

实验 3

第四章 熟悉常用的 HBase 操作

宋 曼

目录

1	实验目的.....	1
2	实验平台.....	1
3	实验内容和要求.....	1

实验 3

第四章 熟悉常用的 HBase 操作

1 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。

2 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

3 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：
 - (1) 列出 HBase 所有的表的相关信息，例如表名；
 - (2) 在终端打印出指定的表的所有记录数据；
 - (3) 向已经创建好的表添加和删除指定的列族或列；
 - (4) 清空指定的表的所有记录数据；
 - (5) 统计表的行数。
2. 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 `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。

3. 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。