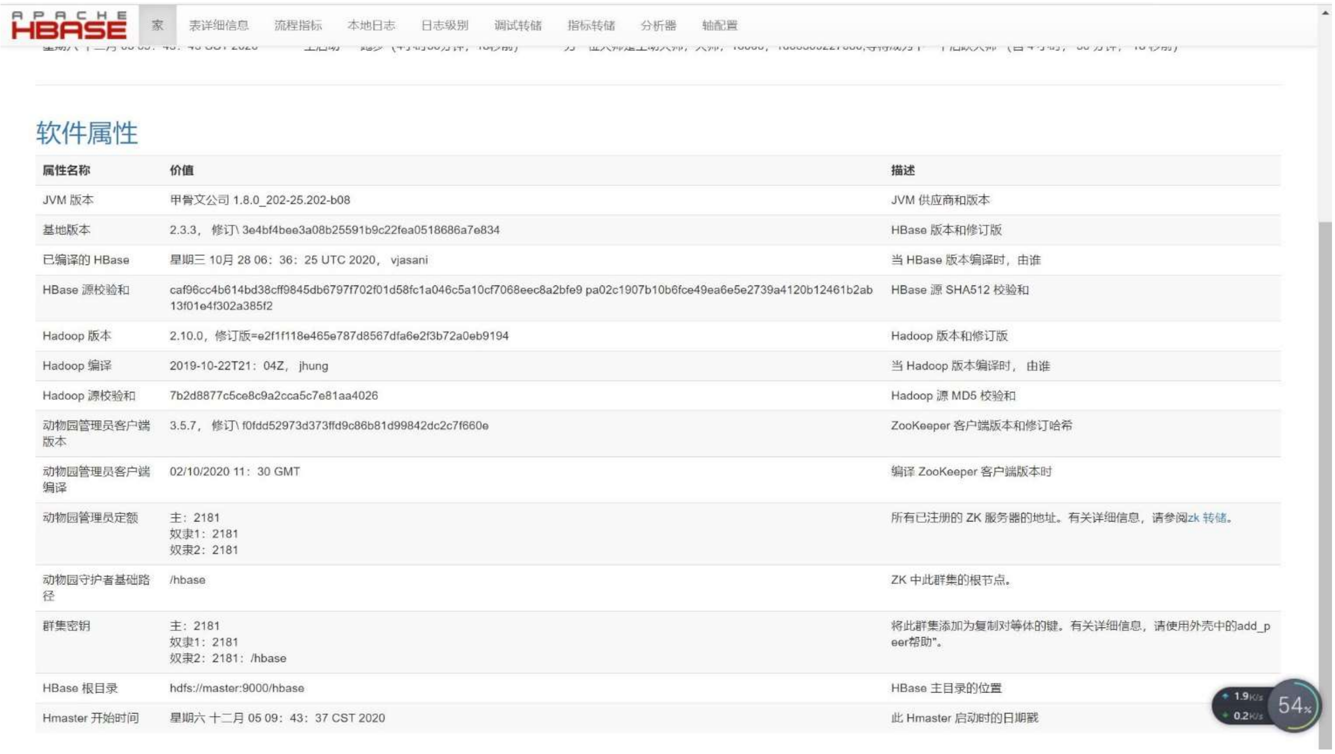
export HBASE_MANAGES_ZK=false

注意：要在hbase文件下，新建data和logs两个文件夹

9.启动hbase

10.验证hbase是否启动成功。

访问：http://xxx.xxx.xxx.xxx:16010/master-status



软件属性

属性名称	价值	描述
JVM 版本	甲骨文公司 1.8.0_202-25.202-b08	JVM 供应商和版本
基地版本	2.3.3， 修订\ 3e4bf4bee3a08b25591b9c22fea0518686a7e834	HBase 版本和修订版
已编译的 HBase	星期三 10月 28 06：36：25 UTC 2020， vjasani	当 HBase 版本编译时，由谁
HBase 源校验和	caf96cc4b614bd38cff9845db6797f702f01d58fc1a046c5a10cf7068eec8a2bfe9 pa02c1907b10b6fce49ea6e5e2739a4120b12461b2ab 13f01e4f302a385f2	HBase 源 SHA512 校验和
Hadoop 版本	2.10.0，修订版=e2f1f118e465e787d8567dfa6e2f3b72a0eb9194	Hadoop 版本和修订版
Hadoop 编译	2019-10-22T21：04Z， jhung	当 Hadoop 版本编译时， 由谁
Hadoop 源校验和	7b2d8877c5ce8c9a2cca5c7e81aa4026	Hadoop 源 MD5 校验和
动物园管理员客户端版本	3.5.7， 修订\ f0fdd52973d373ffd9c86b81d99842dc2c7f660e	ZooKeeper 客户端版本和修订哈希
动物园管理员客户端编译	02/10/2020 11：30 GMT	编译 ZooKeeper 客户端版本时
动物园管理员定额	主：2181 奴隶1：2181 奴隶2：2181	所有已注册的 ZK 服务器的地址。有关详细信息，请参阅zk 转储。
动物园守护者基础路径	/hbase	ZK 中此群集的根节点。
群集密钥	主：2181 奴隶1：2181 奴隶2：2181：/hbase	将此群集添加为复制对等体的键。有关详细信息，请使用外壳中的add_peer帮助"。
HBase 根目录	hdfs://master:9000/hbase	HBase 主目录的位置
Hmaster 开始时间	星期六 十二月 05 09：43：37 CST 2020	此 Hmaster 启动时的日期戳

Hadoopmaster的启动

```
20/12/05 10:42:30 INFO namenode.NameNode: STARTUP_MSG:
/************************************************************
** STARTUP_MSG: Starting NameNode STARTUP_MSG: host =
slave2/39.105.163.95
STARTUP_MSG: args = [-format] STARTUP_MSG: version =
2.10.1
STARTUP_MSG: classpath = /usr/local/cluster/hadoop-2.10.1/etc/hadoop:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/lib/commons-io-2.4.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/common/lib/commons-lang-2.6.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
common/lib/mockito-all-1.8.5.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/
commons-digester-1.8.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jackson- xc-
1.9.13.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jetty-
sslengine-6.1.26.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/servlet-
api-2.5.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/audience-
annotations-0.5.0.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/hamcrest-
core-1.3.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/log4j-1.2.17.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/common/lib/jackson-core-asl-1.9.13.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/lib/gson-2.2.4.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/common/lib/commons-logging-1.1.3.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
common/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/common/lib/commons-cli-1.2.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
common/lib/commons-collections-3.2.2.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
common/lib/asm-3.2.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/slf4j-
api-1.7.25.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jackson-
jaxrs-1.9.13.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jersey-server-1.9.jar:/
usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jersey-core-1.9.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/lib/hadoop-auth-2.10.1.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/lib/protobuf-java-2.5.0.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/lib/paranamer-2.3.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/common/lib/jackson-mapper-asl-1.9.13.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/common/lib/commons-codec-1.4.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
common/lib/jettison-1.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/
commons-compress-1.19.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jetty- util-
6.1.26.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/activation-1.1.jar:/usr/
local/cluster/hadoop-2.10.1/share/hadoop/common/lib/avro-1.7.7.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/lib/jets3t-0.9.0.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/common/lib/hadoop-annotations-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/common/lib/commons-configuration-1.6.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/common/lib/slf4j-log4j12-1.7.25.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
common/lib/junit-4.11.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/
httpclient-4.5.2.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jsp-api-
2.1.jar:/                    usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/guava-
11.0.2.jar:/usr/local/cluster/         hadoop-2.10.1/share/hadoop/common/lib/commons-lang3-
3.4.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/lib/jsch-0.1.55.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/common/lib/httpcore-4.4.4.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/
lib/spotbugs-annotations-3.1.9.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/
json-smart-1.3.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/curator-
framework-2.13.0.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/apacheds- i18n-
2.0.0-M15.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jetty-6.1.26.jar:/
usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/java-xmlbuilder-0.4.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/common/lib/jersey-json-1.9.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/lib/stax2-api-3.1.4.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/common/lib/commons-beanutils-1.9.4.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/common/lib/snappy-java-1.0.5.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
common/lib/stax-api-1.0-2.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/htrace-
core4-4.1.0-incubating.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/
netty-3.10.6.Final.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/api-util-1.0.0-
M20.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/curator-recipes-2.13.0.jar:/
```

```
usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/zookeeper-3.4.14.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/common/lib/nimbus-jose-jwt-7.9.jar:/usr/local/cluster/


hadoop-2.10.1/share/hadoop/common/lib/xmlenc-0.52.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/common/lib/jcip-annotations-1.0-1.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/common/lib/commons-math3-3.1.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
common/lib/api-asn1-api-1.0.0-M20.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/
lib/jaxb-api-2.2.2.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jaxb-
impl-2.2.3-1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/commons-
net-3.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/common/lib/jsr305-3.0.2.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/common/lib/woodstox-core-5.0.3.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/lib/curator-client-2.13.0.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/common/hadoop-common-2.10.1-tests.jar:/usr/local/cluster/ hadoop-
2.10.1/share/hadoop/common/hadoop-nfs-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/common/hadoop-common-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/hdfs:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/commons-io-2.4.jar:/usr/
local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/commons-lang-2.6.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/hdfs/lib/xercesImpl-2.12.0.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/hdfs/lib/servlet-api-2.5.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/
netty-all-4.1.50.Final.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/log4j-
1.2.17.jar:/ usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/jackson-core-asl-
1.9.13.jar:/usr/local/ cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/commons-logging-
1.1.3.jar:/usr/local/cluster/ hadoop-2.10.1/share/hadoop/hdfs/lib/jackson-databind-
2.9.10.6.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/hdfs/lib/commons-cli-1.2.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/hdfs/lib/asm-3.2.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/
okio-1.6.0.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/leveldbjni-all-
1.8.jar:/usr/ local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/okhttp-
2.7.5.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/hdfs/lib/jersey-server-1.9.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/hdfs/lib/jersey-core-1.9.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/
protobuf-java-2.5.0.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/jackson-mapper-
asl-1.9.13.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/commons-codec-1.4.jar:/
usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/jackson-annotations-2.9.10.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/jetty-util-6.1.26.jar:/usr/local/cluster/hadoop-
2.10.1/ share/hadoop/hdfs/lib/guava-11.0.2.jar:/usr/local/cluster/hadoop-
2.10.1/share/hadoop/hdfs/lib/ jetty-6.1.26.jar:/usr/local/cluster/hadoop-
2.10.1/share/hadoop/hdfs/lib/jackson-core-2.9.10.jar:/ usr/local/cluster/hadoop-
2.10.1/share/hadoop/hdfs/lib/htrace-core4-4.1.0-incubating.jar:/usr/ local/cluster/hadoop-
2.10.1/share/hadoop/hdfs/lib/netty-3.10.6.Final.jar:/usr/local/cluster/ hadoop-
2.10.1/share/hadoop/hdfs/lib/hadoop-hdfs-client-2.10.1.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/hdfs/lib/commons-daemon-1.0.13.jar:/usr/local/cluster/ hadoop-
2.10.1/share/hadoop/hdfs/lib/xmlenc-0.52.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/hdfs/lib/xml-apis-1.4.01.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/lib/
jsr305-3.0.2.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/hadoop-hdfs-native-
client-2.10.1-tests.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/hadoop-hdfs-native-
client-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/hadoop-hdfs-rbf-2.10.1-
tests.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/hadoop-hdfs-2.10.1-tests.jar:/usr/
local/cluster/hadoop-2.10.1/share/hadoop/hdfs/hadoop-hdfs-rbf-
2.10.1.jar:/usr/local/cluster/ hadoop-2.10.1/share/hadoop/hdfs/hadoop-hdfs-client-
2.10.1.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/hdfs/hadoop-hdfs-nfs-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/hdfs/hadoop-hdfs-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/hdfs/
hadoop-hdfs-client-2.10.1-tests.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn:/usr/
local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jersey-guice-1.9.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/commons-io-2.4.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/yarn/lib/commons-lang-2.6.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/
commons-digester-1.8.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jackson-
xc-1.9.13.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jetty-sslengine-
```

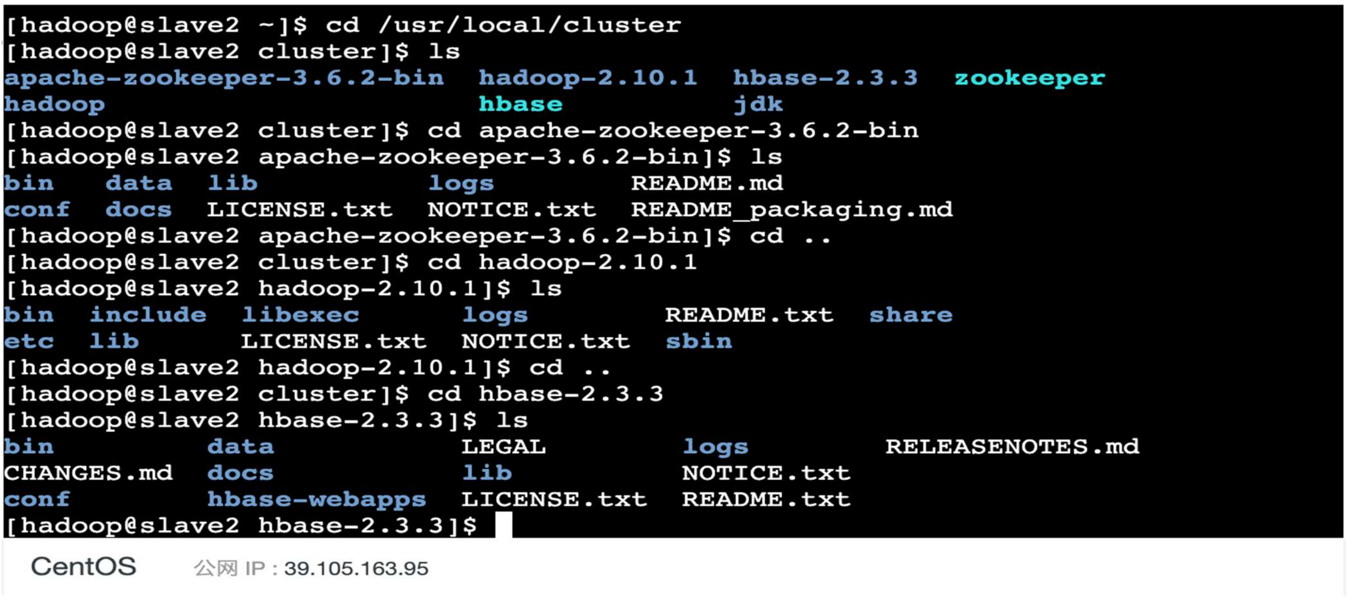
```
6.1.26.jar:/ usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/ehcache-
3.3.1.jar:/usr/local/cluster/ hadoop-2.10.1/share/hadoop/yarn/lib/servlet-api-
2.5.jar:/usr/local/cluster/hadoop-2.10.1/share/ hadoop/yarn/lib/audience-annotations-
0.5.0.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/ yarn/lib/log4j-
1.2.17.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jackson-core- asl-
1.9.13.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/gson-2.2.4.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/yarn/lib/commons-logging-1.1.3.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/apacheds-kerberos-codec-2.0.0-M15.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/commons-cli-1.2.jar:/usr/local/cluster/hadoop-2.10.1/

share/hadoop/yarn/lib/guice-3.0.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/
fst-2.50.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/commons-
collections-3.2.2.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/asm-
3.2.jar:/usr/local/ cluster/hadoop-2.10.1/share/hadoop/yarn/lib/leveldbjni-all-
1.8.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/jackson-jaxrs-1.9.13.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/yarn/lib/jersey-client-1.9.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/
lib/jersey-server-1.9.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jersey-
core-1.9.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/protobuf-java-
2.5.0.jar:/usr/ local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/paranamer-
2.3.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/jackson-mapper-asl-1.9.13.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/geronimo-jcache_1.0_spec-1.0-alpha-1.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/yarn/lib/json-io-2.5.1.jar:/usr/local/cluster/hadoop-
2.10.1/ share/hadoop/yarn/lib/commons-codec-1.4.jar:/usr/local/cluster/hadoop-
2.10.1/share/hadoop/ yarn/lib/jettison-1.1.jar:/usr/local/cluster/hadoop-
2.10.1/share/hadoop/yarn/lib/commons- compress-1.19.jar:/usr/local/cluster/hadoop-
2.10.1/share/hadoop/yarn/lib/jetty-util-6.1.26.jar:/usr/ local/cluster/hadoop-
2.10.1/share/hadoop/yarn/lib/activation-1.1.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/aopalliance-1.0.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/yarn/lib/avro-1.7.7.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/
jets3t-0.9.0.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/commons-
configuration-1.6.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/HikariCP-
java7-2.4.12.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/httpclient-
4.5.2.jar:/usr/ local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jsp-api-
2.1.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/guava-11.0.2.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/yarn/lib/commons-lang3-3.4.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/
javax.inject-1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jsch-
0.1.55.jar:/usr/local/ cluster/hadoop-2.10.1/share/hadoop/yarn/lib/httpcore-
4.4.4.jar:/usr/local/cluster/hadoop-2.10.1/ share/hadoop/yarn/lib/spotbugs-annotations-
3.1.9.jar:/usr/local/cluster/hadoop-2.10.1/share/ hadoop/yarn/lib/json-smart-
1.3.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/ curator-framework-
2.13.0.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/apacheds- i18n-2.0.0-
M15.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jetty-6.1.26.jar:/usr/
local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/java-xmlbuilder-0.4.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/jersey-json-1.9.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/yarn/lib/stax2-api-3.1.4.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/
commons-beanutils-1.9.4.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/snappy-
java-1.0.5.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/mssql-jdbc-6.2.1.jre7.jar:/
usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/stax-api-1.0-2.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/metrics-core-3.0.1.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/yarn/lib/htrace-core4-4.1.0-incubating.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/yarn/lib/netty-3.10.6.Final.jar:/usr/local/cluster/hadoop-
2.10.1/share/hadoop/yarn/lib/ guice-servlet-3.0.jar:/usr/local/cluster/hadoop-
2.10.1/share/hadoop/yarn/lib/api-util-1.0.0- M20.jar:/usr/local/cluster/hadoop-
2.10.1/share/hadoop/yarn/lib/curator-recipes-2.13.0.jar:/usr/ local/cluster/hadoop-
2.10.1/share/hadoop/yarn/lib/zookeeper-3.4.14.jar:/usr/local/cluster/ hadoop-
2.10.1/share/hadoop/yarn/lib/nimbus-jose-jwt-7.9.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/yarn/lib/xmlenc-0.52.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/
```

```
jcip-annotations-1.0-1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/commons-
math3-3.1.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/api-asn1-api-1.0.0-
M20.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/java-util-
1.9.0.jar:/usr/local/ cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jaxb-api-
2.2.2.jar:/usr/local/cluster/hadoop-2.10.1/ share/hadoop/yarn/lib/jaxb-impl-2.2.3-
1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/ lib/commons-net-
3.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/jsr305-3.0.2.jar:/
usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/lib/woodstox-core-5.0.3.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/lib/curator-client-2.13.0.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/yarn/hadoop-yarn-server-nodemanager-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/yarn/hadoop-yarn-server-tests-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/yarn/hadoop-yarn-server-sharedcachemanager-2.10.1.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-applications-distributedshell-2.10.1.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-client-2.10.1.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-server-web-proxy-2.10.1.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-server-applicationhistoryservice-2.10.1.jar:/usr/
local/cluster/hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-server-timeline-

pluginstorage-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-
registry-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-
api-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-server-
common-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-
common-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/yarn/hadoop-yarn-
applications-unmanaged-am-launcher-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
yarn/hadoop-yarn-server-resourcemanager-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/yarn/hadoop-yarn-server-router-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/mapreduce/lib/jersey-guice-1.9.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
mapreduce/lib/commons-io-2.4.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/
lib/hamcrest-core-1.3.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/lib/
log4j-1.2.17.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/lib/jackson-core-
asl-1.9.13.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/lib/guice-3.0.jar:/usr/
local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/lib/asm-3.2.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/mapreduce/lib/leveldbjni-all-1.8.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/mapreduce/lib/jersey-server-1.9.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/mapreduce/lib/jersey-core-1.9.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/mapreduce/lib/protobuf-java-2.5.0.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/mapreduce/lib/paranamer-2.3.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
mapreduce/lib/jackson-mapper-asl-1.9.13.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
mapreduce/lib/commons-compress-1.19.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
mapreduce/lib/aopalliance-1.0.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/lib/
avro-1.7.7.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/lib/hadoop-
annotations-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/lib/
junit-4.11.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/lib/javax.inject-
1.jar:/usr/                  local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/lib/snappy-java-
1.0.5.jar:/usr/local/cluster/                  hadoop-2.10.1/share/hadoop/mapreduce/lib/netty-
3.10.6.Final.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/mapreduce/lib/guice-servlet-3.0.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/mapreduce/hadoop-mapreduce-client-hs-plugins-2.10.1.jar:/usr/
local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/hadoop-mapreduce-client-
jobclient-2.10.1-tests.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/hadoop-
mapreduce-examples-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/mapreduce/
hadoop-mapreduce-client-jobclient-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/hadoop/
mapreduce/hadoop-mapreduce-client-hs-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/share/
hadoop/mapreduce/hadoop-mapreduce-client-shuffle-2.10.1.jar:/usr/local/cluster/hadoop-2.10.1/
share/hadoop/mapreduce/hadoop-mapreduce-client-app-2.10.1.jar:/usr/local/cluster/
hadoop-2.10.1/share/hadoop/mapreduce/hadoop-mapreduce-client-core-2.10.1.jar:/usr/local/
cluster/hadoop-2.10.1/share/hadoop/mapreduce/hadoop-mapreduce-client-common-2.10.1.jar:/
usr/local/cluster/hadoop/contrib/capacity-scheduler/*.jar
STARTUP_MSG: build = https://github.com/apache/hadoop -r 1827467c9a56f133025f28557bfc2c562d78e816;
compiled by 'centos' on 2020-09-14T13:17Z STARTUP_MSG: java = 1.8.0_202
```

```
************************************************************/
20/12/05 10:42:30 INFO namenode.NameNode: registered UNIX signal handlers for [TERM, HUP, INT]
20/12/05 10:42:30 INFO namenode.NameNode: createNameNode [-format]
20/12/05 10:42:31 INFO common.Util: Assuming 'file' scheme for path /usr/local/cluster/hadoop/
hdfs/name in configuration.
20/12/05 10:42:31 INFO common.Util: Assuming 'file' scheme for path /usr/local/cluster/hadoop/
hdfs/name in configuration.
Formatting using clusterid: CID-1fb3843b-63a3-4871-badb-ef1aa60719d8
20/12/05 10:42:32 INFO namenode.FSEditLog: Edit logging is async:true
20/12/05 10:42:32 INFO namenode.FSNamesystem: KeyProvider: null 20/12/05
10:42:32 INFO namenode.FSNamesystem: fsLock is fair: true
20/12/05 10:42:32 INFO namenode.FSNamesystem: Detailed lock hold time metrics enabled: false
20/12/05 10:42:32 INFO namenode.FSNamesystem: fsOwner             = hadoop (auth:SIMPLE)
20/12/05 10:42:32 INFO namenode.FSNamesystem: supergroup           = supergroup
20/12/05 10:42:32 INFO namenode.FSNamesystem: isPermissionEnabled = false 20/12/05
10:42:32 INFO namenode.FSNamesystem: HA Enabled: false


20/12/05 10:42:32 INFO common.Util: dfs.datanode.fileio.profiling.sampling.percentage set to 0.
Disabling file IO profiling
20/12/05 10:42:32 INFO blockmanagement.DatanodeManager: dfs.block.invalidate.limit: configured=1000,
counted=60, effected=1000
20/12/05 10:42:32 INFO blockmanagement.DatanodeManager:
dfs.namenode.datanode.registration.ip-hostname-check=true
20/12/05 10:42:32 INFO blockmanagement.BlockManager:
dfs.namenode.startup.delay.block.deletion.sec is set to 000:00:00:00.000
20/12/05 10:42:32 INFO blockmanagement.BlockManager: The block deletion will start around 2020
Dec 05 10:42:32
20/12/05 10:42:32 INFO util.GSet: Computing capacity for map BlocksMap
20/12/05 10:42:32 INFO util.GSet: VM type    = 64-bit
20/12/05 10:42:32 INFO util.GSet: 2.0% max memory 966.7 MB = 19.3 MB
20/12/05 10:42:32 INFO util.GSet: capacity   = 2^21 = 2097152 entries
20/12/05 10:42:32 INFO blockmanagement.BlockManager: dfs.block.access.token.enable=false 20/12/05
10:42:32 WARN conf.Configuration: No unit for dfs.heartbeat.interval(3) assuming SECONDS
20/12/05 10:42:32 WARN conf.Configuration: No unit for
dfs.namenode.safemode.extension(30000) assuming MILLISECONDS 20/12/05
10:42:32 INFO blockmanagement.BlockManagerSafeMode:
dfs.namenode.safemode.threshold-pct = 0.9990000128746033 20/12/05
10:42:32 INFO blockmanagement.BlockManagerSafeMode:
dfs.namenode.safemode.min.datanodes = 0
20/12/05 10:42:32 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.extension =
30000
20/12/05 10:42:32 INFO blockmanagement.BlockManager: defaultReplication       = 3 20/12/05
10:42:32 INFO blockmanagement.BlockManager: maxReplication                  = 512 20/12/05
10:42:32 INFO blockmanagement.BlockManager: minReplication                 = 1 20/12/05
10:42:32 INFO blockmanagement.BlockManager: maxReplicationStreams           = 2 20/12/05
10:42:32 INFO blockmanagement.BlockManager: replicationRecheckInterval = 3000 20/12/05 10:42:32
INFO blockmanagement.BlockManager: encryptDataTransfer                    = false
20/12/05 10:42:32 INFO blockmanagement.BlockManager: maxNumBlocksToLog            = 1000
20/12/05 10:42:32 INFO namenode.FSNamesystem: Append Enabled: true
20/12/05 10:42:32 INFO namenode.FSDirectory: GLOBAL serial map: bits=24 maxEntries=16777215
20/12/05 10:42:32 INFO util.GSet: Computing capacity for map INodeMap
20/12/05 10:42:32 INFO util.GSet: VM type    = 64-bit
20/12/05 10:42:32 INFO util.GSet: 1.0% max memory 966.7 MB = 9.7 MB
20/12/05 10:42:32 INFO util.GSet: capacity   = 2^20 = 1048576 entries
20/12/05 10:42:32 INFO namenode.FSDirectory: ACLs enabled? false 20/12/05
10:42:32 INFO namenode.FSDirectory: XAttrs enabled? true
20/12/05 10:42:32 INFO namenode.NameNode: Caching file names occurring more than 10 times 20/12/05
10:42:32 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles:
falseskipCaptureAccessTimeOnlyChange: false
20/12/05 10:42:32 INFO util.GSet: Computing capacity for map cachedBlocks
```

```
20/12/05 10:42:32 INFO util.GSet: VM type     = 64-bit
20/12/05 10:42:32 INFO util.GSet: 0.25% max memory 966.7 MB = 2.4 MB
20/12/05 10:42:32 INFO util.GSet: capacity   = 2^18 = 262144 entries
20/12/05 10:42:32 INFO metrics.TopMetrics: NNTop conf:
dfs.namenode.top.window.num.buckets = 10
20/12/05 10:42:32 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10 20/12/05
10:42:32 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
20/12/05 10:42:32 INFO namenode.FSNamesystem: Retry cache on namenode is enabled 20/12/05
10:42:32 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache
entry expiry time is 600000 millis
20/12/05 10:42:32 INFO util.GSet: Computing capacity for map NameNodeRetryCache
20/12/05 10:42:32 INFO util.GSet: VM type     = 64-bit
20/12/05 10:42:32 INFO util.GSet: 0.029999999329447746% max memory 966.7 MB = 297.0 KB

20/12/05 10:42:32 INFO util.GSet: capacity   = 2^15 = 32768 entries


20/12/05 10:42:32 INFO namenode.FSImage: Allocated new BlockPoolId: BP-
713228373-39.105.163.95-1607136152571
20/12/05 10:42:32 WARN namenode.NameNode: Encountered exception during format:
java.io.IOException: Cannot create directory /usr/local/cluster/hadoop/hdfs/name/current
at
org.apache.hadoop.hdfs.server.common.Storage$StorageDirectory.clearDirectory(Storage.java:36 1)
  at   org.apache.hadoop.hdfs.server.namenode.NNStorage.format(NNStorage.java:571)   at
  org.apache.hadoop.hdfs.server.namenode.NNStorage.format(NNStorage.java:592)        at
  org.apache.hadoop.hdfs.server.namenode.FSImage.format(FSImage.java:185)
  at org.apache.hadoop.hdfs.server.namenode.NameNode.format(NameNode.java:1211) at
 org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1655) at
  org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1782)
20/12/05 10:42:32 ERROR namenode.NameNode: Failed to start namenode.  java.io.IOException:
Cannot create directory /usr/local/cluster/hadoop/hdfs/name/current at
org.apache.hadoop.hdfs.server.common.Storage$StorageDirectory.clearDirectory(Storage.java:36 1)
  at   org.apache.hadoop.hdfs.server.namenode.NNStorage.format(NNStorage.java:571)   at
  org.apache.hadoop.hdfs.server.namenode.NNStorage.format(NNStorage.java:592)        at
  org.apache.hadoop.hdfs.server.namenode.FSImage.format(FSImage.java:185)
  at org.apache.hadoop.hdfs.server.namenode.NameNode.format(NameNode.java:1211) at
 org.apache.hadoop.hdfs.server.namenode.NameNode.createNameNode(NameNode.java:1655) at
  org.apache.hadoop.hdfs.server.namenode.NameNode.main(NameNode.java:1782)
20/12/05 10:42:32 INFO util.ExitUtil: Exiting with status 1: java.io.IOException: Cannot create
directory /usr/local/cluster/hadoop/hdfs/name/current
20/12/05 10:42:32 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at slave2/39.105.163.95
************************************************************/
```

总览



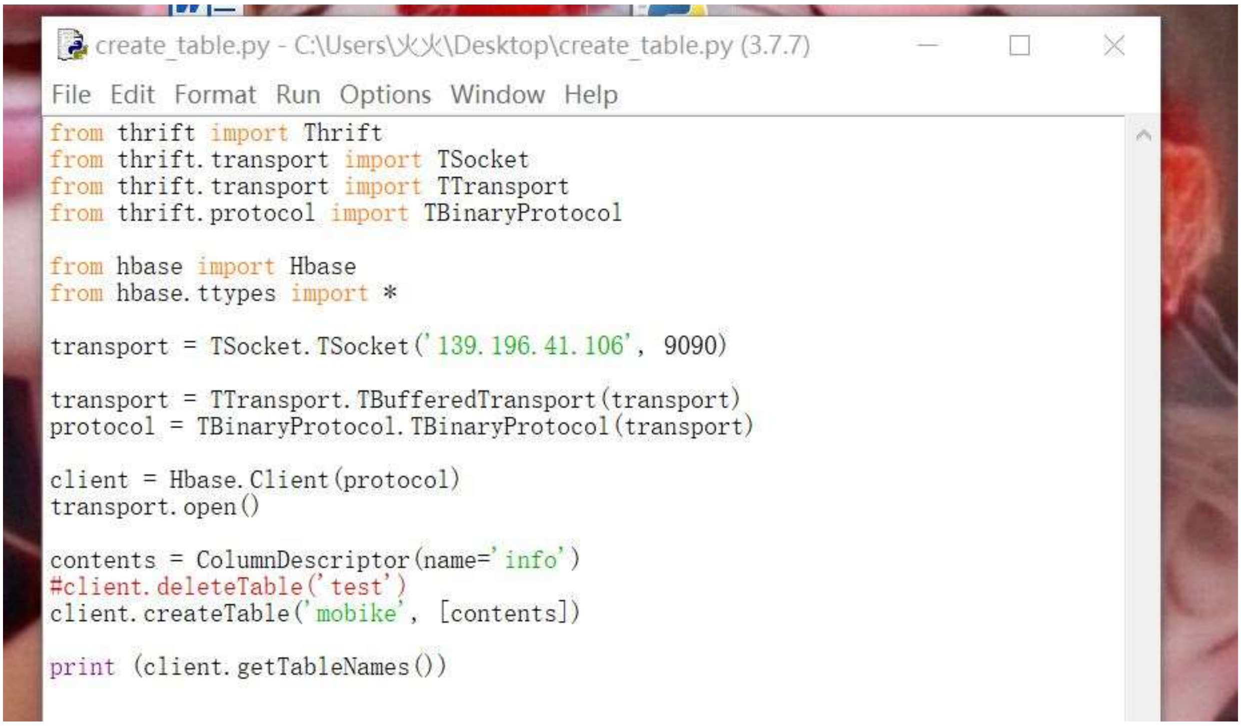
```
[hadoop@slave2 ~]$ cd /usr/local/cluster
[hadoop@slave2 cluster]$ ls
apache-zookeeper-3.6.2-bin  hadoop-2.10.1  hbase-2.3.3  zookeeper
hadoop                      hbase          jdk
[hadoop@slave2 cluster]$ cd apache-zookeeper-3.6.2-bin
[hadoop@slave2 apache-zookeeper-3.6.2-bin]$ ls
bin   data  lib          logs        README.md
conf  docs  LICENSE.txt  NOTICE.txt  README_packaging.md
[hadoop@slave2 apache-zookeeper-3.6.2-bin]$ cd ..
[hadoop@slave2 cluster]$ cd hadoop-2.10.1
[hadoop@slave2 hadoop-2.10.1]$ ls
bin  include  libexec      logs        README.txt  share
etc  lib      LICENSE.txt  NOTICE.txt  sbin
[hadoop@slave2 hadoop-2.10.1]$ cd ..
[hadoop@slave2 cluster]$ cd hbase-2.3.3
[hadoop@slave2 hbase-2.3.3]$ ls
bin         data           LEGAL        logs        RELEASENOTES.md
CHANGES.md  docs           lib          NOTICE.txt
conf        hbase-webapps  LICENSE.txt  README.txt
[hadoop@slave2 hbase-2.3.3]$ 
```

数据处理

本次报告与处理使用语言皆为python 未使用题干中给的java便于与之后大数据与语言的处理形成统一。

Thrift用于HBase中是为了提供跨平台的服务接口，在HBase 中可以使用[hbase-root]/bin/hbase thrift start 命令启动涵盖Thrift的HBase服务端，客户端通过thrift的命令生成不同版本的客户端代码，根据定义的数据格式，对远程HBase服务端进行 操作，是除了REST远程方法调用的另一种途径。

Thirft在此使用为9090端口



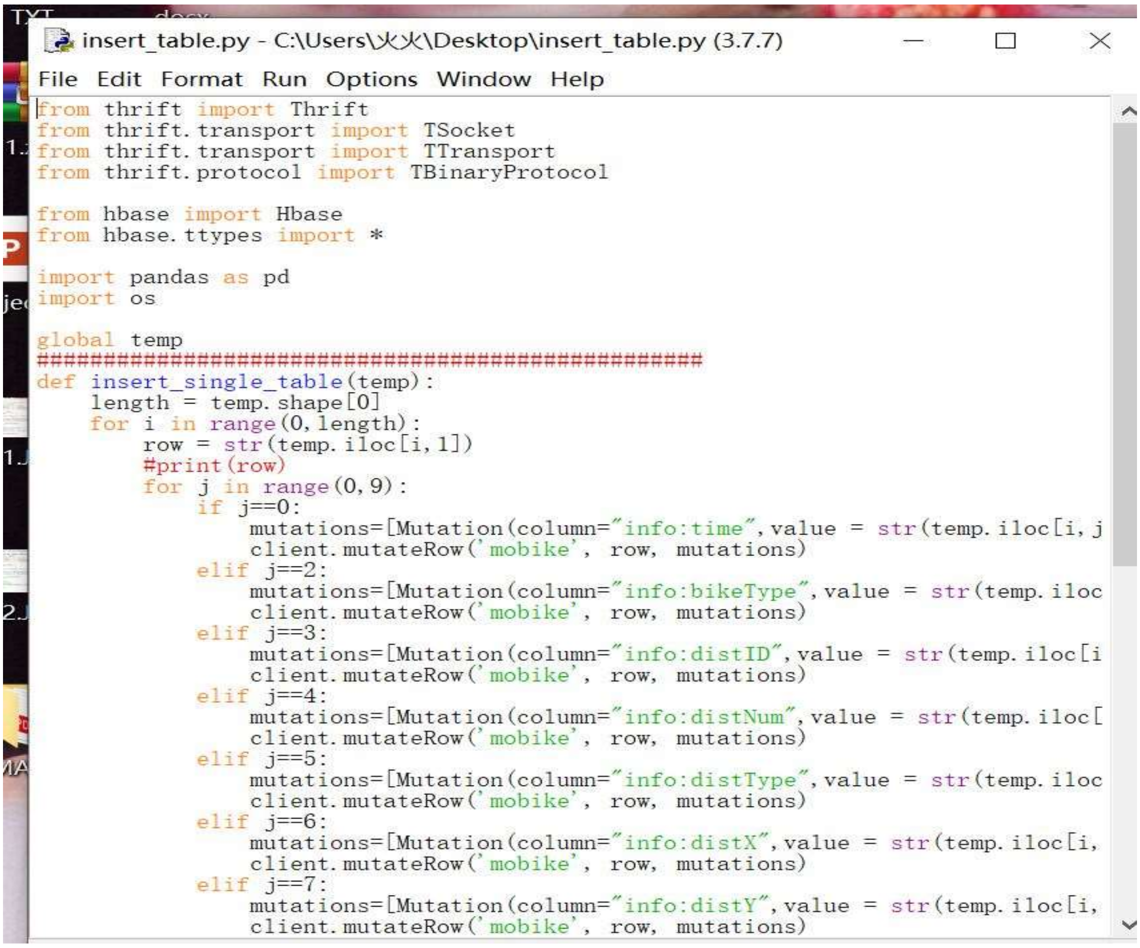
create_table.py - C:\Users\火火\Desktop\create_table.py (3.7.7)
File Edit Format Run Options Window Help

```
from thrift import Thrift
from thrift.transport import TSocket
from thrift.transport import TTransport
from thrift.protocol import TBinaryProtocol

from hbase import Hbase
from hbase.ttypes import *

transport = TSocket.TSocket('139.196.41.106', 9090)

transport = TTransport.TBufferedTransport(transport)
protocol = TBinaryProtocol.TBinaryProtocol(transport)

client = Hbase.Client(protocol)
transport.open()

contents = ColumnDescriptor(name='info')
#client.deleteTable('test')
client.createTable('mobike', [contents])

print (client.getTableNames())
```

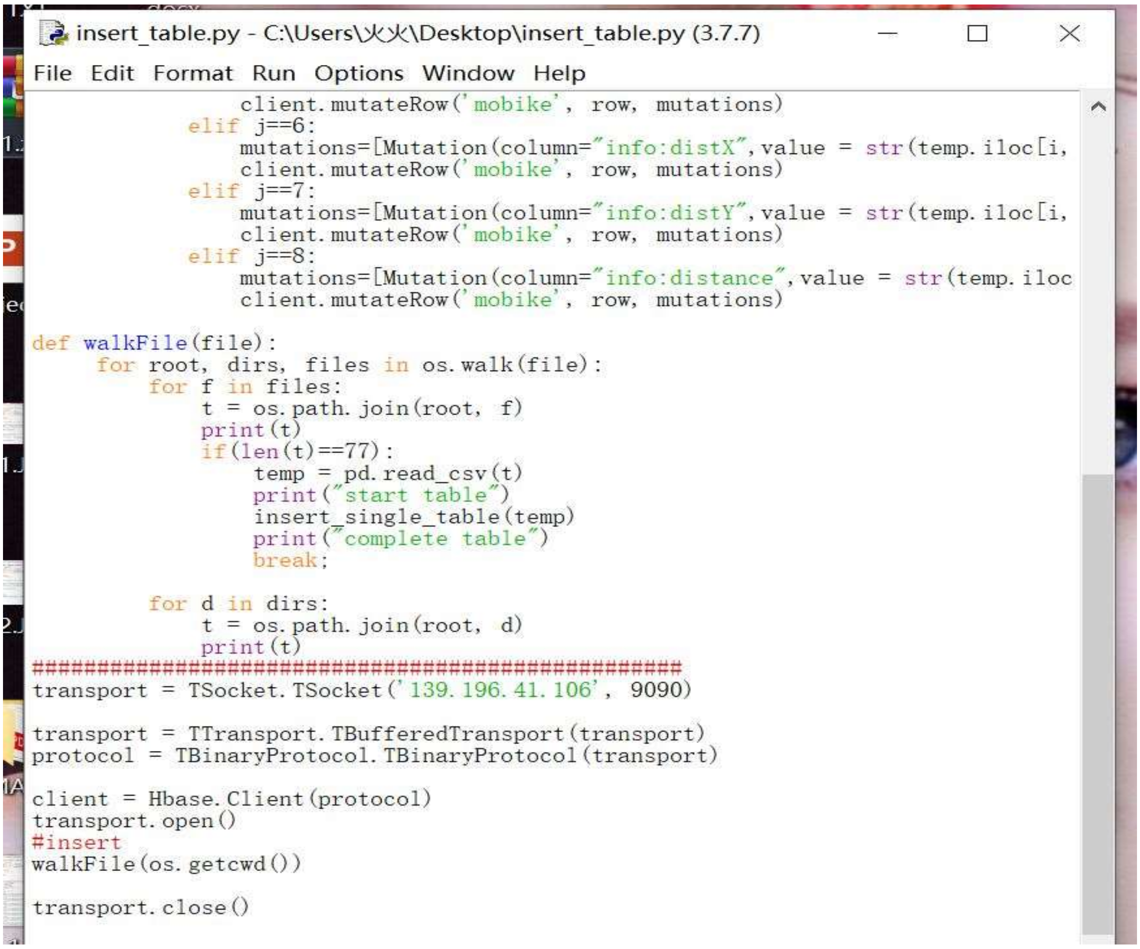
由定义可以看书TSocket继承至虚拟传输类，并且把自己当做模板参数传递过去，所以从虚拟传输类继承下来的虚拟函数（如read_virt）调用非虚拟函数（如read）就是TSocket自己实现的。

先确定infp下面的列簇里是否有time没有就创建time列，先读取并对每一个csv遍历注意内容皆为string，故处理起来比较麻烦

重要步骤:
1.把transport端口打开后 建立一个表
2.Content建立一个列簇
3.把数据放进去


insert_table.py - C:\Users\火火\Desktop\insert_table.py (3.7.7)
File Edit Format Run Options Window Help

```
from thrift import Thrift
from thrift.transport import TSocket
from thrift.transport import TTransport
from thrift.protocol import TBinaryProtocol

from hbase import Hbase
from hbase.ttypes import *

import pandas as pd
import os

global temp
##################################################
def insert_single_table(temp):
    length = temp.shape[0]
    for i in range(0,length):
        row = str(temp.iloc[i,1])
        #print(row)
        for j in range(0,9):
            if j==0:
                mutations=[Mutation(column="info:time",value = str(temp.iloc[i, j
                client.mutateRow('mobike', row, mutations)
            elif j==2:
                mutations=[Mutation(column="info:bikeType",value = str(temp.iloc
                client.mutateRow('mobike', row, mutations)
            elif j==3:
                mutations=[Mutation(column="info:distID",value = str(temp.iloc[i
                client.mutateRow('mobike', row, mutations)
            elif j==4:
                mutations=[Mutation(column="info:distNum",value = str(temp.iloc[
                client.mutateRow('mobike', row, mutations)
            elif j==5:
                mutations=[Mutation(column="info:distType",value = str(temp.iloc
                client.mutateRow('mobike', row, mutations)
            elif j==6:
                mutations=[Mutation(column="info:distX",value = str(temp.iloc[i,
                client.mutateRow('mobike', row, mutations)
            elif j==7:
                mutations=[Mutation(column="info:distY",value = str(temp.iloc[i,
                client.mutateRow('mobike', row, mutations)
```


insert_table.py - C:\Users\火火\Desktop\insert_table.py (3.7.7)
File Edit Format Run Options Window Help

```
                client.mutateRow('mobike', row, mutations)
            elif j==6:
                mutations=[Mutation(column="info:distX",value = str(temp.iloc[i,
                client.mutateRow('mobike', row, mutations)
            elif j==7:
                mutations=[Mutation(column="info:distY",value = str(temp.iloc[i,
                client.mutateRow('mobike', row, mutations)
            elif j==8:
                mutations=[Mutation(column="info:distance",value = str(temp.iloc
                client.mutateRow('mobike', row, mutations)

def walkFile(file):
    for root, dirs, files in os.walk(file):
        for f in files:
            t = os.path.join(root, f)
            print(t)
            if(len(t)==77):
                temp = pd.read_csv(t)
                print("start table")
                insert_single_table(temp)
                print("complete table")
                break;

        for d in dirs:
            t = os.path.join(root, d)
            print(t)
##################################################
transport = TSocket.TSocket('139.196.41.106', 9090)

transport = TTransport.TBufferedTransport(transport)
protocol = TBinaryProtocol.TBinaryProtocol(transport)

client = Hbase.Client(protocol)
transport.open()
#insert
walkFile(os.getcwd())

transport.close()
```

手动添加文件中的内容

Hbase是一个收集bike的ID的集合rokey和列簇的数据库，写作row

列簇的特点：
1.一张表通常有一单独的列簇，而且一张表中的列簇不会超过5个。
2.列簇必须在创建表的时候定义。
3.表的列簇无法改变。
4.每个列簇中的列数是没有限制的。
5.同一列簇下的所有列会保存在一起。
6.列在列簇中是有序的。
7.列在运行时创建。
8.列只有插入后才会存在，空值并不保存。

列簇的作用：
权限控制、存储以及调优都是在列族层面进行的
实际应用中，列族上的控制权限能 帮助我们管理不同类型的应用：我们允许一些应用可以添加新的基本数据、一些应用可以读取基本数据并创建继承的列族、一些应用则只允许浏览数据（甚至可能因 为隐私的原因不能浏览所有数据）。

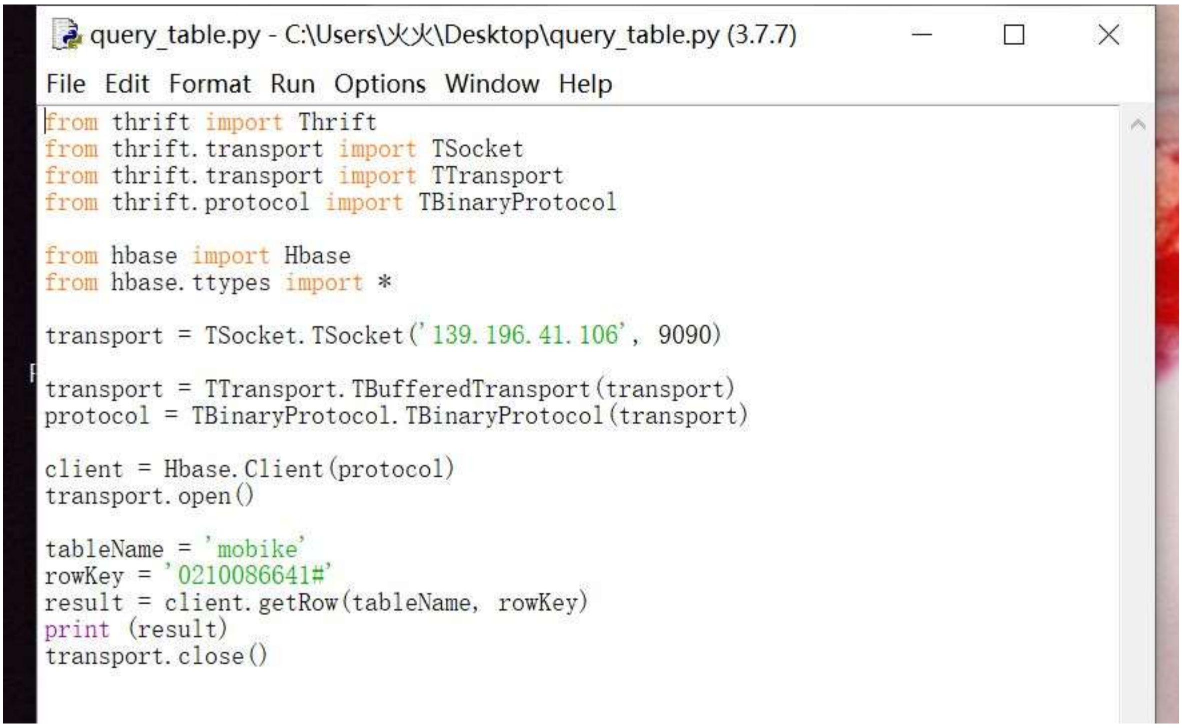
列簇与数据存储的关系为对同一个行键的访问都会落在同样的物理节点上。如果表包含2个列簇，属于两个列簇的文件还是保存在相同的节点上。因此，行键和节点存在一一对应的关系。

Thirft里的 mutation操作

一.什么是mutation
通俗的理解mutations,里面装着一些改变数据方法的集合，这是Veux设计很重要的一点，就是把处理数据逻辑方法全部放在mutations里面，使得数据和视图分离。
二.怎么用mutations?
1.mutation结构
每一个mutation都有一个字符串类型的事件类型(type)和回调函数(handler)，也可以理解为{type:handler()},这和订阅发布有点类似。先注册事件，当触发响应类型的时候调用handker()，调用type的时候需要用到store.commit方法。
2.提交载荷（Payload）
简单的理解就是往handler(state)中传参handler(state,pryload)；一般是个对象
3.commit
提交可以在组件中使用 this.$store.commit('xxx') 提交 mutation，或者使用 mapMutations 辅助函数将组件中的 methods 映射为 store.commit 调用（需要在根节点注入 store）。

4.使用常量替代 Mutation 事件类型

使用常量替代 mutation 事件类型在各种 Flux 实现中是很常见的模式。这样可以使 linter 之类的工具发挥作用，同时把这些常量放在单独的文件中可以对整个 app 包含的 mutation 一目了然


query_table.py - C:\Users\火火\Desktop\query_table.py (3.7.7)
File Edit Format Run Options Window Help

```
from thrift import Thrift
from thrift.transport import TSocket
from thrift.transport import TTransport
from thrift.protocol import TBinaryProtocol

from hbase import Hbase
from hbase.ttypes import *

transport = TSocket.TSocket('139.196.41.106', 9090)

transport = TTransport.TBufferedTransport(transport)
protocol = TBinaryProtocol.TBinaryProtocol(transport)

client = Hbase.Client(protocol)
transport.open()

tableName = 'mobike'
rowKey = '0210086641#'
result = client.getRow(tableName, rowKey)
print (result)
transport.close()
```

class Client(Iface)

1.def enableTable(self, tableName)

2.def disableTable(self, tableName)

3.def isTableEnabled(self, tableName)

4.def compact(self, tableNameOrRegionName)

Compaction 主要起到如下几个作用：

1）合并文件

2）清除删除、过期、多余版本的数据

3）提高读写数据的效率

5.majorCompact（self, tableNameOrRegionName）

Major 操作是对 Region 下的 HStore 下的所有 StoreFile 执行合并操作，最终的结果是整理合并出一个文件。

6.getTableNames(self)

得到数据表名

7.getColumnDescriptors(self, tableName)

得到数据表的列描述（有几个列等等）

8.getTableRegions(self, tableName)

9.createTable(tableName,columnFamilies)

10.deleteTable(self, tableName)

11.get(self, tableName, row, column, attributes)

得到一条记录，attributes 是得到的属性

12.getVer(self, tableName, row, column, numVersions, attributes)

得到特定 table，row,column 的特定的 version 的数目

13.getVerTs(self, tableName, row, column,timestamp, numVersions, attributes)

只有小于或等于特定 timestamp 的记录会被返回

14.getRow（self, tableName, row, attributes）

15.getRowWithColumns(self, tableName, row, columns, attributes)

得到特定的列

16.getRowTs(self, tableName, row, timestamp, attributes)

得到特定时间的 table 中的 row

17.getRowWithColumnsTs(self, tableName, row, columns, timestamp, attributes)

18.getRows(self, tableName, rows, attributes)

19.getRowsWithColumns(self, tableName, rows, columns, attributes)

20.getRowsTs(self, tableName, rows, timestamp, attributes)

21.getRowsWithColumnsTs(self, tableName, rows, columns, timestamp, attributes)

Mutate

22.mutateRow(self, tableName, row, mutations, attributes)
更新或删除数据

23.mutateRowTs(self, tableName, row, mutations, timestamp, attributes)

24.mutateRows(self, tableName, rowBatches, attributes)

25.mutateRowsTs(self, tableName, rowBatches, timestamp, attributes)

26.atomicIncrement(self, tableName, row, column, value)

原子方式增加指定的列值。返回一个递增后的值。

27.deleteAll(self, tableName, row, column, attributes)

删除与给定的row和column匹配的所有单元。

28.deleteAllTs（self, tableName, row, column, timestamp, attributes）

增加条件：删除timestamp 大于等于给定的timestamp值

29.deleteAllRow(self, tableName, row, attributes)

完全删除row