**《数据库管理系统实现》实验指导**

韩京宇

韩丽萍

2025年5月20日

**实验1：DBMS存储实现**

1. 根据mini\_base\_blank\_python3.8的注释，了解整个程序结构，知道各个模块的作用。
2. 确保在自己的机子上能运行，找到main\_db.py中的入口函数。
3. 第一次实验的具体任务如下：
4. 存储下面三个表的模式和实例

students

|  |  |  |  |
| --- | --- | --- | --- |
| s\_id | name | gender | age |
| s01 | Tom | male | 19 |
| s02 | Jack | male | 20 |
| s03 | Lily | female | 17 |

courses

|  |  |  |  |
| --- | --- | --- | --- |
| c\_id | title | semester | credit |
| c01 | database system | fall | 3 |
| c02 | web | spring | 2 |

takes

|  |  |  |
| --- | --- | --- |
| s\_id | c\_id | score |
| s01 | c02 | 90 |
| s02 | c01 | 89 |
| s01 | c02 | 49 |

1. 三个表模式存储到all.sch文件中。
2. 每张表对应一个实例文件，命名是student.ins, course.ins,takes.ins。或者一个实例文件也可以。
3. 在mini\_base\_blank\_python3.8版本上完善代码，对应schema\_db.py和storage\_db.py文件。如果要缓存，完善head\_db.py。
4. 入口函数在main\_db.py，进入各个分支，对上述表中的数据进行查询和增、删、改（需要对入口函数进一步完善）
5. 所有自己新添的函数要有注释，说清楚输入、输出和功能，并对关键步骤和算法做注释。修改的代码文件的开头注释自己的英文名字和联系邮箱，方便代码维护。
6. 一个组完善一份代码，分组由课代表确定后在一周内发给老师。
7. 撰写《数据库管理系统原型的软件说明书》的对应部分。

**实验2：DBMS的SQL语句解析**

1. 支持下列单表查询的SQL语句解析和处理

select \* from students;

select name from students;

select name from students where age=19;

1. 选做支持如下SQL语句

create table courses(c\_id char(20), title char(20),semester char(20),credit integer);

insert into courses values(‘c01’,’database system’,’fall’,3);

delete from courses;

update courses set semester=’spring’ where c\_id=’c01’;

drop table takes;

三、实验步骤

1、完善query\_plan\_db.py,parser\_db.py,lex\_db.py,node\_db.py的对应模块，每个模块的具体功能见注释。

2、在main函数中测试第5选项对应的select-from-where句型。

3、如果实现了选做内容，在main函数中添加对应的入口函数，选项从8开始。

4、参见附件，撰写《数据库管理系统原型的软件说明书》的对应部分。

四、关键技术环节提示

1、涉及模块

lex\_db.py ：词法分析，识别SQL语句中各个符号（ token）

parser\_db.py :语法分析，形成语法树

query\_plan\_db.py :逻辑查询计划生成和执行

main\_db.py:从第5个执行分支进入

2、lex\_db.py模块简介

复习编译原理中词法分析器lex的用法

功能：输入select...from...where语句，识别对应的tokens

定义的tokens：

tokens=('SELECT','FROM','WHERE','AND','TCNAME','EQX','COMMA','CONSTANT','SPACE’)

SQL例句：

select f1,f2 from GOOD where f1='xx' and f2=5

程序识别的token如下

LexToken(SELECT,'select',1,0)

LexToken(TCNAME,'f1',1,7)

LexToken(COMMA,',',1,9)

LexToken(TCNAME,'f2',1,10)

LexToken(FROM,'from',1,13)

LexToken(TCNAME,'GOOD',1,18)

LexToken(WHERE,'where',1,23)

LexToken(TCNAME,'f1',1,29)

LexToken(EQX,'=',1,31)

3、parse\_db.py模块简介

学习yacc（Yet Another Compiler Compiler）的用法

功能：为select...from...where语句创建一棵语法树

原理：

（1）逻辑上通过巴克斯范式递归定义语法

（2）每个函数的开头是语法规则，例子：

'SFW : SELECT SelList FROM FromList WHERE Cond'

（3）每个函数根据语法规则解析传入的token，同时创建树节点，用到了common\_db.py中的Node类定义

4、query\_plad\_db.py模块简介

construct\_logical\_tree() 语法树转换为查询计划树

execute\_logical\_tree() 查询执行

备注：

（1）这两个函数在main\_db.py的第5个分支调用

（2）查询处理涉及四个全局变量，即lex解析器、yacc解析器、全局语法树、逻辑查询计划，它们都在common\_db.py中定义，在main\_db.py中调用。

**实验3：事务持久性实现**

1. 在现有的系统上增加事务持久性的保障。即确保系统出问题时，在出问题前提交的事务对数据库的修改有效。针对insert语句实现持久性保证，也可以实现针对update语句的持久性保证。
2. 核心技术：增加日志文件，根据“提交规则”和“先记后写规则”将数据写入数据库文件或者日志文件，详细参见PPT的事务恢复部分。具体要求

1、增加前像文件、后像文件、活动事务表和提交事务表。其中前像和后像文件是两个一数据块为读写单位的堆文件。

2、每次进行insert或者update时，按照提交规则和先记后写规则写入外存。

3、提交规则的内容：后像必须在事务提交前写入非易失存储器，即（完全）写入数据库或运行记录中。

4、先记后写规则：如果后像在事务提交前写入数据库，则必须首先把前像写入运行记录。

1. 《数据库管理系统原型的软件说明书》中写清楚实现日志的关键数据结构和技术步骤。如果可能，实测事务持久性功能。

**附件1：原型作业分组和评分基本原则**

**一、分组原则**

软件大作业对应平时作业和三个课内实验，占平时成绩的50分，评分方法如下。

1、每个班分成10组，每组最多四个人，超过四人的组要拆分，也可以自己协调进行调整。课代表在第一次实验后把确认后的分组情况汇总成EXCEL表汇报给老师，老师会在QQ群中公布分组。

2、每个小组在最后一次实验课后、考试前一周提交一份软件，包括程序和说明文档。老师根据软件质量给出ABCD四个等级，分别对应47，42，37，32四个分数。

3、每个组提交软件时，同时在程序和说明文档的封面标注加减分模式。假设W代表将来老师给的评分等级，加减分模式是组内协调出的结果，比如有4个同学，下面是几个常见的加减分模式（不限于这些，只要满足后面的规则）。

模式1：张三（W+3）,李四（W-1）,王五（W-1）,赵六（W-1）。表示如果将来得A等级，则张三得50分，李四得46分，王五得46分，赵六得46分。

模式2：张三（W+1）,李四（W+1）,王五（W-1）,赵六（W-1）

模式3：张三（W）,李四（W）,王五（W）,赵六（W）

模式4：张三（W+2）,李四（W+1）,王五（W-1）,赵六（W-2）

每个组的加减分必须满足下列条件：

（1）加减分必须是整数。

（2）W后的正数和负数的代数和等于0。

（3）最大差距小于等于4分。

**二、评分依据**

1、在mini\_base\_blank框架下，实现一个可运行的DBMS原型，选择实现如下的一个或几个模块。

（1）数据的增加、删除或修改（可不支持或支持SQL）

（2）数据的查询（支持或不支持SQL）

（3）普通索引的实现，比如B树，哈希等（要测试用索引和不用索引的对比效果）

（4）日志的实现

2、最后一堂课前每个组把程序打包给老师，包名是“minibase-班级号-组号-同学1-同学2-同学3-同学4.zip”。每个包中包含一个word软件说明书《数据库管理系统实现原型作业》，说明书至少3页，含封面、目录、代码完善模块描述。

（1）封面必须标注班级、组号、每个同学的学号姓名和评分模式对应项。

（2）目录建议根据内容自动生成。

（3）代码完善概述列出在mini\_base\_blank基础上的新增加的或修改的各个模块的内容，列出每个模块的功能，每个模块的关键算法，新增部分或修改部分在文件中的行范围，关键函数或类的输入、输出和功能，也可以列出性能测试数据。

3、任课教师根据软件说明书和程序进行评分，分成ABCD四个等级（优良中差）。其中程序质量得分，参照如下维度进行评分：程序整体功能涵盖度、函数功能全面、函数功能正确、注释输入、注释输出、注释程序步骤、出错处理、结果显示和性能分析。

4、如果要得优，提前告诉老师，申请现场演示程序并回答问题。