

# 数据库概论期末总结

201220014 崔家才

## 第一章 数据库系统概述

### 1.1 基本概念 (概念)

- ◆ 数据库 (数据的集合)、数据库管理系统 (管理数据库的系统软件)、数据库系统 (DB+DBMS+DBA+SP+HP)、数据库管理员

### ~~1.2 数据库系统的发展及趋势~~

### 1.3 数据库系统的特点 (概念)

- ◆ 数据集成化、数据独立性、数据共享、数据冗余、数据的安全性、完整性和一致性、并发控制和故障恢复

### 1.4 数据库内部结构体系 (概念)

- ◆ 数据库的三级结构 (外模式-[逻辑独立性]-概念模式-[物理独立性]-内模式): 三级模式、二级映射

## 第二章 数据模型

### 2.1 数据模型的基本概念 (概念)

- ◆ 概念数据模型 (模型描述)、逻辑数据模型 (具体的 DDL)、物理数据模型 (计算机内部)

### 2.2 数据模型的四个世界 (现实世界、概念世界、信息世界、计算机世界) (概念)

### 2.3 概念世界与概念模型

- ◆ E-R 模型与 E-R 图 (应用★): 实体 (方框)、属性 (圆圈)、联系 (菱形)
- ◆ 扩充 E-R 模型与扩充 E-R 图 (应用★): IS-A 联系 (指向子集的圆圈箭头)、弱实体
- ◆ ~~面向对象模型: 对象、对象标识符、类、方法、超类和子类、聚合和分解、继承和合成、方法、消息、封装~~

### 2.4 信息世界和逻辑模型 (概念)

- ◆ 关系模型: 关系 (表)、属性 (列头)、值域、元组 (行)、关系数据库、关键字 (唯一标识)

### ~~2.5 计算机世界与物理模型~~

- ◆ ~~磁盘组织与文件系统~~
- ◆ ~~逻辑模型的物理存储: 项、记录、文件、索引、集簇~~

## 第三章 关系数据库系统

### ~~3.1 关系数据库系统概述~~

### ~~3.2 关系数据库系统的衡量准则~~

#### ~~◆ 完全关系型的十二条衡量准则~~

#### ~~◆ 空值 (NULL)~~

### 3.3 关系模型数学理论—关系代数

#### 3.3.0 关系模型 (概念)

##### ◆ 关系数据结构

- 表结构：表框架、表的元数（列数）与基数（行数）
- 关系：二维表的性质（有限、无序、唯一、原子）
- 关键字：候选关键字（唯一标识）、主关键字（被选中的候选关键字）、外关键字（其他表的关键字）
- 关系数据库：关系子模式—视图（view）

##### ◆ 关系操纵

- 数据查询：两个关系的合并、单个关系内的元组选择、单个关系内的属性指定
- 数据删除、插入、修改
- 空值的处理（计算结果为空，计数结果为 0）

##### ◆ 关系中的数据约束

- 实体完整性约束（主键不含空）、参照完整性约束（若有外键，必存在于外表）、用户定义的完整性

#### 3.3.1 关系的表示 (概念)

##### ◆ 关系的表示（元组的集合）、迪卡尔乘积

#### 3.3.2 关系操纵的表示 (应用★)

##### ◆ 关系代数中的五种基本运算：选择（ $\sigma_F(R)$ ）、投影（ $\pi_{b_1, b_2, \dots, b_n}(R)$ ）、笛卡儿积（ $R_1 \times R_2$ ）、并（ $R_1 \cup R_2$ ）、差（ $R_1 - R_2$ ）

##### ◆ 基本运算的应用实例

#### 3.3.3 关系模型与关系代数 (概念)

#### 3.3.4 关系代数中的扩充运算 (应用★)

##### ◆ 交（ $\cap$ ）、除法（ $\div$ ）、联接（ $\bowtie_F$ ）与自然联接（ $\bowtie$ ）、外联接

##### ◆ 扩充运算的应用实例

#### 3.3.5 关系代数实例 (应用★)

##### ◆ 综合的关系代数应用实例

### 3.4 关系数据库语言 SQL'92

#### ~~3.4.1 SQL 概貌~~

##### ~~◆ SQL 标准的发展历史、SQL 的基本概念与使用方式、SQL 功能简介~~

### 3.4.2 SQL 数据定义功能 (应用★)

- ◆ SQL 中的数据类型 (各种基本数据类型)、~~Oracle~~数据类型
- ◆ 基本的表结构定义 (CREATE / DROP TABLE) 和修改 (ALTER TABLE ... ADD / DROP) 命令

### 3.4.3 SQL 数据操纵功能 (应用★)

- ◆ SQL 语言与关系代数的关系
- ◆ 映像语句结构 (SELECT ... FROM ... WHERE ... GROUP BY ... HAVING ... ORDER BY ... ASC / DESC)
- ◆
- ◆ 基本查询功能: LIKE、IS NULL 谓词、表的联接查询与自联接查询、ORDER BY 子句的功能
- ◆ 嵌套查询: IN、SOME/ANY/ALL, EXISTS 等谓词、相关子查询与独立子查询
- ◆ 子查询的合并: UNION / INTERSECT / EXCEPT [ALL] 运算
- ◆ 统计查询: 统计与分组统计查询、空值与空集在统计函数中的处理方法
- ◆ 复杂数据查询: 两层的 NOT EXISTS 嵌套结构以实现关系代数中的除法运算的查询功能

### 3.4.4 SQL 的更新功能 (应用★)

- ◆ 删除功能 (DELETE FROM ... WHERE ...)
- ◆ 插入功能: 常量元组的插入、带子查询的元组插入 (INSERT INTO ... VALUES ... / INSERT INTO ... SELECT ...)
- ◆ 修改功能 (UPDATE ... SET ... WHERE ...)

### 3.4.5 视图 (概念)

- ◆ 视图概念 (映像语句的导出表, 虚表)、视图与基表的区别 (是否真实存在于磁盘中)
- ◆ 视图的定义命令: 嵌套定义功能 (CREATE VIEW ... AS SELECT ...)
- ◆ 视图的删除命令: 视图删除中的连锁反应 (DROP VIEW ... 依赖于此的视图也会被删除)
- ◆ ~~视图上的数据访问: 可更新视图~~
- ◆ 视图的优点

## 第四章 数据库的安全性与完整性保护

### ~~4.1 数据库的安全性保护~~

- ◆ ~~数据库的安全与安全数据库~~
- ◆ 数据库安全的基本概念与内容: 主体 (访问者)、客体 (数据) (概念)、身份标识与鉴别、自主访问控制、强制访问控制、隐蔽通道、~~审计~~

◆~~数据库的安全标准：国标 GB17859~~

◆ SQL 对数据库安全的支持：SQL 语言所提供的与数据库安全保护有关的命令（`GRANT SELECT/UPDATE ON ... TO ... / REVOKE SELECT/UPDATE ON FROM ... CASCADE`）（应用★）

#### 4.2 数据库的完整性保护

◆~~数据库完整性保护的功能：目的与常用实现措施~~

◆ 完整性规则的三个内容：实体完整性（主键不空）、参照完整性（外键有源）（概念）、用户定义完整性

◆ 完整性约束的设置、检查与处理：在 SQL 语言的 CREATE TABLE 命令中提供的完整性约束定义子句（`CONSTRAINT ... CHECK / NOT NULL / UNIQUE / PRIMARY KEY ...`）（应用★）

◆~~触发器~~

### 第五章 事务处理、并发控制与故障恢复技术

#### 5.1 事务处理 （概念）

◆ 事务的定义与 A（atomicity 原子性）C（consistency 一致性）I（isolation 隔离性）D（durability 持久性）性质

◆ 事务活动及其状态转换图（活动、预提交、失败、异常中止、提交、回滚.....）

◆ 事务控制及相关的参数设置语句：事务的提交与回滚、事务的读/写类型与隔离级别（`SET ISOLATION LEVEL READUNCOMMITTED / READCOMMITTED - 解决脏读 / READREPEATABLE - 解决不可重复读 / SERIALIZABLE - 解决幻像`）

◆~~事务的语句组成成分~~

#### 5.2 并发控制技术 （概念）

◆ 事务

- 事务的并发性、并发控制
- 调度、串行调度、可串行化调度、冲突（不可交换顺序 - 同一事务或者不同事务同一数据至少一写）与冲突可串行化、冲突可串行化的判定方法（优先图无环）
- 三种数据不一致现象：丢失修改（破坏另一个事务的修改结果）、脏读（读取另一个事务未提交的结果）、不可重复读（同一事务同一数据两次读不一致）

◆ 封锁

- 共享锁、排它锁、所相容矩阵、合适事务
- 基于封锁技术的并发控制实现方法

- 封锁协议：三级封锁协议（一级：写前 X 锁，事务结束后释放，防止丢失修改；二级：一级+读前 S 锁，读后即可释放，防止丢失+脏读；三级：一级+读前 S 锁，事务结束后释放，防止丢失+脏读+不可重复读）、两阶段封锁协议（扩展阶段：申请并获得锁；收缩阶段：释放所有申请获得的锁）
- 合法调度：两阶段封锁协议与冲突可串行化的关系
- 多粒度封锁：封锁粒度与多粒度封锁（封锁粒度小，系统并发度高，调度开销大，反之亦然）、意向锁及其锁相容矩阵、~~多粒度封锁协议~~
- ~~死锁及其解决方法、活锁及其解决方法~~

### 5.3 数据库恢复技术 （概念）

- ◆ 数据库恢复的含义、方法和常用措施
- ◆ 数据库故障的分类
- ◆ 数据库故障恢复三大技术
  - 数据转储：静态转储/动态转储、海量转储/增量转储
  - 日志：
    - 日志的内容、组成、作用与记载原则
    - UNDO 日志的记录格式（记更新前的值）、记载规则（先写日志，后操作；先写磁盘后写日志）以及使用 UNDO 日志的恢复过程（倒着看日志，撤销所有没做完的，并加 ABORT）、~~UNDO 日志中的静止/非静止检查点的设置方法以及使用带检查点的 UNDO 日志的恢复过程~~
    - REDO 日志的记录格式（记更新后的值）、记载规则（先写日志，后改磁盘）以及使用 REDO 日志的恢复过程（顺着看日志，重做已提交的，没做完的加 ABORT）、~~REDO 日志中的（非静止）检查点的设置方法以及使用带检查点的 REDO 日志的恢复过程~~
    - UNDO 日志与 REDO 日志的区别与不足
    - UNDO/REDO 日志的记录格式（更新前后的值都记录）、记载规则（先写日志，后改磁盘）以及使用 UNDO/REDO 日志的恢复过程（先 UNDO, 后 REDO）、~~UNDO/REDO 日志中的检查点的设置方法以及使用带检查点的 UNDO/REDO 日志的恢复过程~~
  - 事务的撤销（UNDO）与重做（REDO）
- ◆ 恢复策略：小型（UNDO 即可）/中型（UNDO+REDO）/大型故障（先利用备份，再 UNDO+REDO）的恢复策略
- ◆ ~~数据库镜像~~

## 第六章 数据库中的数据交换

## 6.1—概述

- ◆—数据交换的五种方式

## 6.2 数据交换的管理 (概念)

- ◆—会话管理
- ◆ 连接管理 (CONNECT TO / DISCONNECT)
- ◆ 游标管理：游标的定义、打开、使用与关闭命令、可滚动游标的定义及其在数据更新命令中的使用
- ◆—诊断管理
- ◆ 动态 SQL (没有硬编码的 SQL 语句)

## 6.3—数据交换的流程

## 6.4—数据交换的四种方式

- ◆—嵌入式 SQL：嵌入式 SQL 的编程方式及其程序结构
- ◆—自含式 SQL：函数与过程
- ◆—调用层接口：ODBC/JDBC 函数的基本使用流程
- ◆—Web 方式

# 第七章 数据库的物理组织

## 7.1—概论

## 7.2—数据库的物理存储介质

- ◆—计算机物理存储介质的层次划分

## 7.3—磁盘存储器及其结构

- ◆—计算机磁盘存储器的物理组织方式

## 7.4—文件组织

- ◆—文件的组织结构
- ◆—文件中的记录在磁盘上的分配方式
- ◆—定长记录与变长记录的组织方式

## 7.5—文件记录组织

- ◆—堆文件、顺序文件、散列文件、聚集文件

## 7.6 索引技术与散列技术 (概念)

- ◆ 顺序文件的组织方式 (按主键排序)
- ◆ 索引文件的组织方式
  - 在顺序文件上的索引技术：稠密索引(每个记录索引)、稀疏索引(每个磁盘块索引，用磁盘块第一条记录的主键)、多级索引(稀疏索引+稠密索引；如果是顺序文件，还可以稀疏索引+稀疏索引) (应用★)
  - 非顺序文件中的索引技术(桶索引)
  - ◆—具有重复键值的索引、多维索引



◆ B/B+树文件

- 树的结构与各个节点（能容纳的 **key** 的个数叫做秩）的组成内容（索引只放在叶子结点，内结点的键只用来定位）
- B+树上的搜索（内结点二叉搜索，叶结点顺序搜索）、~~插入、删除算法~~
- B+的特点及其与 B 树的区别（B+树数据只在叶子结点，B 树数据可以在任意结点）

◆ HASH 文件：散列索引文件的组织方式（哈希函数与桶）

7.7 数据库与文件

## 第八章 关系数据库规范化理论

### 8.1 概述 （概念）

- ◆ 模式设计质量的评价指标：数据冗余度、插入/删除等更新异常

### 8.2 规范化理论

#### 8.2.1 函数依赖

- ◆ 各种函数依赖的定义：完全/部分 FD、平凡/非平凡 FD、直接/传递 FD （概念）
- ◆ Armstrong 公理系统 （概念）
- ◆ 使用函数依赖定义的关键字（完全函数依赖）（概念）
- ◆ 属性集闭包（传递闭包）与关键字（闭包为全部属性的最小属性集）的计算算法 （应用★）

#### 8.2.2 与函数依赖有关的范式 （概念）

- ◆ 范式：1NF（属性不可分割）、2NF（1NF+非主属性完全依赖于关键字）、3NF（2NF+非主属性直接依赖于关键字）、BCNF（若  $X \rightarrow Y$ ，则 X 包含关键字）
- ◆ 各级范式的分解方法（将不满足要求的函数依赖拆出去形成两个关系）

#### 8.2.3 多值依赖与第四范式

- ◆ ~~多值依赖、与多值依赖有关的推理规则、4NF~~

### 8.3 规范化所引起的一些问题

- ◆ 函数依赖的逻辑蕴涵、函数依赖集的等价 （概念）、最小函数依赖集及其判定条件 （概念）、最小函数依赖集的计算算法（去除部分依赖、去除多余依赖）（应用★）
- ◆ 模式分解的无损联结性（分解后还可以原样连接）、依赖保持性（闭包相同）及其判定方法 （概念）
- ◆ 直接到 3NF 且满足无损联结性和依赖保持性的模式分解算法（用最小依赖，对每一个函数依赖，看是否需要分出一个新的关系）（应用★）

#### ~~8.4 关系数据库规范化的非形式化判别法~~

### 第九章 数据库设计

#### 9.1 数据库设计概述 (概念)

- ◆ ~~数据库设计的基本任务、数据库的生命周期~~

#### ~~9.2 数据库设计的需求分析~~

- ◆ ~~需求说明书~~

#### 9.3 数据库的概念设计

- ◆ 数据库概念设计的过程 (概念)
- ◆ 视图集成的原理、策略与步骤、冲突的解决办法 (概念)

#### 9.4 数据库的逻辑设计

- ◆ E-R 模型和扩充 E-R 模型向关系模型的转换 (1:1 全参与的 1 合并; 1:n 全参与的 n 合并; m:n 不合并) (应用★)

#### ~~9.5 数据库的物理设计~~

- ◆ ~~存储结构、存取路径、集簇、索引~~

### 第十章 数据库管理

#### ~~9.1 数据库管理概述~~

#### ~~9.2 数据库管理的内容~~

#### ~~9.3 数据库管理员 DBA~~

#### ~~9.4 数据库性能配置和优化 (缓冲池、数据重组、查询优化)~~

#### ~~9.5 数据库管理系统~~