

参考教程: [Apache HBase™ Reference Guide](#)

操作系统: Linux Ubuntu、Windows 10

## 单机模式

安装hbase, 修改hbase-env.sh的内容, 主要是自己的Java路径

启动hbase。成功:

```
wangsky@ubuntu:~/hadoop/hbase$ jps
15432 Jps
15118 HMaster
wangsky@ubuntu:~/hadoop/hbase$ ./bin/hbase shell
2020-11-21 05:47:52,537 WARN [main] util.NativeCodeLoader: Unable to load native hbase-hadoop library for your platform... using builtin-java classes where applicable
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
For Reference, please visit: http://hbase.apache.org/2.0/book.html#shell
Version 2.2.5, rf76a601273e834267b55c0cda12474590283fd4c, 2020年 05月 21日 星期四 18:34:40 CST
Took 0.0125 seconds
hbase(main):001:0>
```

## 伪分布式

更改hbase-site里的配置, 完成伪分布式

```
wangsky@ubuntu:~/hadoop/hbase$ ./bin/start-hbase.sh
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
localhost: running zookeeper, logging to /home/wangsky/hadoop/hbase/bin/../logs/hbase-wangsky-zookeeper-ubuntu.out
localhost: running master, logging to /home/wangsky/hadoop/hbase/bin/../logs/hbase-wangsky-master-ubuntu.out
localhost: running regionserver, logging to /home/wangsky/hadoop/hbase/bin/../logs/hbase-wangsky-regionserver-ubuntu.out
wangsky@ubuntu:~/hadoop/hbase$ jps
25456 Jps
25287 HRegionServer
25208 HMaster
25147 HQuorumPeer
wangsky@ubuntu:~/hadoop/hbase$
```

📝 关闭虚拟机再重启hdfs时发现namenode无法正常启动, 打开log发现:

```

2020-11-23 10:44:04,414 INFO org.apache.hadoop.util.ExitUtil: Exiting with status 1:
org.apache.hadoop.hdfs.server.common.InconsistentFSStateException: Directory /tmp/hadoop-root/dfs/name is in an inconsistent state: storage
directory does not exist or is not accessible.
2020-11-23 10:44:04,415 INFO org.apache.hadoop.hdfs.server.namenode.NameNode: SHUTDOWN_MSG:
/*****
SHUTDOWN_MSG: Shutting down NameNode at ubuntu/127.0.1.1
*****/

```

原因可能是因为哪一次关闭出现了问题，导致tmp文件夹下某个文件在关机的时候被删除，但是hadoop没有得知这个信息（我估计就是stop-all引发的问题）

解决方法：将core-site.xml里面设置tmp-dir为其他文件夹，并且重新初始化

```

root@ubuntu:/home/wangsky/hadoop/hadoop-3.2.1/sbin# ./start-dfs.sh
Starting namenodes on [localhost]
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Starting datanodes
localhost: Warning: Permanently added 'localhost' (ECDSA) to the list of known hosts.
Starting secondary namenodes [ubuntu]
ubuntu: Warning: Permanently added 'ubuntu' (ECDSA) to the list of known hosts.
root@ubuntu:/home/wangsky/hadoop/hadoop-3.2.1/sbin# jps
10864 DataNode
11265 Jps
11128 SecondaryNameNode
10652 NameNode
root@ubuntu:/home/wangsky/hadoop/hadoop-3.2.1/sbin#

```

然后就会出现Hmaster消失的问题：

```

wangsky@ubuntu:~/hadoop$ jps
57763 HRegionServer
57620 HQuorumPeer
58101 Jps
wangsky@ubuntu:~/hadoop$

```

换用外部zookeeper之后：

```

wangsky@ubuntu:~/hadoop/hbase$ ./bin/start-hbase.sh
running master, logging to /home/wangsky/hadoop/hbase/bin/../logs/hbase-wangsky-master-ubuntu.out
: running regionserver, logging to /home/wangsky/hadoop/hbase/bin/../logs/hbase-wangsky-regionserver-ubuntu.out
wangsky@ubuntu:~/hadoop/hbase$ jps
60353 HRegionServer
59859 QuorumPeerMain
60403 Jps
60271 HMaster
wangsky@ubuntu:~/hadoop/hbase$

```

在确保了一切配置都正确之后，过一段时间，Hmaster仍然消失，查一下log发现报错和之前一样

```

2020-11-25 10:19:05,082 WARN [master/ubuntu:16000] master.ActiveMasterManager: Failed get of master address: java.io.IOException: Can't get
master address from ZooKeeper; znode data == null
2020-11-25 10:19:05,107 INFO [main-EventThread] zookeeper.ClientCnxn: EventThread shut down for session: 0x10000aef1490001
2020-11-25 10:19:05,107 INFO [master/ubuntu:16000] zookeeper.ZooKeeper: Session: 0x10000aef1490001 closed
2020-11-25 10:19:05,107 INFO [master/ubuntu:16000] regionserver.HRegionServer: Exiting; stopping=ubuntu,16000,1606299530040; zookeeper
connection closed.
2020-11-25 10:19:05,117 ERROR [main] master.HMasterCommandLine: Master exiting
java.lang.RuntimeException: HMaster Aborted
    at org.apache.hadoop.hbase.master.HMasterCommandLine.startMaster(HMasterCommandLine.java:244)
    at org.apache.hadoop.hbase.master.HMasterCommandLine.run(HMasterCommandLine.java:140)
    at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:76)
    at org.apache.hadoop.hbase.util.ServerCommandLine.doMain(ServerCommandLine.java:149)
    at org.apache.hadoop.hbase.master.HMaster.main(HMaster.java:2945)

```

都是 Failed to get of mater address

尝试了各种各样的方法都没法成功之后，我选择在另一个虚拟机上重装hadoop和hbase，我觉得应该是版本不匹配的问题

在网上找到了一个版本匹配情况：

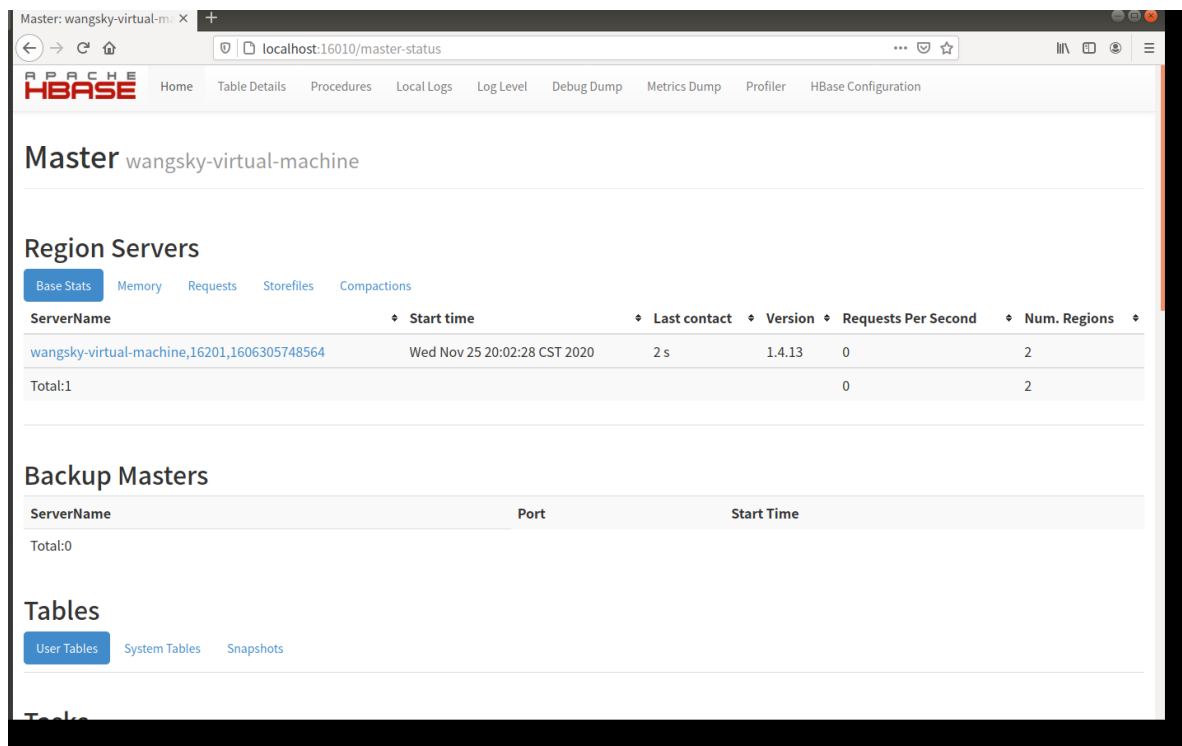
	HBase-1.3.x	HBase-1.4.x	HBase-1.5.x	HBase-2.1.x	HBase-2.2.x	HBase-2.3.x
Hadoop-2.4.x	✓	✗	✗	✗	✗	✗
Hadoop-2.5.x	✓	✗	✗	✗	✗	✗
Hadoop-2.6.0	✗	✗	✗	✗	✗	✗
Hadoop-2.6.1+	✓	✗	✗	✗	✗	✗
Hadoop-2.7.0	✗	✗	✗	✗	✗	✗
Hadoop-2.7.1+	✓	✓	✗	✓	✗	✗
Hadoop-2.8.[0-2]	✗	✗	✗	✗	✗	✗
Hadoop-2.8.[3-4]	!	!	✗	✓	✗	✗
Hadoop-2.8.5+	!	!	✓	✓	✓	✗
Hadoop-2.9.[0-1]	✗	✗	✗	✗	✗	✗
Hadoop-2.9.2+	!	!	✓	!	✓	✗
Hadoop-2.10.0	!	!	✓	!	!	✓
Hadoop-3.0.[0-2]	✗	✗	✗	✗	✗	✗
Hadoop-3.0.3+	✗	✗	✗	✓	✗	✗
Hadoop-3.1.0	✗	✗	✗	✗	✗	✗
Hadoop-3.1.1+	✗	✗	✗	✓	<a href="https://log.csdn.net/qq_32474">https://log.csdn.net/qq_32474</a>	

按照这个表，分别安装了hadoop2.7.2和hbase1.4.13，配置好之后终于成功了 🎉

```

wangsky@wangsky-virtual-machine:~/hadoop/hbase$ jps
14868 DataNode
15866 HRegionServer
15770 HMaster
15692 HQuorumPeer
15084 SecondaryNameNode
16125 Jps
14687 NameNode
wangsky@wangsky-virtual-machine:~/hadoop/hbase$ jps
16148 Jps
14868 DataNode
15866 HRegionServer
15770 HMaster
15692 HQuorumPeer
15084 SecondaryNameNode
14687 NameNode
wangsky@wangsky-virtual-machine:~/hadoop/hbase$

```



# windows

考虑到编程的方便性，这里选择在Windows上再安装一遍，便于使用idea做maven的项目管理

📝 windows下最好还是安装1.x版本，安装2.x版本时会出现一些问题

📝 安装和配置都和linux下没有本质区别，只是脚本文件是cmd不是sh

jps显示Hmaster正常

```
D:\Hadoop\hbase-1.6.0\bin>hbase shell
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:D:/Hadoop/hbase-1.6.0/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:D:/Hadoop/hadoop-2.7.7/share/hadoop/common/lib/slf4j-log4j12-1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
HBase Shell
Use "help" to get list of supported commands.
Use "exit" to quit this interactive shell.
Version 1.6.0, r5ec5a5b115ee36fb28903667c008218abd21b3f5, Fri Feb 14 12:00:03 PST 2020
hbase(main):001:0> exit

D:\Hadoop\hbase-1.6.0\bin>jps
1248 Jps
2336 BootLanguagServerBootApplication
1720 XMLServerLauncher
17608 HMaster
20984 NameNode
10620 org.eclipse.equinox.launcher_1.6.0.v20200915-1508.jar
5068 DataNode

D:\Hadoop\hbase-1.6.0\bin>
```

可以看到hdfs下的hbase目录

## Browse Directory

/hbase							Go!
Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	wangs	supergroup	0 B	2020/11/22 上午10:15:36	0	0 B	<a href="#">.tmp</a>
drwxr-xr-x	wangs	supergroup	0 B	2020/11/22 上午10:15:47	0	0 B	<a href="#">MasterProcWALS</a>
drwxr-xr-x	wangs	supergroup	0 B	2020/11/22 上午10:14:52	0	0 B	<a href="#">WALS</a>
drwxr-xr-x	wangs	supergroup	0 B	2020/11/22 上午10:15:39	0	0 B	<a href="#">data</a>
-rw-r--r--	wangs	supergroup	42 B	2020/11/22 上午10:14:20	1	128 MB	<a href="#">hbase.id</a>
-rw-r--r--	wangs	supergroup	7 B	2020/11/22 上午10:14:19	1	128 MB	<a href="#">hbase.version</a>
drwxr-xr-x	wangs	supergroup	0 B	2020/11/22 上午10:14:21	0	0 B	<a href="#">oldWALS</a>

Hadoop, 2018.

# Java代码实现

IDE：Idea

OS：Windows 10

代码管理：Maven

代码见github仓库 <https://github.com/wangskyGit/FBDPcode.git>

pom.xml中需要增加的依赖如下,版本号与自己安装的hbase一致：

```
<dependencies>
  <dependency>
    <groupId>org.apache.hbase</groupId>
    <artifactId>hbase-client</artifactId>
    <version>1.6.0</version>
  </dependency>
</dependencies>
```

运行结果：

```
start insert data .....
end insert data .....
start insert data .....
end insert data .....
start insert data .....
end insert data .....
rowkey:001
列: Courses/Chinese====值:80
列: Courses/Math====值:90
列: Courses/Physics====值:95
列: Description/Height====值:176
列: Description/Name====值:Li Lei
列: Home/Province====值:Zhejiang
列: ID/ID====值:001
rowkey:002
列: Courses/Chinese====值:88
列: Courses/Math====值:77
列: Courses/Physics====值:66
列: Description/Height====值:183
列: Description/Name====值:Han Meimei
列: Home/Province====值:Beijing
列: ID/ID====值:002
rowkey:003
列: Courses/Chinese====值:90
列: Courses/Math====值:90
列: Courses/Physics====值:90
列: Description/Height====值:162
列: Description/Name====值:Xiao Ming
列: Home/Province====值:Shanghai
列: ID/ID====值:003
rowkey:001
名字: Li Lei====Home:Zhejiang
rowkey:002
名字: Han Meimei====Home:Beijing
rowkey:003
名字: Xiao Ming====Home:Shanghai
start insert data .....
end insert data .....
column added
start insert data .....
end insert data .....

Process finished with exit code 0
|
```

运行成功!

# shell 运行

1. 创建表格，并插入数据

```

=> []
hbase(main):002:0> create 'students','ID','Description','Courses','Home'
0 row(s) in 4.4600 seconds

=> Hbase::Table - students
hbase(main):003:0> put 'students','001','Description:Name','Li Lei'
0 row(s) in 0.0670 seconds

hbase(main):004:0> put 'students','001','Description:Height','176'
0 row(s) in 0.0040 seconds

hbase(main):005:0> put 'students','001','Courses:Chinese','80'
0 row(s) in 0.0030 seconds

hbase(main):006:0> put 'students','001','Courses:Math','90'
0 row(s) in 0.0030 seconds

hbase(main):007:0> put 'students','001','Courses:Physics','95'
0 row(s) in 0.0040 seconds

hbase(main):008:0> put 'students','001','Home:Province','Zhejiang'
0 row(s) in 0.0060 seconds

hbase(main):009:0> put 'students','002','Description:Name','Han Meimei'
0 row(s) in 0.0060 seconds

hbase(main):010:0> put 'students','002','Description:Height','183'
0 row(s) in 0.0050 seconds

hbase(main):011:0> put 'students','002','Courses:Chinese','88'
0 row(s) in 0.0050 seconds

hbase(main):012:0> put 'students','002','Courses:Math','77'
0 row(s) in 0.0030 seconds

hbase(main):013:0> put 'students','002','Courses:Physics','66'
0 row(s) in 0.0050 seconds

hbase(main):014:0> put 'students','002','Home:Province','Beijing'
0 row(s) in 0.0040 seconds

hbase(main):015:0> put 'students','003','Description:Name','Xiao Ming'
0 row(s) in 0.0040 seconds

hbase(main):016:0> put 'students','003','Description:Height','162'
0 row(s) in 0.0030 seconds

hbase(main):017:0> put 'students','003','Courses:Chinese','80'
0 row(s) in 0.0030 seconds

hbase(main):018:0> put 'students','003','Courses:Math','90'
0 row(s) in 0.0040 seconds

hbase(main):019:0> put 'students','003','Courses:physics','95'
0 row(s) in 0.0040 seconds

hbase(main):020:0> put 'students','003','Home:Province','Shanghai'
0 row(s) in 0.0040 seconds

```

## 2. scan students

```
hbase(main):022:0> scan 'students'
ROW                                COLUMN+CELL

001                                column=Courses:Chinese, timestamp=1606121043804, value=80
001                                column=Courses:Math, timestamp=1606121055806, value=90
001                                column=Courses:Physics, timestamp=1606121069133, value=95
001                                column=Description:Height, timestamp=1606121012463, value=176
001                                column=Description:Name, timestamp=1606120992488, value=Li Lei
001                                column=Home:Province, timestamp=1606122183571, value=Zhejiang
002                                column=Courses:Chinese, timestamp=1606122306523, value=88
002                                column=Courses:Math, timestamp=1606122331598, value=77
002                                column=Courses:Physics, timestamp=1606122346306, value=66
002                                column=Description:Height, timestamp=1606122246569, value=183
002                                column=Description:Name, timestamp=1606122222577, value=Han Meimei
002                                column=Home:Province, timestamp=1606122375046, value=Beijing
003                                column=Courses:Chinese, timestamp=1606122463044, value=80
003                                column=Courses:Math, timestamp=1606122475966, value=90
003                                column=Courses:physics, timestamp=1606122488875, value=95
003                                column=Description:Height, timestamp=1606122440069, value=162
003                                column=Description:Name, timestamp=1606122423898, value=Xiao Ming
003                                column=Home:Province, timestamp=1606122512759, value=Shanghai

3 row(s) in 0.2410 seconds
```

### 3. 查询学生来自的省

```
hbase(main):024:0> scan 'students',{COLUMNS=>'Home:Province'}
ROW                                COLUMN+CELL

001                                column=Home:Province, timestamp=1606122183571, value=Zhejiang
002                                column=Home:Province, timestamp=1606122375046, value=Beijing
003                                column=Home:Province, timestamp=1606122512759, value=Shanghai

3 row(s) in 0.1800 seconds

hbase(main):025:0> █
```

### 4. (4) 增加新的列Courses:English, 并添加数据;

```
hbase(main):025:0> put 'students','001','Courses:English','100'
0 row(s) in 0.0060 seconds
```

### 5. 增加新的列族Contact和新列 Contact:Email, 并添加数据;

```
hbase(main):027:0> disable('students')
0 row(s) in 4.5710 seconds

hbase(main):028:0> alter 'students','Contact'
Updating all regions with the new schema...
1/1 regions updated.
Done.
0 row(s) in 3.5320 seconds

hbase(main):029:0> enable('students')
0 row(s) in 4.4490 seconds

hbase(main):030:0> put 'students','001','Contact:email','wsq_super_handsome@truth.com'
0 row(s) in 0.0050 seconds

hbase(main):031:0> █
```



## 6. 删除students表

```
hbase(main):031:0> disable('students')
0 row(s) in 4.4130 seconds

hbase(main):032:0> drop('students')
0 row(s) in 2.9540 seconds
```