

# 《数据库原理》大作业要求

## 一、概述

基于给出的 ThssDB 框架，补充、扩展、实现一个关系型数据库管理系统。

## 二、核心模块

### 2.1 通信模块

系统整体采用客户端/服务器架构，RPC 采用 Apache Thrift。框架中的 `client` 和 `server` 包分别用来管理客户端和服务端，`thrift` 包下的 `rpc.thrift` 文件中给了一个定义返回当前时间（`getTime`）服务的例子，还定义了连接数据库（`connect`）、断开连接（`disconnect`）和执行语句（`executeStatement`）这三个服务，需要同学们自行实现。此外，可以根据自身需要定义其他服务。

#### 基础要求：

1. 实现 `thrift/rpc.thrift` 中的 `connect`、`disconnect` 和 `executeStatement` 服务。

### 2.2 异常处理模块

框架中的 `exception` 包用来管理数据库中涉及到的各种异常类。已给出的两个异常类 `DuplicateKeyException` 和 `KeyNotExistException` 是 B+ 树中用到的。同学们需要在实现下面几个模块时做好异常处理，并在最终的大作业报告中对自定义的异常类予以说明。

### 2.3 存储模块

完成存储引擎，实现对数据的基本管理。框架中的 `schema` 包定义了一些基本概念：

**Column 类：**一张表的某一列的元信息。成员变量包括该列的名字（`name`）、数据类型（`type`）、是否为主键（`primary`）、是否可以为空

(notNull) 以及最大长度 (maxLength, 仅限于 String 类型)。

Entry 类: 某一行记录在表中某一列的数据。

Row 类: 某一行记录在表中所有列的数据。

### 基础要求:

1. 利用 java 的序列化和反序列化实现记录的持久化;
2. 实现对记录的增加、删除、修改、查询;
3. 支持五种数据类型: Int, Long, Float, Double, String。

### 进阶要求:

1. 实现高效的文件存储格式, 如页式存储等。

## 2.4 元数据管理模块

在存储引擎的基础上, 实现元数据管理引擎。框架中的 schema 包定义了一些基本概念:

Table 类: 表的信息。一个 Database 可以有多张 Table。

Database 类: 数据库的信息。用户可以创建多个 Database。

Manager 类: 管理 Database。

此外, 框架中的 index 包涉及索引的元数据, 使用 B+ 树实现。

BPlusTree 类: 用主键 entry 做 key, 用真实的 row 作为 value。

BPlusTreeNode 类: B+ 树结点抽象类。

BPlusTreeInternalNode 类: B+ 树内部结点类。存储 key 数组和 children 数组来支持 B+ 树的索引。

BPlusTreeInternalNode 类: B+ 树叶结点类。存储 key 数组和 value 数组来支持 B+ 树的索引和真实数据的存储, 同时维护一个 next 指针以支持所有叶子结点可以高效遍历。

BPlusTreeIterator 类: B+ 树的迭代器类。通过在 B+ 树最底层的叶子结点维护一个链表就可以高效的实现范围查询和遍历。

### 基础要求:

1. 实现表的创建、删除、修改;
2. 实现数据库的创建、删除、切换;
3. 实现表和数据库的元数据 (有哪些数据库, 数据库里有哪些表, 每个表的结构如何) 的持久化。
4. 重启数据库时从持久化的元数据中恢复系统信息。

## 2.5 查询模块

实现查询引擎，包括 SQL 解析和查询执行两部分。其中，`parser` 包给出了利用 `antlr4` 实现 SQL 解析的例子。框架中的 `query` 包提供了一些可能用到的基础类：

**MetaInfo 类：**存储一个表的名称和所有列，用来在 `Queryresult` 中构建用户想要获取的列的索引。

**QueryResult** 类：选择运算中选择某些列或全部列的结果，需要对歧义列名或不存在列名、不存在表名等错误情况进行捕捉。

**QueryTable** 类：选择运算中选择表的方法类，用迭代器的方式从当前表中不断获取新的行并判断其是否满足 **where** 中条件。

### 基础要求:

1. 支持下列查询语句（不区分大小写）：
  - 创建表：CREATE TABLE *tableName*(*attrName1* *Type1*, *attrName2* *Type2*,..., *attrNameN* *TypeN* NOT NULL, PRIMARY KEY(*attrName1*)), 实现 “NOT NULL”和 “PRIMARY KEY”这两个关键字, *Type* 为 Int, Long, Float, Double, String（必须给出长度）之一。主键仅在某一列定义。  
示例：CREATE TABLE *person* (*name* String(256), *ID* Int not null, PRIMARY KEY(*ID*))
  - 删除表：DROP TABLE *tableName*  
展示某张表的模式信息(每一列的名字及其数据类型，格式自定)：SHOW TABLE *tableName*
  - INSERT INTO [*tableName*(*attrName1*, *attrName2*,..., *attrNameN*)] VALUES (*attrValue1*, *attrValue2*,..., *attrValueN*), 字符串需要用单引号包围。  
示例：INSERT INTO *person* VALUES ('Bob', 15)  
INSERT INTO *person*(*name*) VALUES ('Bob')会提示字段 *ID* 不能为空。
  - DELETE FROM *tableName* WHERE *attrName* = *attValue*
  - UPDATE *tableName* SET *attrName* = *attrValue* WHERE *attrName* = *attrValue*
  - SELECT *attrName1*, *attrName2*, ... *attrNameN* FROM *tableName* [ WHERE *attrName1* = *attrValue* ]
  - SELECT *tableName1*.*AttrName1*, *tableName1*.*AttrName2*..., *tableName2*.*AttrName1*, *tableName2*.*AttrName2*,... FROM *tableName1* JOIN *tableName2* ON *tableName1*.*attrName1* = *tableName2*.*attrName2* [ WHERE *attrName1* = *attrValue* ]

上述语句中，Where子句至多涉及一个比较，并且关系为“<”，“>”，“=”

<=",">=","=","<>"之一。Select 子句不包含表达式。Join 至多涉及 2 张表，On 子句的限制同 Where 子句。

进阶要求：

1. 应用课程中介绍的查询优化技术
2. 支持多列主键；
3. where 条件支持逻辑运算符（and/or）；
4. 实现三张表以上的 join；
5. 实现 outer join 等其他类型的 join；
6. 其他标准 SQL 支持的查询语法。

## 2.6 事务与恢复模块

实现简单的事务及恢复功能，支持小规模并发。

基础要求：

1. 实现 begin transaction 和 commit；采用普通锁协议，实现 read committed 的隔离级别即可；
2. 实现单一事务的 WAL 机制，要求实现写 log 和读 log，在重启时能够恢复记录的数据即可。

进阶要求：

1. 实现多事务的并发和恢复机制；
2. 实现更高级的隔离级别（repeatable read/serializable）；
3. 实现 rollback、savepoint 等功能。
4. 实现 2PL 或 MVCC 协议。

## 2.7 说明

框架所提供的类及方法均为建议和参考，大家可以自行选择利用或者不利用或者部分利用这些类及方法，非常鼓励大家设计出性能更优的架构。

## 三、作业要求

1. 组队人数：2~3 人
2. 逐步完成系统的各个模块，逐次提交设计文档，最终提交需要提交设计文

档、展示 PPT 及源代码。

提交项	截止日期
存储模块	5 月 3 日
元数据管理模块	5 月 10 日
查询模块	5 月 21 日
事务与恢复模块	5 月 31 日
最终提交	6 月 7 日

3. 项目检查（机动安排）

评分依据（满分 100 分，完成额外功能可以加分，最高不超过 110 分）：  
以教材给出的大学数据库为测试的 **schema**

基础要求完整、正确（90 分）；

以 HSQLDB 为基准，评测性能（10 分）；

进阶要求，酌情给分（不超过 10 分）。

4. 加分项：所有其他的扩展（支持更多的 SQL、性能优化、权限认证、触发器、约束等）都可以作为加分项。

5. 课堂展示（第 16 周）：PPT（系统介绍）。