



海量数据计算研究中心

数据库系统 总结与要点

主讲：程思瑶

海量数据计算研究中心





概览

数据库系统

基础篇

数据库系统概述

关系数据库系统

结构化查询语言SQL

设计篇

概念数据库设计

逻辑数据库设计

物理数据库设计

实现篇

物理存储结构

查询处理

查询优化

并发控制

数据库恢复





数据库系统：基础篇

● 数据库系统概述

- 数据、数据库、数据库管理系统、数据抽象与数据模型，层次模型、网状模型
- 重点：
 - ✓ DBS定义与特点,文件系统与数据库系统的区别
 - ✓ 三级抽象(视图抽象、逻辑抽象、物理抽象)
 - ✓ 二级映射
 - ✓ 数据独立性(物理数据独立性、逻辑数据独立性)





数据库系统：基础篇



- 关系数据库系统

- 关系模型、关系运算

- 重点：

- ✓ 关系数据库的数据结构(元组、属性、域、码(超码、候选码、外码))
 - ✓ 完整性约束规则(实体完整性、参照完整性、用户自定义完整性)、
 - ✓ 关系代数(6种基本操作，关系代数书写查询)





数据库系统：基础篇

- 结构化查询语言SQL

- SQL-数据定义、查询、数据库修改、事物、触发器，嵌入式SQL简介
- 重点：
 - ✓ SQL数据定义(关系定义、视图定义、完整性约束的SQL语言表示),
 - ✓ SQL查询(单表、连接、分组、聚集),
 - ✓ SQL数据修改





数据库系统：设计篇

● 概念数据库设计

- 数据库设计概述、需求分析、概念数据库设计方法
- 重点：
 - ✓ ER图
 - ✓ 实体
 - ✓ 属性
 - ✓ 码
 - ✓ 联系：映射基数、参与约束(全域关联、部分关联)





数据库系统：设计篇

● 逻辑数据库设计

- 初始关系模式形成，关系模式规范化、优化，定义完整性和安全性约束，定义子模式，性能估计
- 重点：
 - ✓ ER图—>关系表，
 - ✓ 关系数据库的规范化：函数依赖、完全函数依赖、部分函数依赖、传递函数依赖、Armstrong公理系统、求属性闭包、求候选码、求极小函数依赖集，
 - ✓ 关系数据规范形式：1NF、2NF、3NF、BCNF、无损连接性、函数依赖保持性、判别方法、关系模式的分解算法





数据库系统：设计篇



- 物理数据库设计

- 影响物理数据库设计的因素，关系模式选择存取方法，设计关系、索引等数据库文件的物理存储结构
- 重点：
 - ✓ 影响物理数据库设计的因素、
 - ✓ 聚集存储方法





数据库系统：实现篇

- 物理存储结构

- 数据库存储设备，磁盘文件，Hash文件，索引文件
B+树文件索引，B树文件索引
- 重点：
 - ✓ 磁盘缓冲区处理技术、
 - ✓ 磁盘容错技术(RAID1,2,4,5)
 - ✓ 磁盘文件(文件，文件记录、文件存储方式(连续、链式、索引))
 - ✓ Hash文件(动态Hash技术)
 - ✓ 索引文件(主索引、聚集索引、辅助索引、多级索引、B+树索引结构)





数据库系统：实现篇



- 查询处理

- 关系代数操作实现算法

- 重点：

- ✓ 选择、

- ✓ 投影、

- ✓ 连接(Nest-Loop-Join(优化), Sort-Merge-Join、Hash Join)、

- ✓ 集合操作





数据库系统：实现篇



● 查询优化

- 关系表达式的等价转换规则，表达式结果大小的估计，启发式关系代数优化算法，复杂性估计方法
- 重点：
 - ✓ 关系代数的等价转换规则
 - ✓ 表达式结果大小的估计(均匀分布)、处理时间的估计
 - ✓ 启发式关系代数优化算法(查询树、优化树)





数据库系统：实现篇

- 并发控制

- 事务概念，事务的并发执行和调度，并发控制协议（基于锁的协议、基于时间戳的协议、多版本机制快照隔离）
- 重点：
 - ✓ 事物的概念(ACID)、数据库的一致性
 - ✓ 串行调度、可串行调度、冲突可串行化的判定方法(定义，冲突图)
 - ✓ 基于锁的并发控制协议(锁的概念与种类，锁的相容关系、两段锁协议、死锁的判定)
 - ✓ 基于时间戳的并发控制协议，两者的比较





数据库系统：实现篇

● 数据库恢复

- 数据库恢复必要性，使用日志的数据库恢复技术使用检查点的数据库恢复技术，恢复算法，缓冲技术
- 重点：
 - ✓ 数据库故障分类
 - ✓ 使用日志的数据库恢复技术(日志的内容，redo、undo、推迟更新技术、即时更新技术)
 - ✓ 使用检查点的数据库恢复技术(日志的内容，哪些事物需要redo、哪些需要undo)、
 - ✓ 恢复算法(*先写日志，后更新数据库)





Thanks

