

课程编号：22010110

课程名称：数据库概论

英文名称：Introduction to Databases

前导课程：程序设计、离散数学、数据结构、操作系统

任课教师：胡伟

周讲课/实习时数：4/0

学分：3

课程简介

“数据库概论”是为南京大学计算机科学与技术系本科生开设的方向核心课程，主要介绍数据库的基本概念与关系数据库系统的基本理论。通过对相关概念和理论的学习，可以让学生掌握如何去设计、建立、操作和管理一个具体的数据库系统，从而为以后的数据库理论研究、课程实践以及信息系统开发工作建立一个良好的基础。

课程目的与要求

通过对该课程的学习使学生能够：

- 1) 熟悉数据库领域内的一些专用名词术语，了解数据库的基本概念与思想
- 2) 掌握数据库技术的基本理论—数据模型，了解三个不同层次上的数据模型概念，掌握三种常用的概念数据模型—实体联系模型、扩充实体联系模型和面向对象数据模型的建模方法
- 3) 掌握关系模型的基本数学理论—关系代数，以及关系数据库的设计理论—关系规范化理论
- 4) 了解数据库的设计方法与设计过程，掌握关系数据库的逻辑数据模型的设计方法
- 5) 学会使用交互式 SQL 语言来建立和访问关系数据库中的数据，并定义系统的数据完整性和访问控制要求
- 6) 建立初步的数据库的数据交换概念，了解如何在应用程序中访问数据库

课程内容与学时分配：

第一章 数据库系统概述（2 学时）

1.1 基本概念

◆ 数据库、数据库管理系统、数据库系统、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

- ◆ 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

- ◆ 数据库的三级结构：三级模式、二级映射

第二章 数据模型（6 学时）

2.1 数据模型的基本概念

- ◆ 概念数据模型、逻辑数据模型、物理数据模型

2.2 数据模型的四个世界

2.3 概念世界与概念模型

- ◆ E-R 模型与 E-R 图：实体、属性、联系
- ◆ 扩充 E-R 模型与扩充 E-R 图：IS-A 联系、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

- ◆ 关系模型：关系、属性、值域、元组、关系数据库、关键字

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（14 学时）

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

- ◆ 关系数据结构
 - 表结构：表框架、表的元数与基数
 - 关系：二维表的性质
 - 关键字：候选关键字、主关键字、外关键字
 - 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

◆ 关系中的数据约束

- 实体完整性约束、参照完整性约束、用户定义的完整性

3.3.1 关系的表示

◆ 关系的表示、迪卡尔乘积

3.3.2 关系操纵的表示

◆ 关系代数中的五种基本运算：选择、投影、笛卡儿积、并、差

◆ 基本运算的应用实例

3.3.3 关系模型与关系代数

3.3.4 关系代数中的扩充运算

◆ 交、除法、联接与自然联接、外联接

◆ 扩充运算的应用实例

3.3.5 关系代数实例

◆ 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

◆ SQL 中的数据类型、Oracle 数据类型

◆ 基本的表结构定义和修改命令

3.4.3 SQL 数据操纵功能

◆ SQL 语言与关系代数的关系

◆ 映像语句结构

◆ 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

◆ 删除功能

◆ 插入功能：常量元组的插入、带子查询的元组插入

◆ 修改功能

3.4.5 视图

- ◆ 视图概念、视图与基表的区别
- ◆ 视图的定义命令：嵌套定义功能
- ◆ 视图的删除命令：视图删除中的连锁反应
- ◆ 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（6 学时）

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、客体、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

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、参照完整性、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
- ◆ 触发器

第五章 事务处理、并发控制与故障恢复技术（10 学时）

5.1 事务处理

- ◆ 事务的定义与 ACID 性质
- ◆ 事务活动及其状态转换图
- ◆ 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
- ◆ 事务的语句组成成分

5.2 并发控制技术

- ◆ 事务
 - 事务的并发性、并发控制
 - 调度、串行调度、可串行化调度、冲突与冲突可串行化、冲突可串行化的判定方法
 - 三种数据不一致现象：丢失修改、脏读、不可重复读
- ◆ 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、两阶段封锁协议
- 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、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

- ◆ 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、记载规则以及使用 UNDO 日志的恢复过程、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、记载规则以及使用 REDO 日志的恢复过程、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、记载规则以及使用 UNDO/REDO 日志的恢复过程、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/中型/大型故障的恢复策略
- ◆ 数据库镜像

第六章 数据库中的数据交换（2 学时）

6.1 概述

- ◆ 数据交换的五种方式

6.2 数据交换的管理

- ◆ 会话管理
- ◆ 连接管理
- ◆ 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

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（4 学时）

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

- ◆ 顺序文件的组织方式
- ◆ 索引文件的组织方式
 - 在顺序文件上的索引技术：稠密索引、稀疏索引、多级索引
 - 非顺序文件中的索引技术
 - 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的组成内容
 - B+树上的搜索、插入、删除算法
 - B+的特点及其与 B 树的区别
- ◆ HASH 文件：散列索引文件的组织方式

7.7 数据库与文件

第八章 关系数据库规范化理论（8 学时）

8.1 概述

- ◆ 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
- ◆ Armstrong 公理系统
- ◆ 使用函数依赖定义的关键字
- ◆ 属性集闭包与关键字的计算算法

8.2.2 与函数依赖有关的范式

- ◆ 范式：1NF、2NF、3NF、BCNF
- ◆ 各级范式的分解方法

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、最小函数依赖集及其判定条件、最小函数依赖集的计算算法
- ◆ 模式分解的无损联结性、依赖保持性及其判定方法
- ◆ 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

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（6 学时）

9.1 数据库设计概述

- ◆ 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
- ◆ 视图集成的原理、策略与步骤、冲突的解决办法

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

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（2 学时）

10.1 数据库管理概述

10.2 数据库管理的内容

10.3 数据库管理员 DBA

10.4 数据库性能配置和优化（缓冲池、数据重组、查询优化）

10.5 数据库管理系统

上机实验（8 学时）

1. 实验目的

- 1) 加深对数据库课程的理解
- 2) 通过实验掌握一种数据库产品的 SQL 语言的使用方法与操作能力
- 3) 通过数据库内核实验, 深入理解数据库分层架构的设计与实现

2. 实验要求

- ◆ 实验一 (2 学时): 熟悉一种数据库产品的操作使用环境, 掌握交互方式下 SQL 语言的数据操纵功能
- ◆ 实验二 (2 学时): 掌握高级程序设计语言与数据库的数据交换, 以及简单的存储过程和存储函数

系统实验每学年选做 2 项:

- ◆ 系统一 (2 学时): 掌握数据库存储引擎与多种页面调度算法
- ◆ 系统二 (2 学时): 掌握火山模型执行引擎, 实现数据库基本算子, 了解外部排序实现方式
- ◆ 系统三 (2 学时): 掌握基本数据分析算子, 理解内连接、外连接以及聚合算子的实现方式, 了解不同连接方式的优劣与适用的查询条件
- ◆ 系统四 (2 学时): 掌握 B+树索引结构及动态更新, 了解索引结构 (哈希索引等) 在不同查询类型上的优势

3. 实验方法

- 1) 所有实验应由学生个人独立完成 (不推荐学生以组为单位完成)
- 2) 所有实验都在学院教学用计算机机房或学生个人电脑上完成
- 3) 教师及助教检查学生的实验报告和代码, 并给出成绩, 计入课程总分中

课程考核方法: 课后作业 (10%) + 上机实验 (20%) + 期末考试 (70%)

教材

[1] 徐洁磐, 柏文阳, 刘奇志. 数据库系统实用教程. 高等教育出版社, 2006

参考书

- [2] 王珊, 萨石煊. 数据库系统概论 (第 5 版). 高等教育出版社, 2014
- [3] A. Silberschatz, H.E. Korth, S. Sudarshan. Database System Concepts (6th Ed.). 杨冬青 等译. 机械工业出版社, 2012
- [4] 施伯乐, 丁宝康, 汪卫. 数据库系统教程 (第 3 版). 高等教育出版社, 2008
- [5] J.D. Ullman, J. Widom. A First Course in Database Systems (3rd Ed.). 岳丽华 等译. 机械工业出版社, 2009
- [6] H. Garcia-Molina, J.D. Ullman, J. Widom. Database System Implementation (2nd Ed.). 杨冬青 等译. 机械工业出版社, 2010