第二次实验 熟悉常用的HBase操作

# 实验目的

1. 理解HBase在Hadoop体系结构中的角色；

2. 熟练使用HBase操作常用的Shell命令；

3. 熟悉HBase操作常用的Java API。

# 实验平台

操作系统：Linux

Hadoop版本：2.6.0或以上版本

HBase版本：1.1.2或以上版本

JDK版本：1.6或以上版本

Java IDE：Eclipse

# 实验内容和要求

1. 编程实现以下指定功能，并用Hadoop提供的HBase Shell命令完成相同任务：
2. 列出HBase所有的表的相关信息，例如表名；
3. 在终端打印出指定的表的所有记录数据；
4. 向已经创建好的表添加和删除指定的列族或列；
5. 清空指定的表的所有记录数据；
6. 统计表的行数。
7. 现有以下关系型数据库中的表和数据，要求将其转换为适合于HBase存储的表并插入数据：

学生表（Student）

|  |  |  |  |
| --- | --- | --- | --- |
| 学号（S\_No） | 姓名（S\_Name） | 性别（S\_Sex） | 年龄（S\_Age） |
| 2015001 | Zhangsan | male | 23 |
| 2015003 | Mary | female | 22 |
| 2015003 | Lisi | male | 24 |

课程表（Course）

|  |  |  |
| --- | --- | --- |
| 课程号（C\_No） | 课程名（C\_Name） | 学分（C\_Credit） |
| 123001 | Math | 2.0 |
| 123002 | Computer Science | 5.0 |
| 123003 | English | 3.0 |

选课表（SC）

|  |  |  |
| --- | --- | --- |
| 学号（SC\_Sno） | 课程号（SC\_Cno） | 成绩（SC\_Score） |
| 2015001 | 123001 | 86 |
| 2015001 | 123003 | 69 |
| 2015002 | 123002 | 77 |
| 2015002 | 123003 | 99 |
| 2015003 | 123001 | 98 |
| 2015003 | 123002 | 95 |

同时，请编程完成以下指定功能：

（1）createTable(String tableName, String[] fields)

创建表，参数tableName为表的名称，字符串数组fields为存储记录各个域名称的数组。要求当HBase已经存在名为tableName的表的时候，先删除原有的表，然后再创建新的表。

（2）addRecord(String tableName, String row, String[] fields, String[] values)

向表tableName、行row（用S\_Name表示）和字符串数组files指定的单元格中添加对应的数据values。其中fields中每个元素如果对应的列族下还有相应的列限定符的话，用“columnFamily:column”表示。例如，同时向“Math”、“Computer Science”、“English”三列添加成绩时，字符串数组fields为{“Score:Math”,”Score；Computer Science”,”Score:English”}，数组values存储这三门课的成绩。

（3）scanColumn(String tableName, String column)

浏览表tableName某一列的数据，如果某一行记录中该列数据不存在，则返回null。要求当参数column为某一列族名称时，如果底下有若干个列限定符，则要列出每个列限定符代表的列的数据；当参数column为某一列具体名称（例如“Score:Math”）时，只需要列出该列的数据。

（4）modifyData(String tableName, String row, String column)

修改表tableName，行row（可以用学生姓名S\_Name表示），列column指定的单元格的数据。

（5）deleteRow(String tableName, String row)

删除表tableName中row指定的行的记录。

1. 利用HBase和MapReduce完成如下任务：

假设HBase有2张表，表的逻辑视图及部分数据如下所示：

表 逻辑视图及部分数据

|  |  |
| --- | --- |
| 书名（bookName） | 价格（price） |
| Database System Concept | 30$ |
| Thinking in Java | 60$ |
| Data Mining | 25$ |

要求：从HBase读出上述两张表的数据，对“price”的排序，并将结果存储到HBase中。

# 实验报告

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 《大数据技术基础》课程机房上机实验报告 | | | | |
| 题目： |  | 姓名 |  | 日期 |
| 实验环境： | | | | |
| 实验内容与完成情况： | | | | |
| 出现的问题： | | | | |
| 解决方案（列出遇到的问题和解决办法，列出没有解决的问题）： | | | | |