数据库应用-概念设计+逻辑设计

• p28 引言

- p32 数据库技术的发展如何支付方式的发展?
- p33 数据库中存放的是"谁"的数据?"谁"又是如何描述和表示的?
 - p33 直观描述
 - p34 抽象描述: 用概念来描述事或物及相互联系
 - 内涵: "类"
 - 外延: "对象"
 - p36 二维表描述
 - 内涵: 列
 - 外延: 行
- p37 数据管理发展过程
- p42 数据模型
 - 概念及其满足的要求
 - 两个层次
 - 概念模型(信息模型):面向客观事件建模→ER模型
 - 数据模型(逻辑模型):面向计算机建模
 - p44 两大阶段
 - 概念模型的作用
- p514个基本概念
 - 数据 Data
 - 数据库DB
 - p57 与文件的区别
 - 数据库管理系统DBMS
 - DML
 - DDL
 - 数据库系统DBS
- p63 数据库的开发运行环节
- p67 数据抽象(数据视图技术: 物理层、逻辑层、视图层)
- p68 数据库存储技术
- p69 数据库访问技术
- p70 数据库控制技术
- p71 数据库维护技术

- p72 应用技术开发
- p73 开发环节总结
- p77 数据库需求分析(书上好像没有)
 - p81 需求分析
 - 基本任务
 - 基本方法
 - 结果
 - 其他任务
 - p85 数据流程图
 - 基本要素
 - 描述的内容: 表达了数据、处理过程、及其相互作用关系抽象的拓扑结构
 - 分层数据流程图
 - p93 数据字典
 - 6个组成部分
 - p101 UML类图的作用
 - UML: 统一建模语言
 - 类图
 - 用例图
 - 活动图(时序图、顺序图)
 - p105 数据库需求分析
 - 数据库需求分析 vs 需求分析
 - 核心任务
 - 其他任务
- p109 ER模型 < 概念模型
 - p113 数据库设计过程
 - 基础条件
 - 核心阶段
 - 后期工作
 - p114 ER模型三要素: 符号(书172)
 - 实体集
 - 矩形
 - 联系集
 - 菱形
 - 虚线连出矩形,作为其属性

- 四种对应类型
 - 画法: 有一的要添加箭头(1...n, 1方画箭头), 没有一的直接画直线
- 属性
- p117 联系集类型和弱实体集
- p119 属性类型和多元联系集
- p121 练习
- p124 ER模型设计注意问题(书163)
 - p128 用实体还是实体集
 - p131 确定属性的位置
 - p135 用属性还是联系集 & 复杂属性处理方法
 - p139 用实体集还是联系集
 - p141 用二元联系还是多元联系
- p143 扩展ER模型
 - 特化: 自顶向下
 - 概化: 自底向上
 - 聚集(用的不多)
- 例题:作业1▶(根据一段话绘制ER图)
- p147 关系模型
 - p150 数据模型
 - 三大基本要素: 数据结构、数据操作、约束条件
 - 三大基本模型
 - p152 层次模型:基本概念+数据操作+优缺点
 - p161 网状模型:基本概念+数据操作+优缺点
 - p168 关系模型
 - 关系,元组,属性→表,行,列
 - 实体集→数据对象
 - 联系集→联系表
 - 1NF: 关系的属性具有原子性
 - p171 码与模式图
 - 书23 超码 vs 候选码 vs 主码
 - 书25 外码,参照关系
 - p174 关系模型的完整性约束
 - 实体完整性规则: 主键约束
 - 参照完整性规则:外键约束
 - 域完整性规则: 创建时添加check语句

- p179|p184+书28 关系代数▶
 - 基本运算(各有各的查询优势):选择,投影,更名,笛卡尔积,并,差
 - 附加运算(可用基本运算表示,查询能力不会增强):交,自然连接,赋值运算,外连接
 - 扩展运算(无法用基本运算表示):广义投影,聚集函数
- p201 ER模型如何转为关系模型
 - 实体集转换
 - 简单实体集
 - 弱实体集: 主码→弱联系对应主实体的主码+弱实体集的分辨符
 - 复杂实体集: 横向(复合属性)和纵向(多值属性)扩展
 - 联系集转换
 - 一对一、多对多、一对多
 - 弱实体联系(冗余模式,可忽略)
 - p215 一元、二元、三元联系转换
 - 模式合并
 - 扩展ER图转换
 - 特化和概化
 - 聚集
 - p209 ER图转换为关系模式的规则总结
 - 格式: 实体名字(属性1, 属性2)
 - 联系集转换
 - 有1的: 多方, 需要放置一方的主键和所有联系属性
 - 多对多: 生成新的关系(主键1,主键2,联系属性),其主键为(主键1,主键 2)
 - p220 例题▶

• p222 数据库操作SQL

- p226 数据定义语言DDL
 - create table
 - create database
- p229 数据操作语言DML
 - p229 查询
 - 投影: SELECT
 - 选择: where
 - 聚集函数: (where ...) group by ... (having ...)
 - 嵌套子查询: in, some, all, exist

- 集合运算: Union, Except, Intersact
- 更名 as
- p234 insert:3种
- delete
- update
- p236 表间连接
 - 自然连接 vs 等值连接
 - 连接: (inner join | left outer join | right outer join | full outer join)
 - 方式: (natural | on | using)
- p233 示例:用SQL语句完成需求▶
- p238 高级SQL
 - p242 视图的定义及其创建
 - p243 完整性约束
 - 实体完整性: primary key (A1,A2,...,Ak)
 - 参照完整性: foreign key (B1,B2,...,Bk) references B
 - 域完整性: check
 - p244 创建断言(数据更新时,保持谓词为真)
 - p245 授权
 - 表(关系)上、视图上、授权图及其动态管理
 - 索引(ppt上没讲,自己查)
- p254 逻辑模型设计(优化-范式与函数依赖)
 - p259 关系模型设计中潜在的问题: 修改异常,添加异常,删除异常
 - 函数依赖基础知识
 - 【问题】冗余→【解决】模式分解(带来的新问题: p296)
 - p265 定义
 - p273 函数依赖理论
 - Armstrong公理:
 - 自反律: $X \supset Y 则 X \rightarrow Y$
 - 增广律: $X \to Y$ 则 $XZ \to YZ$
 - 传递律: X o Y, Y o Z则X o Z
 - Armstrong推理:
 - 合并律: $X \to Y, X \to Z$ 则 $X \to YZ$
 - 分解律: $X \to YZ$ 则 $X \to Y, X \to Z$
 - 伪传递律: $: X \to Y, WY \to Z$ 则 $WX \to Z$
 - p281 函数依赖与键

- 闭包计算
 - p282 函数依赖集闭包

• 算法1: p282, 指数级

● 算法2: p290, 先计算每个属性组合的属性集闭包 ▶

• p284 属性依赖集闭包

• 算法: p285

• 3个用途: p289

- 正则覆盖
- p295 模式分解案例与存在的问题
- 模式分解的两个标准

基本要求: p297

- 无损连接分解
 - 定义: p299

• 其他判断方式: p302

- 保持函数依赖
 - 定义: p304

• 其他判断方式: p306

- 关系模型好坏的评价标准/关系模式优化依据: 范式
- 关系模式的常用范式
 - p272 1NF:每个属性都是原子属性
 - p272 2NF
 - 任何满足第二范式的关系满足第一范式
 - 所有非主属性必须依赖于整个候选码而不能依赖于候选码的部分属性。
 - BCNF
 - 条件 p315
 - 函数依赖平凡,即左边包括右边
 - 函数依赖不平凡,则左边属性集是超键
 - 其他判断方式 p316+p317
 - 不满足BCNF后的分解 p319
 - 分解: 若u→w不满足BCNF,则分解R为uw和R-w
 - p324 无损分解,但不一定保持函数依赖,典例p322
 - p325 总结
 - p344 BCNF分解树 ►
 - 3NF
 - 解决BCNF的问题: p326

- 比BCNF弱的一个关系,是分解和性能的折中
- 条件
 - 左边包括右边,即函数依赖平凡
 - 左边是超键
 - 如果左边不是超键,则需要右边是超键的一部分
- 表现:不存在传递依赖
 - 理解2:在满足2NF的条件下,不包含非主属性对码的传递函数依赖(即码应该直接决定非主属性,不能间接决定)
- 不满足3NF后的分解 p332
 - 求正则覆盖
 - 合并律
 - 删除无关属性
 - 每个函数依赖做成子模式
 - 判断任意候选键是否需要做成子模式
 - 删除是其他子模式子集的子模式
- 双保持:保持无损连接,保持函数依赖
- p335 BCNF VS 3NF
- 典例: p350 ►
- 36分开始, 没听到...

数据库原理-物理设计

- p353 物理设计
 - p353 数据库存储技术
 - 硬盘组织方式 p356
 - 定长、变长
 - 文件记录组织方式 p357
 - 堆、顺序、散列、多表簇集
 - 物理设计的任务 p362
 - 数据字典的存放 p364
 - 数据库缓冲区 p368
 - 作用
 - 管理过程
 - 替换策略
 - p374 索引与散列
 - 索引的基本概念 p378
 - 顺序索引
 - 按照顺序分 p379
 - 主索引(顺序一致)
 - 辅助索引(顺序不一致)→一定是稠密索引
 - 按照大小分 p380
 - 稠密索引
 - 稀疏索引→一定是主索引
 - 多级索引 p381
 - 辅助索引 p382: 存在间接指针层
 - B+树 p383(了解)
 - 散列索引
 - 散列函数
 - 分类
 - 静态索引 p387
 - 动态索引 p388 → 根据搜索码二进制前i位分裂或合并桶
- p395 查询处理
 - 查询处理过程 p399: 关系代数和SQL语句的相互转换
 - 查询代价的度量 p400

- 选择操作的实现 p401
 - 全表扫描
 - 索引(或散列)扫描
- 连接操作 p405
 - 嵌套循环方法
 - 排序-合并方法(归并方法)
 - 索引连接方法
 - Hash join方法
- 表达式的计算 p415
 - 物化计算
 - 流水线计算
- 数据库SQL引擎 p417: SQL解析(词法分析、语法分析、语义分析)+优化(查询重写)
- 查询树的绘制
- 查询优化概述 p435
- 查询优化总目标 p438
- 查询优化步骤 p439
- 表达式转换的等价规则 p452, 书331
- 查询优化三个准则
 - 选择下移策略: 先选择, 减少列
 - 投影下移策略: 先投影, 减少行
 - 选择连接顺序优化策略: 小关系连接优先策略
- 执行计划 p463: 确切定义每个运算应使用的算法+运算之间的执行应该如何协调
- 物理查询优化:索引的选择(了解即可)
- 连接运算结果大小的估算(了解即可)
- openGauss的查询重写技术 p468

• p478 事务管理

- 事务 p483, 书356
 - 基本概念
 - 事务基本知识: 一系列指令形成的逻辑单元, 并只能原子执行(要么全部执行, 要么不执行)
 - 事务执行出错后:回滚
 - CAID特性
 - 事务状态图
 - 并发调度的好处书360
- p485 调度、串行调度、并行调度、可串行化调度

- p486 冲突指令、冲突可串行化、调度优先图、冲突可串行化判定方法(优先图无环)
- p490 事务的隔离性和原子性
 - 可恢复调度与无级联调度
 - 事务隔离性级别
 - 隔离性依次降低:可串行化、可重复读、已提交读、未提交读
- p495 隔离性级别的实现方式
 - 两阶段封锁协议
 - 时间戳排序协议
 - 快照隔离技术
- 书71+369 SQL中的事务定义
- p496 事务处理
 - 显示事务
 - 自动提交事务
 - 隐性事务
- p510 并发控制
- 锁
 - p516 锁的概念和基本种类
 - 共享锁S, 排他锁X
 - 死锁(双等待-僵局)、活锁(单相思-饿死)
 - p528 锁转换
 - p529 锁管理器
 - p541+书379 死锁
 - p541 死锁检测:等待图有环
 - p542 死锁解决: 找事物回滚
 - p543 死锁预防
- p516 基于锁的协议
 - p517 封锁协议: lock、unlock
 - p521 两阶段封锁协议:保证冲突可串行化,但不保证不会死锁,不保证可恢复性和无级 联回滚
 - 增长阶段、缩减阶段
 - p525 严格两阶段封锁协议(排他锁在事务提交后释放):保证无级联回滚→也保证了 可恢复性
 - p527 强两阶段封锁协议(所有锁在事务提交后释放): 按照锁的提交顺序串行化
 - p534 树形协议:保证冲突可串行化,保证不会死锁,不保证可恢复性和无级联回滚
 - 基本思路

- 协议内容
- p540 优缺点
- p544 多粒度锁与封锁协议: 保证可串行性
 - 基本概念: 书381+p544
 - p545 类型:显示锁+低层隐式锁+高层意向锁
 - p547 加锁示例
 - p548 协议(了解)
- p553 基于时间戳的协议 ►: 保证冲突可串行化,保证不会死锁,不保证可恢复性和无级联回滚
 - 没有锁, 所以保证不会死锁
 - p554 协议内容
- p559 快照隔离
 - p559 概述
 - p560 方法: 先提交者获胜、先更新者获胜
- 基于有效项检查的协议(书386,了解)

p566 备份和恢复

- p570 故障分类:事务故障、系统崩溃、磁盘故障、其他故障
- p571 稳定存储器的实现
 - 易失性存储器、非易失性存储器、稳定存储器
 - RAID → 远程备份系统
 - p572 数据访问
- 数据恢复与原子性
 - p578 检查点
 - 基于日志的恢复机制
 - redo
 - 完成了: commit, abort
 - 不需要重新生成日志
 - undo
 - 未完成: 只有start
 - 需要从下往上回滚并生成日志
 - p581 基于检查点进行恢复
 - checkpoint
 - checkpoint前已完成的事务不管
- p582 远程备份系统
 - 基本工作原理

- 关键技术
- p588 体系结构
 - p593 集中式体系结构
 - 基本组成要素
 - 分类: 单用户系统、多用户系统
 - p595+p606 粗粒度并行VS细粒度并行
 - p596 CS(客户-服务器体系结构)
 - 两层: 前端和后端
 - 三层:表现层、业务逻辑层、数据访问层
 - BS (浏览器-服务器体系结构)
 - 数据模型的三层模式
 - p600 外模式:界面、数据库用户的数据视图
 - p601 模式: 概念模式, 关系、逻辑结构→公共数据视图
 - p602 内模式: 物理结构, 如存储结构等, 数据的内部表示方式
 - 三层模式-两层映射:一方面保证了数据的独立性(逻辑独立性与物理独立性),另一方面,应用程序所需要的数据视图被分隔开,同样保证了数据的安全性
 - p604 服务器系统体系结构
 - 事务服务器和数据服务器
 - 6个进程
 - p606 并行系统
 - 衡量要素
 - DBMS分类
 - p609 分布式系统
 - 结构说明
 - 主要用途
 - 注意: 异构分布式数据库、保持事务原子性的方法(如: 两阶段提交法,2PC)
 - p612-p614 并行系统 VS 分布式系统

• p617 特殊数据库

- 特殊数据库
 - 复杂数据类型
 - SOL
 - 支持的复杂数据类型
 - 复杂关系表的定义
 - 复杂关系表上插入数据
 - 复杂关系表上查询

- 对象标识和引用类型
- XML
 - 定义
 - XML vs HTML
 - 基本要素:元素、属性
 - 名字空间
- XML数据文档模式
 - 文档类型定义DTD
 - schema模式
 - XML文档树模式
- Xpath语言
- XQuery语言
- 云数据库
 - 概念
 - 特点
 - 举例
 - 云数据库vs自建数据库对比
 - 影响
 - 云数据库vs分布式数据库
 - 类型
 - 托管
 - 云原生
- NoSQL
- NewSQL