**北京邮电大学软件学院**

**2020-2021学年第二学期实验报告**

**课程名称：** 大数据原理与技术

**项目名称： 实验 3 HBase 表设计及编程**

**项目完成人：**

**姓名：\_\_\_ \_王衔飞\_\_\_\_**

**学号：\_\_\_ \_\_\_**

**指导教师： 孙鹏飞**

**日 期： 2021年5 月 30 日**

1. **实验内容**

现有以下关系型数据库中的表和数据，请设计适合于以下数据存储的 HBase 的表并给出

表结构（行键、列族、列等）：

学生表（Student）

|  |  |  |  |
| --- | --- | --- | --- |
| 学号（S\_No） | 姓名（S\_Name） | 性别（S\_Sex） | 年龄（S\_Age） |
| 2015001 | Zhangsan | male | 23 |
| 2015002 | 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 fields 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. **实验环境**

Ubuntu21.04

Openjdk 11

Hadoop 3.3.0

Hbase 2.4.3

1. **实验结果和问题分析**

首先启动hadoop(start-all.sh)，然后运行hbase（start-hbase.sh）。执行jps命令，可以发现hbase已经成功运行。

文本

描述已自动生成

在intellij idea 中新建maven项目，在pom.xml中引入hadoop和hbase相关依赖，如下：

图片包含 表格

描述已自动生成

创建类HBaseOperation用于执行数据库相关操作，构造函数如下。

图形用户界面, 文本, 应用程序

描述已自动生成

编写函数createTable(String tableName, String[] fields)，按照实验要求，在创建之前会先判断hbase中是否已经该表，如果存在则先删除该表再重新创建新表。

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

编写函数void addRecord(String tableName, String row, String[] fields, String[] values)。根据行键获取到put对象后，根据用户输入的field进行不同的操作。如果用户在field中设置了列族和列限定符的话（以是否包含冒号进行判断），那么该方法将往指定的表中的单元格添加新的数据。如果用户没有设置列限定符，只设置了列族，那么就往指定表中的列族添加数据。

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

编写函数scanColumn (String tableName,String column)。输出查找到的结果，如果找不到就抛出NullPointerException。

图形用户界面, 文本, 应用程序

描述已自动生成

编写函数void modifyData(String tableName,String row,String column,String value)。将value的值通过put方法放到单元格中实现数据的修改。

图形用户界面, 文本, 应用程序, 电子邮件

描述已自动生成

编写deleteRow(String tableName,String rowKey)。根据行键获得Delete对象后直接调用delete方法即可实现删除行。

图形用户界面, 文本, 应用程序

描述已自动生成

测试：

根据hbase的特点将实验给出的关系数据库转换为如下hbase数据库表：

StudentTable表

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 行键\列族 | studentInfo | | | Math | | ComputerScience | |
| S\_Name | S\_No | S\_Sex | S\_Age | C\_Name | SC\_Score | C\_Name | SC\_Score |

|  |  |
| --- | --- |
| English | |
| C\_Name | SC\_Score |

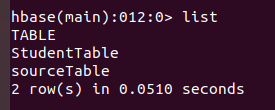
sourceTable表

|  |  |  |
| --- | --- | --- |
| 行键\列族 | sourceInfo | |
| C\_Name | C\_No | C\_Credit |

在main函数中调用创建表和添加记录的方法并运行。

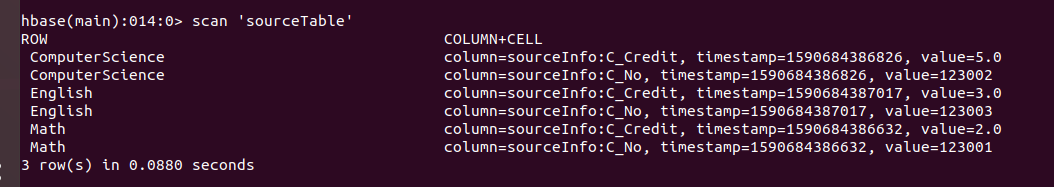
文本

描述已自动生成



文本

描述已自动生成



测试scanColumn方法。

运行代码 hbase.scanColumn("StudentTable", "English");

图形用户界面, 文本, 应用程序

描述已自动生成

输出符合预期。

测试修改数据功能,将mary的英语成绩改成666

hbase.modifyData("StudentTable", "Mary", "English:SC\_Score", "666");

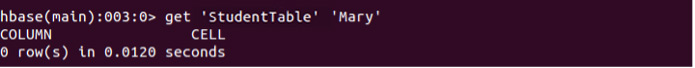
图形用户界面, 应用程序

描述已自动生成

最后测试删除行功能

hbase.deleteRow("StudentTable","Mary")

删除Mary的成绩发现查不到了



实验心得：通过本次实验我了解到了如何对Hbase进行基本操作