

金融大数据实验3

姓名：盛祺晨 学号：191220093

金融大数据实验3

环境配置

设计表格

SHELL

设计合适的表

查询选修Computer Science的学生的成绩

增加新的列族和新列Contact:Email,并添加数据

删除学号为2015003的学生的选课记录

删除所创建的表

JAVA

运行结果

设计并创建合适的表

插入数据

查询选修Computer Science的学生的成绩

增加新的列族和新列Contact:Email,并添加数据

删除学号为2015003的学生的选课记录

删除所创建的表

问题描述

环境配置

Mac本机配置环境会出现RegionServer不存在的情况，我更换了hadoop和HBase的版本号、使用外部zookeeper都不能解决，最终采用了申请BDKit的方法，非常有效。

对于在本地配置中出现的问题，将在问题部分描述。

拉取Hadoop3.1.4+Hbase2.2.6的镜像。并能正确进去shell编程。

Master

sqc191220093-master

Regions in Transition

1 region(s) in transition. 0 region(s) in transition for more than 60000 milliseconds.

Region	State	RIT time (ms)	Retrie
1588230740	hbase:meta,,1.1588230740 state=OPEN, ts=Sun Nov 28 05:52:36 UTC 2021 (0s ago), server=192-168-219-155.sqc191220093-slave-2.spark-cluster.svc.cluster.local,16020,1638028384313	298	0

prev

1

next

Region Servers

Backup Masters

ServerName	Port	Start Time
Total:0		

设计表格

rowKey(S_No)	info			Course(123001)				Course(123002....)	Contact
	S_Name	S_Sex	S_Age	C_No	C_Name	C_credit	C_score	C_No.....C_score	S_email
2015001	Li Lei	Male	23	123001	Math	2.0	86	...	lilie@qq.com
2015002	Han Meimei	Female	22						hmm@qq.com
2015003	Zhang San	Male	24	123001	Math	2.0	98		zs@qq.com

SHELL

设计合适的表

```
> create 'student','S_No','info','course1','course2','course3'
# 基础信息
> put 'student','2015001','info:S_Name','Li Lei'
> put 'student','2015001','info:S_Sex','male'
> put 'student','2015001','info:S_Age','23'

> put 'student','2015002','info:S_Name','Han Meimei'
> put 'student','2015002','info:S_Sex','female'
> put 'student','2015002','info:S_Age','22'

> put 'student','2015003','info:S_Name','Zhang San'
> put 'student','2015003','info:S_Sex','male'
> put 'student','2015003','info:S_Age','24'
# 课程信息
# student 2015001
> put 'student','2015001','course1:C_No','123001'
> put 'student','2015001','course1:C_Name','Math'
> put 'student','2015001','course1:C_Credit','2.0'
> put 'student','2015001','course1:C_Score','86'
```

```

> put 'student','2015001','course3:C_No','123003'
> put 'student','2015001','course3:C_Name','English'
> put 'student','2015001','course3:C_Credit','3.0'
> put 'student','2015001','course3:C_Score','69'
# student 2015002
> put 'student','2015002','course2:C_No','123002'
> put 'student','2015002','course2:C_Name','Computer Science'
> put 'student','2015002','course2:C_Credit','5.0'
> put 'student','2015002','course2:C_Score','77'

> put 'student','2015002','course3:C_No','123003'
> put 'student','2015002','course3:C_Name','English'
> put 'student','2015002','course3:C_Credit','3.0'
> put 'student','2015002','course3:C_Score','99'
# student 2015003
> put 'student','2015003','course1:C_No','123001'
> put 'student','2015003','course1:C_Name','Math'
> put 'student','2015003','course1:C_Credit','2.0'
> put 'student','2015003','course1:C_Score','98'

> put 'student','2015003','course2:C_No','123002'
> put 'student','2015003','course2:C_Name','Computer Science'
> put 'student','2015003','course2:C_Credit','5.0'
> put 'student','2015003','course2:C_Score','95'

```

查询选修Computer Science的学生的成绩

```

hbase(main):045:0> scan 'student',{COLUMN=>'course2:C_Score'}
ROW                                COLUMN+CELL

  2015002                            column=course2:C_Score, timestamp=1638083414463,
value=77
  2015003                            column=course2:C_Score, timestamp=1638083712088,
value=95
2 row(s)
Took 0.0275 seconds

```

增加新的列族和新列Contact:Email,并添加数据

```

#联系方式
> alter 'student','Contact'
> put 'student','2015001','Contact:Email','lilei@qq.com'
> put 'student','2015002','Contact:Email','hmm@qq.com'
> put 'student','2015003','Contact:Email','zs@qq.com'

```



```

info:S_Age                                timestamp=1638083345604, value=24

info:S_Name                                timestamp=1638083335992, value=Zhang San

info:S_Sex                                timestamp=1638083340980, value=male

1 row(s)
Took 0.0787 seconds

```

删除所创建的表

```

> disable 'student'
> drop 'student'
> list
TABLE

0 row(s)
Took 0.0278 seconds

=> []

```

```

hbase(main):058:0> get 'student','2015003'
COLUMN                                CELL
Contact:Email                        timestamp=1638083767501, value=zs@qq.com
course2:C_Credit                      timestamp=1638083707487, value=5.0
course2:C_Name                        timestamp=1638083702874, value=Computer Science
course2:C_No                          timestamp=1638083698085, value=123002
course2:C_Score                       timestamp=1638083712088, value=95
info:S_Age                           timestamp=1638083345604, value=24
info:S_Name                           timestamp=1638083335992, value=Zhang San
info:S_Sex                           timestamp=1638083340980, value=male
1 row(s)
Took 0.0787 seconds
hbase(main):059:0> disable 'student'
Took 2.3601 seconds
hbase(main):060:0> drop 'student'
Took 1.5248 seconds
hbase(main):061:0> list
TABLE
0 row(s)
Took 0.0278 seconds
=> []
hbase(main):062:0> █

```

JAVA

将程序打包后上传到BDKit后，在BDKit中运行

```
hadoop jar Hbase-1.0-SNAPSHOT.jar Hbase
```

运行结果

结果如下

```
PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE 1: bash + □ ☒ ∨ ×

insert recored 2015003 to table student ok.
insert recored 2015003 to table student ok.
insert recored 2015003 to table student ok.
insert recored 2015003 to table student ok.
insert recored 2015003 to table student ok.
insert recored 2015003 to table student ok.
insert recored 2015003 to table student ok.
end creating !
searching scores of course Computer Science ...
keyvalues={2015002/course2:C_Score/1638090027197/Put/vlen=2/seqid=0}
RowKey:2015002 Family:course2 Qualifier:C_Score Value:77
keyvalues={2015003/course2:C_Score/1638090027317/Put/vlen=2/seqid=0}
RowKey:2015003 Family:course2 Qualifier:C_Score Value:95
end searching !
add new column family Contact:Email
add column family Contact ok
insert recored 2015001 to table student ok.
insert recored 2015002 to table student ok.
insert recored 2015003 to table student ok.
end adding !
deleting courses of student 2015003...
delete 2015003:course1:C_No susscess
delete 2015003:course1:C_Name susscess
delete 2015003:course1:C_Credit susscess
delete 2015003:course1:C_Score susscess
delete 2015003:course2:C_No susscess
delete 2015003:course2:C_Name susscess
delete 2015003:course2:C_Credit susscess
delete 2015003:course2:C_Score susscess
delete 2015003:course3:C_No susscess
delete 2015003:course3:C_Name susscess
delete 2015003:course3:C_Credit susscess
delete 2015003:course3:C_Score susscess
end deleting
deleting the total table...
2021-11-28 09:00:31,138 INFO client.HBaseAdmin: Started disable of student
2021-11-28 09:00:33,426 INFO client.HBaseAdmin: Operation: DISABLE, Table Name: default:student, procId: 85 complete
d
2021-11-28 09:00:33,937 INFO client.HBaseAdmin: Operation: DELETE, Table Name: default:student, procId: 88 completed
deletetable student ok.
end deleting the total table
[root@sqc191220093-master FBDP_HBase]#
```

和上述shell中的思想一样，先连接hbase

```
//配置Configuration
configuration = HBaseConfiguration.create();
configuration.set("hbase.zookeeper.quorum","localhost");
configuration.set("hbase.zookeeper.property.clientPort","2181");

connection = ConnectionFactory.createConnection(configuration);
admin = connection.getAdmin();
```

设计并创建合适的表

```
HTableDescriptor tableDescriptor = new HTableDescriptor(TableName.valueOf(tableName));
for (String familyName : familyNames) {
    tableDescriptor.addFamily(new HColumnDescriptor(familyName)); // 增加列族
}
admin.createTable(tableDescriptor); //创建表
```

插入数据

```
Table table = connection.getTable(TableName.valueOf(tableName)); // 获取table
Put put = new Put(Bytes.toBytes(rowKey)); // 定位rowkey
put.addColumn(Bytes.toBytes(family), Bytes.toBytes(column), Bytes.toBytes(value)); // 增加列族, 列, 值
table.put(put);
```

查询选修Computer Science的学生的成绩

```
Table table = connection.getTable(TableName.valueOf(tableName)); // 获取表的名字
Scanner scanner = table.getScanner(Bytes.toBytes(family), Bytes.toBytes(qualifier)); // 根据列族: 列名来定位
printScanResults(); // 打印结果
```

增加新的列族和新列Contact:Email,并添加数据

```
TableName tableName1 = TableName.valueOf(tableName); // 获取表名
HColumnDescriptor hColumnDescriptor = new HColumnDescriptor(columnName);
admin.addColumn(tableName1, hColumnDescriptor); // 增加表名: 列族名
```

删除学号为2015003的学生的选课记录

```
Table table = connection.getTable(TableName.valueOf(tableName)); // 获取表名
Delete deleteColumn = new Delete(Bytes.toBytes(rowKey)); // 找到rowkey
deleteColumn.addColumn(Bytes.toBytes(familyName), Bytes.toBytes(columnName)); // 要删除的column增加列族名: 列名
table.delete(deleteColumn); // 删除
```

删除所创建的表

```
admin.disableTable(TableName.valueOf(tableName)); // 将表禁用
admin.deleteTable(TableName.valueOf(tableName)); // 将表删除
```

问题描述

在后面使用了BDKit之后基本没有问题产生

在本地配置hbase环境时, 出现了以下问题:

1.stop-hbase.sh 一直处于等待状态

参考文章: https://blog.csdn.net/weixin_45462732/article/details/106909501

输入: hbase-daemon.sh stop master关闭master, 就可以stop-hbase.sh正常关闭hbase

2.jps中出现Hmaster和HRegionServer，但是点击localhost:16010网站后，发现并没有连接成功，于是采用外部zookeeper解决。

```
[chengqichendeMacBook-Pro:bin shengqichen$ jps
41921 NameNode
42354 ResourceManager
42163 SecondaryNameNode
43652 HMaster
42453 NodeManager
42026 DataNode
42842 HQuorumPeer
43979 Jps
43807 HRegionServer
chengqichendeMacBook-Pro:bin shengqichen$
```

参考文章：<https://www.cxybb.com/article/dc666/82866442>

3.进程中有Hmaster和HRegionServer，点击localhost:16010网站后，能连接成功，但是

```
hbase:001:0> create 'test','cf'

ERROR: org.apache.hadoop.hbase.ipc.ServerNotRunningYetException: Server is not running yet
    at org.apache.hadoop.hbase.master.HMaster.checkServiceStarted(HMaster.java:2817)
    at org.apache.hadoop.hbase.master.MasterRpcServices.isMasterRunning(MasterRpcServices.java:1205)
    at org.apache.hadoop.hbase.shaded.protobuf.generated.MasterProtos$MasterService$2.callBlockingMethod(MasterProtos.java)
    at org.apache.hadoop.hbase.ipc.RpcServer.call(RpcServer.java:392)
    at org.apache.hadoop.hbase.ipc.CallRunner.run(CallRunner.java:133)
    at org.apache.hadoop.hbase.ipc.RpcExecutor$Handler.run(RpcExecutor.java:354)
    at org.apache.hadoop.hbase.ipc.RpcExecutor$Handler.run(RpcExecutor.java:334)

For usage try 'help "create"'
```

发现RegionServer并没有run起来，不能解决，于是采用BDkit。