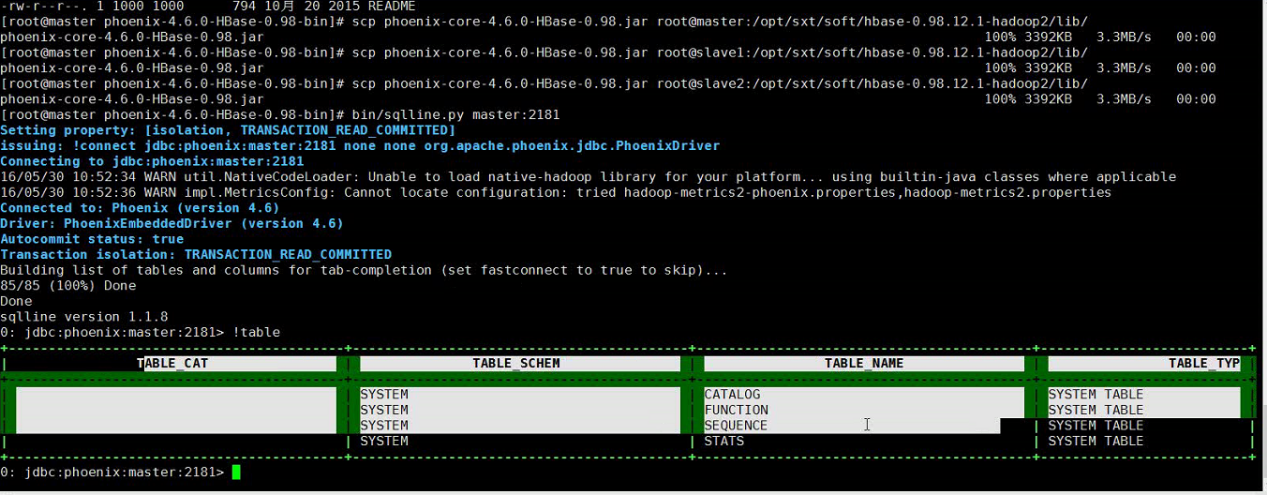
适用范围：

规则的hbase表，类似关系型数据库的表。

一：安装phoenix



启动成功



创建一个表吧~

二：一些类sql语句

Upsert into （插入数据，修改数据）

upsert into test.Person (IDCardNum,Name,Age) values (100,'小明',12);

插入一条数据

upsert into test.person (idcardnum,sex) values (100,'男');

把id为100的性别修改为男

技术博客，有更多的基本操作！

https://segmentfault.com/a/1190000002936080

这个相当于hive的 操作，讲一个表的表数据导入另一个表：

UPSERT INTO TRAVEL SELECT ROWKEY,SP,EP,ST,ET,NEXT VALUE FOR SEQ FROM TMP\_TRAVEL;

三，安装squirrel 可视化工具

双机jar包执行安装。



其中安装完driver和aliases时，最好先test。如果报错，有可能是本地/system32/drivers/hosts没有配置

分页操作：

先创建不带计数器的表。导入数据。

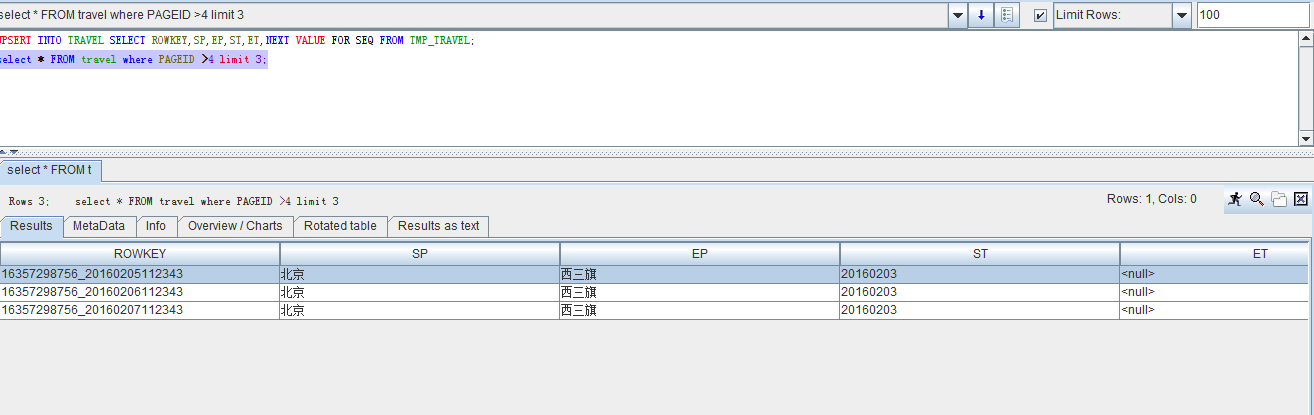
创建带计数器的表，创建计数器表。

使用上边的导入表数据方式：

UPSERT INTO TRAVEL SELECT ROWKEY,SP,EP,ST,ET,NEXT VALUE FOR SEQ FROM TMP\_TRAVEL;

然后查询：select \* from travel pageid>4 limit 3;

查询出从5开始的三条数据，结果如下。



四



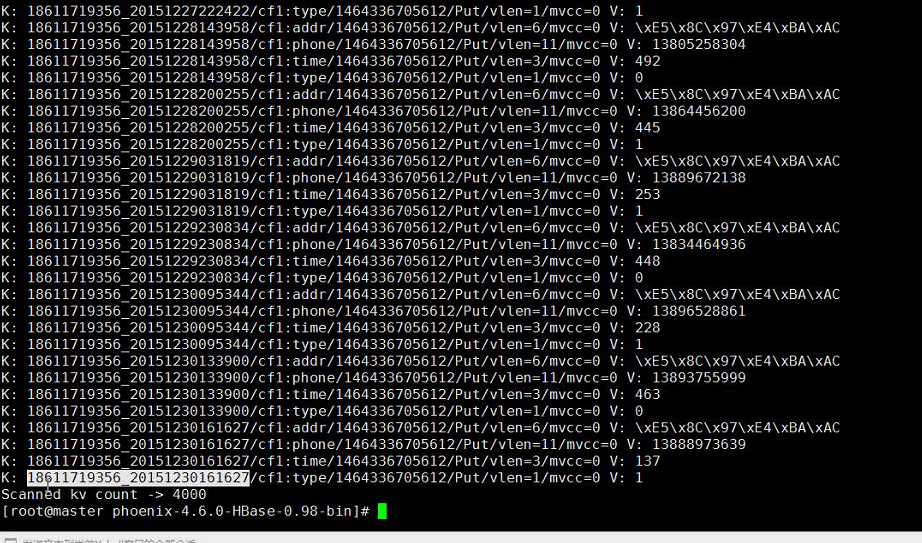
用jdbc链接。Spring的xml文件方式。

1. Protobuf

一个跨语言的将二进制码转化为对象，再把对象转化为二进制的工具。用来把cell合并成对象。这样可以节省磁盘空间。比如：person表，列属性有4个，10000条数据，那么在hdfs中，就存了4000条，每个rowkey都有4条数据。如果把4条，放在同一个对象中，那么就存1000条，节省不少磁盘空间。

安装：

1. tar -zxvf protub.gz
2. ./configure make && make install
3. 在默认路径usr/local中。自己写一个person的文件，然后执行，就会生成一个java文件。通过hbase的java代码创建表插入数据时，调用这个文件里面的方法，就可以插入合并了的对象数据。



在hdfs的存储方式图

4.代码：

略。

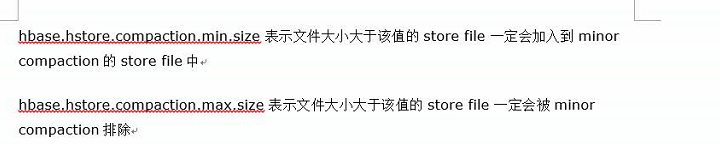
六：hbase的优化！

散列原则！

1. 预分区（region）。
2. 设计row key ，尽量让它散列化。取反，或hash（为了让他们进入不同的分区）

比如phone。Rowkey是时间和号码。最好使用号码\_时间，而不是时间\_号码。

3.使用minor（用的少）



1. 使用major（用的多）

触发它的条件有：

Hbase.hregin.majoucompaction 默认为24小时执行 Major可以开启，在深夜执行，让store file合并。

Major\_compact命令

1. 线程池htablepool

Configuration conf = HBaseConfiguration.create()

HtablePool pool = new HTablePool(conf,10);

设置10个链接，减少多个线程链接握手时消耗的时间。

6.查询时的速度

Hbase遵循的原则，快速从海量数据中，查询到我们所需要的信息。

速度排名：

get rowkey 直接获取

Scan rowkey-rowkey 扫描一片

Scan singleColumnValueFilter 扫描过滤全部

等等等许多，了解一些即可。

六：mapreduce操作hbase

关键语句

TableMapReduceUtil.initTableReducerJob("user", HbaseReducer.class, job);

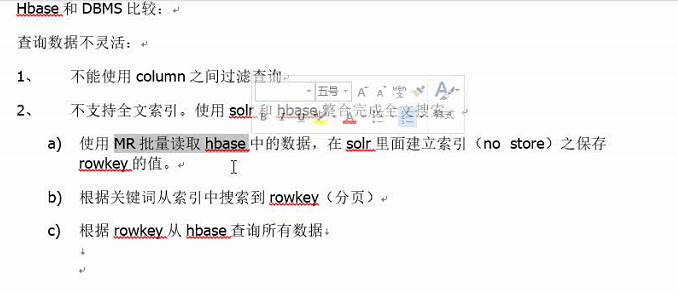
HbaseReducer extends TableReduce

Put put = new Put(Bytes.*toBytes*(key.toString()));  
put.add("cf1".getBytes(),"name".getBytes(),sb.toString().getBytes());

详情看idea中的testHbase2项目

插入一个有意义的总结：

★★★★★



火车票业务:

1. 查询某两地之间的列车有哪些，时间，价钱，中间停靠站，到达时间，开车时间
2. 查询某趟火车所有的停靠站
3. 改签，改签还有哪些车，改签到站
4. 列车表

Rowkey为： 首发地名\_终点地名\_列车号

列族为cf1： 列属性名：中途停靠站名，值：到此站的时间。（含首尾站与时间）

1. 通过1的列车表，可以直接查询到。
2. 改签时，查询当前rowkey的value>当前时间 的列的属性名。

一张表就完成了.......插入数据复杂，但是查询，很方便啊。