

数据库系统概论项目报告

潘子睿
2020010960

余任杰
2020010966

2022 年 12 月 9 日

目录

1	项目架构	2
1.1	支持功能	2
1.1.1	数据结构	2
1.1.2	SQL语句	2
1.1.3	其他	2
1.2	代码测试	2
2	编译与运行	2
3	系统功能	2
3.1	页式文件系统	2
3.2	记录管理	3
3.3	索引管理	3
3.4	系统管理	3
3.5	SQL解析	3
4	参考资料	3

1 项目架构

SimDB是一个简单的关系型数据库系统，其能够支持一些基本的SQL语句，完成对数据库的插入、删除、更新、查找等操作。项目的最底层是一个页式文件管理系统，在其上构建了记录管理和索引管理两个模块，分别用来维护数据库中的一条条记录以及在某些记录上建立的索引。系统管理模块通过调用记录管理和索引管理两个模块的接口，实现了具体操作数据库的各项功能。最后，通过Antlr4解析SQL指令，并将其交给系统管理模块执行。

本项目源代码目前位于<https://github.com/pzrain/SimDB>，并将在本课程结课后开源。

1.1 支持功能

1.1.1 数据结构

数据库当前支持的数据结构包括：

- 整型（INT）
- 浮点型（FLOAT）
- 字符串型（VARCHAR）

其中VARCHAR仅支持定长字符串。

1.1.2 SQL语句

1.1.3 其他

1.2 代码测试

2 编译与运行

本项目基于CMake进行自动构建。需要Antlr4依赖，以及编译器支持C++17特性。使用时，将CMakeLists.txt中的ANTLR4_RUNTIME_DIRECTORY的值设置为Antlr4运行时库antlr4-runtime.h所在的目录，并将ANTLR4_RUNTIME_SHARED_LIBRARY的值设置为Antlr4运行时动态链接库libantlr4-runtime.so。

成功配置好Antlr4后，在项目根目录下执行以下命令：

```
./run.sh -c
```

即可自动进行编译并运行，生成的可执行文件SimDB会被放置在./bin目录下。

3 系统功能

3.1 页式文件系统

本部分代码位于./src/filesystem，需要注意的是，本部分直接使用了课程实验文档附

录提供的参考实现代码。

数据库是被设计用来存储大量数据的系统，数据库中一个文件的大小甚至可能超过计算机的内存。因此，需要一个页式文件管理系统来管理数据库的各个文件，以及一个缓存机制，将操作的多个页面缓存在内存中，只在需要时进行替换和写回，以提高读写的效率。参考实现中使用的替换算法为最近最少使用算法（LRU）。

3.2 记录管理

3.3 索引管理

3.4 系统管理

3.5 SQL解析

4 参考资料

以下列出了本项目完成过程中所参考的部分资料，包括课程的实验文档、往届的实现，以及相关网站等。

- 往届实现的数据库项目
- CS346 Redbase project
- B^+ 树的定义与实现
- 使用antlr4解析SQL
- 课程实验文档