

数据库系统概论中期报告

2017011475 潘俊臣

2017011474 郑林楷

小组人员分工

已经完成部分的分工：

共同完成的部分：

索引模块和记录管理模块的设计，索引模块部分的实现。

分别完成的部分：

郑林楷：记录管理模块部分的实现，部分系统管理模块的功能。

潘俊臣：中期汇报提交。

计划未来的分工：

共同完成的部分：

系统管理模块的进一步设计和实现

查询解析系统的设计和实现

分别完成的部分：

郑林楷：模块测试

潘俊臣：文档工作

项目进度汇报

完成了数据库索引模块和记录管理模块的设计和实现。

完成了小部分系统管理模块。

基础部分设计文档：

整体格式

Folder

每个 Folder 对应一个数据库。

`database.udb`: JSON 格式，储存数据库相关信息。

Sheet

每个前缀对应一个数据表。

默认有一个以 RID 为偏移量的文件 `[Sheet].usid` 用来储存数据。

`[Sheet].usim`: JSON 格式，储存数据表相关信息：

- Name
- Comment (描述, 默认为无)
- RID 计数器 (包括被删除了的记录, 递增)
- record_num (未被删除的记录数量)
- 列信息
 - 每一列的信息
 - Name
 - Key (Primary, Foreign, [Common])
 - Unique (Yes, [No])
 - Default (默认值, 默认为无)
 - Null ([Yes], No)
 - Comment (描述, 默认为无)
- 外键约束
 - 每个外界约束的信息
 - Name
 - 本表所对应的列名
 - 其他表以及所对应的的列名
- 索引信息
 - 每个索引的信息
 - 索引键 (多个)
 - 索引值计算表达式
 - 附加信息 (多个)
 - B+ 树相关信息
 - Page 数量
 - 根 Page 编号
 - 下一个 DELETED Page 编号 (-1 表示 Null)
 - 叶子结点 Record 大小
 - 非叶子结点 Record 大小

Data

命名格式为 `[Sheet].usid`

Page size: 65536 bits, 8192 Bytes, 8KB

由 Record 按 RID 排序而成, Sheet 可在已知 RID 的情况下 $O(1)$ 找到数据。

Index

命名格式为 `[Sheet]_{num}.usid`

Page size: 65536 bits, 8192 Bytes, 8KB

Page:

Macro	bits	Desc
NEXT_DEL_PAGE	32	下一个 DELETED Page 编号 (非全 0 表示 DELETED, 全 1 表示 Null)
LEFT_PAGE	32	左边 Page 编号 (全 1 表示 Null)
RIGHT_PAGE	32	右边 Page 编号 (全 1 表示 Null)
RECORD_COUNT	16	Record 数量
RECORD_SIZE	32	RECORD长度
MIN_CHILD	32	最左边的儿子的 Page 编号
RECORDS	?(?)	各个记录

CHILD_PAGE 在非叶子节点才有。

Each record in index page:

Macro	bits	Desc
RECORD_ID	32	Record ID
KEY	?	索引值
CHILD	32	当前 Record 右边儿子的 Page 编号

PS: SORTED_QUEUE 前后放置两个极小极大值, 极小值为 0000, 极大值为 1FFF

类型

- INT
- CHAR()
- VARCHAR()
 - 最多 65536 个字节

命令

DESC table_name;

- 查看表 table_name 的定义

CREATE DATABASE db_name;

- 创建名为 db_name 的数据库

CREATE TABLE table_name (key_name TYPE ATTRI, ~);

- 创建名为 table_name 的表

INSERT INTO table_name VALUES (~);

- 向表内插数据

操作思想

对列的操作直接重构主索引 (RID 为索引值) 和有关的次索引即可。

嵌套查询：在索引搜索后进行全局搜索。

多表连接：有序的话就一一配对？复杂度 $O(n)$ ？

聚集查询：新建索引即可。

模糊查询：有些能建立索引搜索，有些就还是全局吧。

存在的问题和困难

无

对课程的建议

建议大作业的要求再细化一些。大作业文档很多部分的描述都很模糊，尤其是数据范围这一块，初期设计的时候很迷茫。