

**系统能力综合培养实验报告**

|  |  |  |  |
| --- | --- | --- | --- |
| 学 院： | 计算机学院 | | |
| 专 业： | 计算机科学与技术 | | |
| 班 级： | ACM1601班、计算机1609班 | | |
| 学 号： |  | U201614774 |  |
| 姓 名： | 黄佳汇 | 杨祖容 | 王通 |
| 指导教师： | 谢美意、左琼 | | |

|  |  |
| --- | --- |
| 分数 |  |
| 教师签名 |  |

2019年 月 日

目 录

[1 实验任务概述 1](#_Toc493799738)

[2 记录管理模块 2](#_Toc493799739)

[2.1 模块概述 2](#_Toc493799740)

[2.2 模块设计 3](#_Toc493799741)

[2.3 模块实现 3](#_Toc493799742)

[2.4 实验过程 3](#_Toc493799743)

[3 系统管理模块 3](#_Toc493799744)

[3.1 模块概述 3](#_Toc493799745)

[3.2 模块设计 3](#_Toc493799746)

[3.3 模块实现 3](#_Toc493799747)

[3.4 实验过程 3](#_Toc493799748)

[4 查询处理模块 3](#_Toc493799749)

[4.1 模块概述 3](#_Toc493799750)

[4.2 模块设计 3](#_Toc493799751)

[4.3 模块实现 3](#_Toc493799752)

[4.4 实验过程 3](#_Toc493799753)

[5 索引管理模块 4](#_Toc493799754)

[5.1 模块概述 4](#_Toc493799755)

[5.2 模块设计 4](#_Toc493799756)

[5.3 模块实现 4](#_Toc493799757)

[5.4 实验过程 4](#_Toc493799758)

[6 实验总结 4](#_Toc493799759)

[6.1 工作总结 4](#_Toc493799760)

[6.2 改进方案 4](#_Toc493799761)

[附录 4](#_Toc493799762)

（目录统一采用小四号宋体/Times New Roman和1.25倍行距）

# 实验任务概述

## 1.1 实验目的

1)了解数据库、操作系统、编译、数据结构等专业知识在关系数据库管理系统实现中的综合运用；

2)了解关系数据库管理软件的体系结构和设计方法；

3)了解关系数据库中数据和元数据的组织方法；

4)掌握关系数据的底层存储实现技术；

5)掌握基本关系操作在关系数据上的实现算法；

6)掌握基本SQL语句的实现方法；

7)掌握基本的查询优化技术。

## 1.2 实验要求

基于HustBase系统的总体设计架构，根据预先给定的系统框架、部分模块和接口要求，设计并实现系统中的其余模块功能，完成一个具有基本数据定义、数据操纵和数据查询功能的单用户关系数据库管理系统。

实验课时安排为四周，按周次划分为四个阶段。除去已预先提供的用户界面、语法分析和页面管理模块，要求学生自己设计实现记录管理、系统管理、查询处理和索引管理模块以及所有模块的联调，完成整个HustBase系统的开发。具体课程内容安排如表2.1所示。

表2.1 实验内容安排

|  |  |
| --- | --- |
| **周 次** | **内 容 安 排** |
| 第一周 | 学习HustBase的系统结构和工作原理，在已提供模块的基础上，开始记录管理模块、系统管理模块和索引管理模块的编码。 |
| 第二周 | 完成建库、建表功能的编码和联调；开始数据查询模块的编码。 |
| 第三周 | 完成数据操纵和数据查询功能的编码和联调。 |
| 第四周 | 完成索引管理功能，并与其他相关模块结合，完成系统整体联调。 |

HustBase系统的开发环境为Windows 操作系统及 Visual Studio 2010，开发语言为C语言。为保证良好的模块独立性、便于共同开发及阶段性检查，每个模块必须按文档规定的格式对外提供调用接口，但是模块内部的实现方案不做统一规定，由学生自行设计。

具体要求：

1. 实验内容采取2-3人小组的形式合作完成，每个学生在小组中均应承担合理的工作量，基本原则为“共同设计，分工编码”。
2. 考核方式为现场演示及答辩和测试程序验证，最终成绩根据系统完成情况、演示答辩情况和实验报告质量综合评分，同组学生的成绩相同。

课程结束后，以小组为单位提交系统的源码、编译后的程序及实验报告。

## 1.3 HustBase体系结构介绍

HustBase是一个用于实践教学的简化版数据库管理系统，其设计目标是支持简单的SQL语句，提供基本的数据定义、数据操纵、数据查询和查询优化功能。为了便于学生在为期四周的实验周期内掌握系统框架并完成实验内容，与实际的商用DBMS相比，HustBase做了大量的功能简化，舍弃了如多用户、事务、故障恢复等常见特性，支持的SQL语法也非常有限。

HustBase的体系架构如图3.1所示。系统由页面管理、记录管理、索引管理、系统管理、查询处理、语法分析和用户界面等模块构成。各模块功能如下：

（1）页面管理：数据库中所有的数据、元数据和索引数据均以文件的形式存储在磁盘上。Windows操作系统支持的文件结构为流文件，而数据库中的数据是以记录为单位来进行存取的，为了实现对磁盘文件的高效I/O，需要将流文件转换为记录文件。这个过程分为两步：首先，将连续的流文件划分为由若干个固定大小的页面，将其抽象为分页文件；其次，将分页文件的页面划分为若干个固定大小的记录插槽或索引项插槽，将其抽象为记录文件或索引文件。本模块提供第一步所需功能，即将流文件抽象为分页文件，并提供以页面为单位的文件读写接口，具体包括：创建、销毁、打开和关闭分页文件；遍历指定文件中的所有页面；从指定文件中读取一个特定页面；在指定文件中添加、删除及修改页面等。

（2）记录管理：系统表和数据表中的数据均以记录为单位进行存取。本模块在页面管理模块的基础上，将文件中每个页面又划分为若干个固定大小的记录插槽，实现记录的插入、删除、修改和查找，为上层的数据操纵及数据查询功能提供支持。具体包括：创建、销毁、打开和关闭记录文件；插入、删除、修改记录；查找符合指定条件的记录等。

（3）索引管理：利用索引，系统可以为查询提供快速访问路径。本模块在页面管理模块的基础上，将文件中每个页面又划分为若干个固定大小的索引项插槽，提供对索引项的管理。索引文件与记录文件的不同之处在于，索引采用B+树结构来组织索引项，以便提高关键字查找的效率。具体包括：创建、销毁、打开和关闭索引文件；索引项的插入、删除、修改以及查找等。



图1.1 HustBase系统体系架构图

（4）系统管理：本模块提供对数据定义（CREATE/DROP TABLE，CREATE/DROP INDEX）和数据操纵（INSERT、DELETE、UPDATE）功能的支持。系统管理模块的实现依赖于记录管理模块和索引管理模块。

（5）查询处理：本模块用于提供对数据查询（SELECT）功能的支持，并向客户端返回最终的查询结果集。查询处理模块的功能依赖于记录管理模块和索引管理模块。

（6）语法分析：本模块提供对SQL命令的语法分析，并将分析结果以语法树的形式提供给系统管理模块和查询处理模块，以便其做进一步的分析和处理。

（7）用户界面：本模块为用户提供操作界面，接收用户输入的SQL命令，并向用户输出SQL命令的执行结果。

在图3.1显示的各模块中，蓝色模块的功能代码已经提供，红色模块则需要学生自行设计实现。学生根据自己的实际能力，也可以对蓝色模块的代码进行一定程度的修改以扩展系统功能，但原则上不应修改本文档中规定的模块接口。

## 1.4 小组成员分工

黄佳汇：任务A

杨祖容：任务B

王通 ：任务C

简要介绍本实验的实验目的、HustBase系统、实验具体任务要求、小组成员分工等信息。

**关于正文排版要求：**

正文统一采用小四号宋体/Times New Roman和1.25倍行距，段前、段后均0磅间隔。

正文应论述清楚，文字简练通顺，插图简明，书写整洁。文中图、表按制图要求绘制。

**关于正文章节标题的排版要求：**

报告章标题称为一级标题，章内小节标题依次分为二级标题、三级标题等。一级标题的编号用数字1，2，…编制；二级标题的编号用1.1，1.2，…编制；三级标题的编号用1.1.1，1.2.1，… 编制；四级及以后各级标题可依此类推。建议标题不超过3级（如1.1.1），超出部分可根据需要使用(1)，①，A，a)，…等形式描述。

标题编号与标题文字之间均用空格隔开，如：“1 引言”、“2.1 需求分析”。报告正文的一级标题（章）须另起一页居中排版。

1级标题宋体小三号字加黑，单倍行距，段前、段后均12磅间隔。

2级标题宋体四号字加黑，1.25倍行距，段前6磅，段后0磅间隔。

3级标题宋体小四号字加黑，单倍行距，段前3磅、段后8磅间隔。

**关于正文中的图的排版要求：**

正文中所有插图要求图面整洁，布局合理，线条粗细均匀，圆弧连接光滑，尺寸标注规范。所有曲线、图表、线路图、流程图、程序框图、示意图等不准徒手画，必须按国家规定标准或工程要求采用计算机或手工绘制。

所有插图均应有图号和图名。图号按章编，如第2章的图为图2.1、图2.2、…，第3章的图为图3.1、图3.2、…等。图名是插图的名称，扼要概括图的内容，字数不宜太多。图号和图名写在图下方，并相对于图居中排版。少数图有图注，图注写在图下面且字号应比图号、图名的字小一号，图名和图注后面均不加标点符号。

所有插图均应在正文中予以引用。引用某插图时，一般写为“…见图x.y”或“图x.y是…”。正文中的插图均须安排在文中第一次引用到该图的正文下面，要求先见文，后见插图，且图一般不跨页绘制。

图中文字、图号和图名，统一采用小五号宋体。

**关于正文中的表的排版要求：**

表格由表号、表名、表头、表身等组成。表号按章编，如第2章的表为表2.1、表2.2、…，第3章的表为表3.1、表3.2、…等。表名是表格的名称，扼要概括表的内容，字数不宜太多。表号、表名放在表的正上方，相对于表体居中排版。表号及表名后不加标点。表头包括栏头、行头，与表身一起构成表格的主体。表中的竖称为栏，横格称为行。表身的内容，一般包括：数据、文字、公式和表图等。表内的数据对应位要对齐。少数表有表注，表注写在表下面且字号应比表号、表名的字小一号。

所有表格均应在正文中予以引用。引用某表格时，一般写为“…见表x.y”或“表x.y是…”。表格应尽量靠近正文的叙述，一般应先见文，后见表，表不跨节。表格允许转页。表格转页部分可以不写表号和表名，但要重复书写表头，并在表头右上角写“（续）”字标注。

表中文字、图号和图名，统一采用小五号宋体。

**关于正文中公式的排版要求：**

公式一般另行居中写，公式末不加标点。若公式前有文字，如例、解、证、假定等，文字顶格写，公式仍居中写。一行写不下时，公式允许转行。公式转行需处理得当，做到既意义正确，又使版面美观匀称。

公式要有编号，公式编号按章编，如第2章的公式为(2.1)、(2.2)、…，第3章的公式为(3.1)、(3.2)、…等。公式编号写在公式右侧行末顶边线，并加圆括号。

公式一般应在正文中予以引用，引用时以公式编号指示公式。正文中常有公式中表示量的符号说明，采用“式中”二字作为标志。一般可写成接排形式，如“式中，A指……；B指……”。

**关于实践报告的目录格式要求：**

目录是课程实验报告的纲要。正文的各级标题（一般最多取三级）、附录应编入目录，但目录本身不出现在其中。

目录要求层次清晰，含标题及对应的起始页号。目录的最后 “附录”无章节号。

课程实验报告正文、参考文献和附录页面，使用“1，2，3，… ”编连续页码。页码应标在页面的右下角。

目录中正文的各级标题名称、参考文献和附录及其对应的起始页号，务必与报告中正文的各级标题名称、附录及其对应的起始页号保持一致。

# 记录管理模块

## 模块概述

数据表中的数据由若干条记录构成，最终存储在磁盘文件中。本模块在分页文件的基础上，进一步将页面文件改造为记录文件，每个页面的数据区被划分为若干个记录插槽，每个记录插槽用于存放一条记录，并对外提供以记录为单位的记录文件读写接口。记录管理模块依赖于下层的页面管理模块，页面管理模块以页面为单位进行文件读写，而记录管理模块则以记录为单位进行文件读写，每条记录以页面号+插槽号为唯一标识。为简化实现，规定如下：1. 记录文件中存储的记录为定长记录；2. 同一个文件中的记录格式相同，即一个数据表对应一个文件；3. 记录不能跨页，且单条记录的长度不超过页面数据区大小

## 模块设计



图2.1 记录文件结构示意图

记录文件是在页面文件的基础上改造的，所有每个记录文件同样由若干个页面构成，其中第0页为页面信息控制页，存放页面使用情况信息。

1）记录信息控制页

根据图2.1，记录文件的第1页为记录信息控制页。该页的主要功能是控制管理整个页面的数据页分配，包括当前记录总数、每个记录的长度以及每页可以装载的记录数，从pData[0]开始，存放一个数据结构：

typedef struct {

int nRecords; //当前文件中包含的记录数

int recordSize; //每个记录的大小

int recordsPerPage; //每个页面可以装载的记录数量

} RM\_FileSubHeader;

接着，从pData[sizeof(RM\_FileSubHeader)]开始，存放一个位图，用来标记记录文件中哪些页面为满页，哪些页面为非满页。非满页是指页面中还存在空闲的记录插槽的那些页面。位图中相应位为0表示非满页，为1表示满页。该位图与第0页中的位图相似，但是所表示的含义不同。位图的初始值为3，以表示页面信息控制页以及记录信息控制页是满页。

2）数据页

从分页文件的第2页开始为数据页面，每个数据页面被划分为若干个记录插槽，记录在数据页面中采用顺序存储方式。此外，为了区分当前页面中哪些插槽已经存放了记录数据，哪些插槽为空，还需要记录相关的控制信息。

在每个数据页面中，从pData[0]开始存放一个int 大小的值，用来表示当前数据页中包含的记录数，所有的数据页的记录数量之和即是记录信息控制页中的总记录数值。

然后存放一个位图，用于管理该页面中记录插槽的使用情况。如果一个数据页面能够存放n个记录，则需要一个大小为 Sup(n/8) 字节的位图。若某个插槽含有有效记录，位图中相应的位为1；若某个插槽为空插槽，相应的位为0。位图后的空间按记录长度划分为若干个插槽。注意，位图的大小取决于一个页面能够存放的记录数，而一个页面能够存放的记录数又是由记录的大小决定。因此，每个记录文件在创建时都需要根据其存储的记录长度来计算位图所需的字节数。

## 模块实现

### 2.3.1文件操作

使用一个记录文件前，需要先打开文件（RM\_OpenFile），获取一个记录文件句柄，后续的文件操作均以该句柄为标识，不再使用该文件时，要关闭文件（RM\_CloseFile）释放资源。

### 2.3.2记录操作

在以上文件结构设计方案的基础上，记录管理模块需要实现对记录文件的各种读写操作：

1）插入记录

插入记录时，首先在文件中找到一个非满页（若没有则需要申请一个新的页面），在该页中找到一个空插槽并插入记录。记录插入后，要更新控制页，即将文件包含的记录数加1，同时将插槽位图中的相应位置1。如果记录插入后，该页面已经没有空插槽，则将该页面标记为满页，即在记录文件的控制页的位图上将该页面所对应的位置1。

2）删除记录

删除记录时，对于一个给定的记录ID，先检查该记录ID的有效性，然后找到相应的记录。记录删除并不需要将页面中记录所占用的区域清零，只需要将插槽位图中将相应的位置0，并把该页面标记为非满页。同时将控制页中的记录个数减1。

3）修改记录

修改记录时，对给定的记录ID，先检查该记录ID的有效性，然后找到相应的记录，用新的记录值替换原记录值。

4）记录扫描

根据用户给定的条件扫描记录文件的所有数据页，并返回其中符合条件的记录。此操作分两步完成，用户先初始化一个指定条件的文件扫描（打开扫描），然后逐个获取符合条件的记录。

## 实验过程

（设计、编码和调试中遇到的主要问题及所采取的解决方案）

# 系统管理模块

## 模块概述

## 模块设计

## 模块实现

## 实验过程

# 查询处理模块

## 模块概述

## 模块设计

## 模块实现

## 实验过程

# 索引管理模块

## 模块概述

## 模块设计

## 模块实现

## 实验过程

# 实验总结

## 工作总结

（逐条概括、总结此次实验的主要工作，阐述此次实验的心得体会。）

## 改进方案

（在目前HustBase总体设计方案的基础上，指出现有方案的不足，给出自己的改进思路和大致方案）

# 附录