

数据库课程项目

一、任务说明

实现一个单用户的关系数据库管理系统。该项目分为四个功能模块：

1. 记录管理模块：该模块是DBMS的文件系统，管理存储数据库记录以及元数据的文件。该模块依赖于我们预先给定的一个页式文件I/O系统，在此基础上扩展而成。
2. 索引模块：为存储在文件中的记录建立B+树索引，加快查找速度。
3. 系统管理模块：实现基本的数据定义语言（DDL），实现解析器来解析命令行。
4. 查询解析模块：解析SQL语句，能将输入的SQL语句解析成关系代数表达式，并生成查询执行计划，访问文件系统执行查询，输出查询结果。

其中，记录管理模块、系统管理模块和查询解析模块为必做内容，索引模块为选做内容。在上述功能的基础上，同学们可自行查阅资料，对该系统进行个性化的功能扩展及性能优化，内容包括但不限于：

1. 查询优化：基于对查询计划代价的估计，为给定查询选择最有效的查询执行计划。
2. 支持属性域约束和外键约束。
3. 创建数据表时支持更多的数据类型。例如decimal,date等。
4. 支持三个或以上表的连接。
5. 支持更多SQL语句。例如聚集查询AVG,SUM,MIN,MAX,GROUP BY等。
6. 支持模糊查询。例如LIKE关键字以及“%,”,”?”等通配符。
7. 提供类似MySQL front的图形化UI。

上述功能扩展均属于选做内容。如果同学们实现了个性化的扩展，请在项目报告中详细叙述实现内容及原理，并在当面验收时说明，以便助教评分。

二、作业要求

1. 本项目要求随着课堂进度的进行，逐步完成系统的各个模块，并按照指定的截止时间提交相应模块的代码备查。截止时间初步确定如下：

系统模块	截止日期
记录管理模块	10月26日
系统管理模块	11月16日
查询解析模块	12月14日

请在截止日期当天24:00前提交各模块的代码至网络学堂。

除此之外，要在期中（第11周）进行一次项目进度汇报，内容包括目前进展以及存在的问题等；在期末（第16周）进行项目验收并提交项目报告。

2. 本项目的评分标准如下：

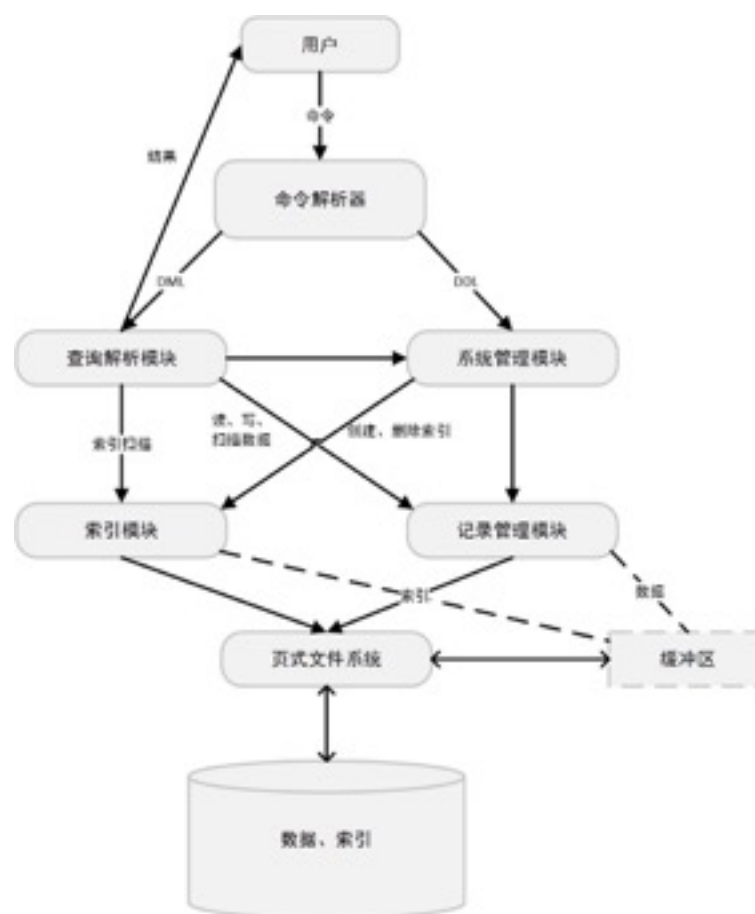
考核项目	占分比重
日常提交	10%
期中报告	10%
项目报告文档	10%
期末验收	70%

具体说明：

- (1) “日常提交”部分的分值由同学们在指定截止时间内提交相应模块的代码的情况决定，旨在鼓励大家按照进度完成工作，期末验收以最后一次提交的代码为准，与本部分提交过的代码无关。
- (2) “期中报告”部分的分值由同学们期中汇报的表现决定。

- (3) 期末在提交代码的同时也要提交项目报告文档，其内容包括但不限于系统架构设计、主要模块设计原理、主要模块接口说明、实验结果、小组分工、参考文献等。
- (4) “期末验收”部分的成绩由最后一次提交的代码的运行结果，以及当面提问的结果为准，与“日常提交”环节提交的代码无关。
- (5) 期末验收原则上必须通过测试，否则可能导致挂科。
3. 本项目的开发过程中允许同学们参考和使用已有的开源项目和工具，但必须在最后的报告中详细指所借鉴的内容，并注明详细出处。
4. 本项目的最终验收方式为当面提问以及黑盒测试。同学们提供命令行接口并输入助教提供的SQL语句，系统输出正确的运行结果。附件中给出的数据模式以及SQL语句仅供同学们验证基本正确性，验收时会使用更多SQL语句进行测试。
5. 有关大作业的系统架构设计，有疑问的同学可以参考一下斯坦福大学CS346的课程项目 <http://web.stanford.edu/class/cs346/>

三、系统架构



五、各模块介绍

本章介绍了各模块的详细信息以及系统设计的一些建议。其中“建议”内容仅供参考，请同学们批判性的接受。

1. 记录管理模块

数据库管理系统中需要实现的第一个模块是记录管理模块，该模块需要实现一些类和方法来对存储记录的文件进行管理。该模块必须实现的基本功能如下，可以直接调用我们提供的页式文件I/O系统中的方法来实现下列功能。

- 新建文件、删除文件、打开文件、关闭文件。
- 插入记录、删除记录、更新记录、获取属性值满足特定条件的记录。

建议：

- (1) 每个文件保存的记录是定长的，不同文件保存的记录长度可以不同。
- (2) 文件的第一页可以用来存储和文件相关的信息，比如文件中记录的长度、文件每一页存储的记录个数、文件中的记录个数。
- (3) 每一条记录都应该有RID来作为该记录的唯一标示。该标示不应随着记录的插入、删除、更新而改变。
- (4) 当插入一条记录时，对文件中的各页依次扫描来查找空闲的位置会导致效率低下。我们建议使用一个位数组来记录每页的空闲空间。
- (5) 扫描文件获取属性值满足特定条件的记录是十分关键的，后续SQL语句的解析会依赖这一操作。

2. 索引模块

在该模块需要为文件中的记录建立B+树索引以提高查询速度，每一个文件会与多个索引相关联（一个属性一个索引）。这些索引也存储在文件中，因此你可以直接调用我们提供的页式文件I/O系统中的方法来实现部分功能。该模块必须实现的基本功能如下：

- 创建索引、删除索引、打开索引、关闭索引
- 索引中插入节点、删除节点、获取属性值满足特定条件的节点。

3. 系统管理模块

该模块以及查询解析模块要求实现解析器来解析用户的命令并执行一系列相关的操作。在该模块中必须解析实现的用户命令如下：

- 创建数据库 `CREATE DATABASE DBname`
删除数据库 `DROP DATABASE DBname`
切换数据库 `USE DATABASE DBname`
列出现有的所有数据库以及其包含的所有表名 `SHOW DATABASE DBname`
DBname是用户命名的数据库名称。本项目中的数据库管理系统要管理多个数据库，实现在各个数据库之间切换。
- 创建表 `CREATE TABLE tableName(attrName1 Type1, attrName2 Type2,..., attrNameN TypeN NOT NULL, PRIMARY KEY(attrName1))`，注意要实现“NOT NULL”和“PRIMARY KEY”这两个关键字。
删除表 `DROP TABLE tableName`
列出现有的所有表以及其模式信息 `SHOW TABLE tableName`

建议：

- (1) 创建、删除以及切换数据库可以通过创建可执行文件并分别命名为create、drop以及use来实现。所有和数据库DBname相关的文件应保存在和这个数据库相关的目录中，这个目录可以命名为DBname。执行命令create时通过mkdir命令创建这个目录，执行命令drop时删除这个目录以及目录下的所有文件，执行命令use时切换当前目录到DBname目录下。
- (2) 创建删除表可以通过在指定的数据库目录下新建和删除文件来实现。

4. 查询解析模块

该模块要求实现解析器来解析用户的命令并执行一系列相关的操作。在该模块中必须解析实现下列四个常见语句的常见语法选项，

- `INSERT INTO [tableName(attrName1, attrName2,..., attrNameN)] VALUES (attrValue1, attrValue2,..., attrValueN)`

- DELETE FROM *tableName* WHERE *whereClauses*
- UPDATE *tableName* SET *tableName.attrName* = *attExpression*
- SELECT *tableName.AttrName* FROM *tableName* WHERE *whereClauses*

其中SELECT要支持2个表的连接操作。SET 子句的表达式包括四则运算，WHERE子句的条件包括常见条件表达式，NULL判断，整数和字符串比较，模式匹配等。需要考虑主键约束以及数据类型的约束。

建议：

- (1) 这一模块会调用前三个模块的函数，是本数据库管理系统的核心模块。实现这一模块的功能有多种途径，请务必仔细选择以免影响系统效率。请对输入命令进行检查并尽可能的考虑多种情况，例如主键为空，数据类型不符，外键依赖等。
- (2) 解析完用户的命令后，首先转换成关系代数表达式，并生成查询执行计划，然后访问文件系统执行查询，输出查询结果。
- (3) 请注意系统的鲁棒性和各种边界条件的处理。