

# 数据库系统原理

陈岭

浙江大学计算机学院

- **教材:**

- Database Systems Concepts 6<sup>th</sup> edition
- By Abraham Silberschatz, Henry F. Korth and S. Sudarshan
- Higher Education Press, McGraw-Hill Companies

- **参考书目:**

- Database Management Systems 3<sup>rd</sup> edition  
By Ramakrishnan and Gehrke
- Database Systems: The Complete Book  
By Garcia-Molina, Ullman and Widom
- 数据库系统概论（第四版），萨师煊 王珊，高等教育出版社，2006
- 数据库系统原理教程，王珊 陈红，清华大学出版社，2003
- 数据库课程设计，陈根才 孙建伶 林怀中 周波，浙江大学出版社，2007  
(实验参考书)

# 1

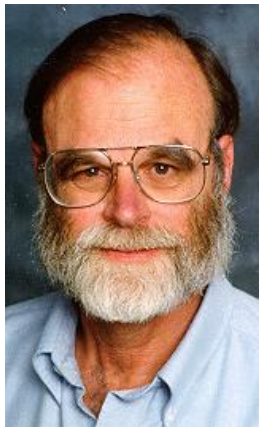
## 引论

- ❑ 数据库系统目的
- ❑ 数据视图
- ❑ 数据模型
- ❑ 数据库语言
- ❑ 数据库管理员
- ❑ 数据库用户
- ❑ 事务管理
- ❑ 存储管理
- ❑ 数据库体系结构

# 数据库系统目的

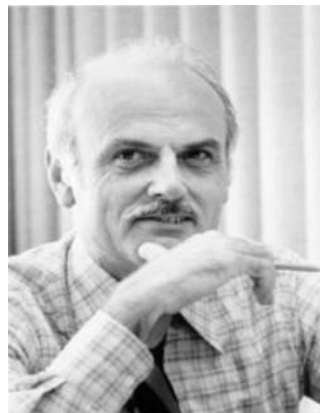
- ❑ 数据处理和管理是计算机应用最重要的领域，数据库系统的知识对于计算机学者至关重要。
- ❑ 数据库涉及社会生活的方方面面：
  - 银行业：所有交易
  - 航空公司：预订，时间表
  - 大学：注册，成绩
  - 销售：客户，产品，购买
  - 制造业：生产，库存，订单，供应链
  - 电子政务，电子商务，... ..
- ❑ 你可能学习了很多课程，但数据库系统能够让你找到一份好工作

# 图灵奖获得者



1998, James Gray——事务、锁、日志和二阶段提交

1981, Edgar F. Codd——关系数据库



1972, Charles W. Bachman——网状数据库

# 学些什么？

我们将从以下三方面学习数据库相关知识：

## □ 数据库模型与设计

- 从现实生活中抽象出数据模型，再将其转换为适合目标DBMS（数据库管理系统）的形式：表、视图。

## □ 编程：使用数据库

- 查询、更新数据（SQL）

## □ 数据库管理系统实现

- 数据库管理系统的工作机制及设计

# 学生分数表的设计

学号	姓名	专业	DB平时	DB期末	DB总评成绩	OS平时	OS期末	OS总评成绩
3023001093	黄毅照	混合班	85	95	90			85
3011112340	周朝威	计算机科学与技术	80	90	85			88
3020621034	徐鑫	计算机科学与技术	90	90	90			85
3020831035	薄延嵩	计算机科学与技术	70	80	75			90
3021131123	胡俊	计算机科学与技术	70	70	70			75
3022112002	蒋永丽	计算机科学与技术	80	90	85			80
3022112003	顾娉婷	计算机科学与技术	90	90	90			85

这张表的设计好吗？为什么？

## 另外一种表的设计

Students

Sid	Sname	Ssex	Sage	Specialty
3023001093	黄毅照	M	21	No
3011112340	周朝威	F	20	Cs
3020621034	徐鑫	M	18	Cs
3020831035	薄延嵩	M	19	Cs
3021131123	胡俊	F	22	Cs

Courses

cid	Cname	credit
1	DB	4
2	OS	5
3	English	4
4	Math	4

Enrolled

sid	cid	grade1	grade2	grade3
3023001093	1	90		
3023001093	2	85		
3020621034	1	90		
3020831035	1	75		
3021131123	2	75		

这种表的设计好吗？

为什么？

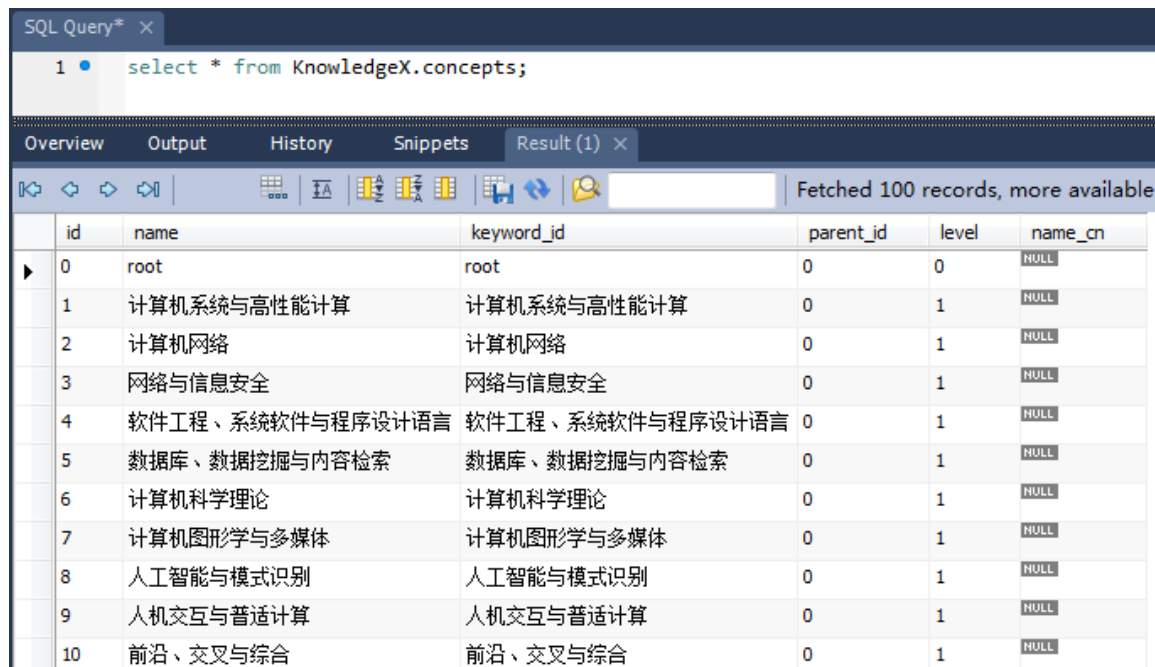




# 数据库访问

## 方法1：利用数据库管理系统提供的交互工具访问数据库

如：MySQL WorkBench, SQL Server 查询分析器, ORACLE Sql\*Plus, Work Sheet



The screenshot shows a SQL query tool interface. At the top, a tab labeled 'SQL Query\*' contains a query: `1 • select * from KnowledgeX.concepts;`. Below the query editor, there are tabs for 'Overview', 'Output', 'History', 'Snippets', and 'Result (1)'. The 'Result (1)' tab is active, displaying a table of results. The table has columns: 'id', 'name', 'keyword\_id', 'parent\_id', 'level', and 'name\_cn'. The results show a hierarchical structure of concepts, starting with 'root' at level 0, and various computer science topics at level 1. A status bar at the top right of the result area indicates 'Fetched 100 records, more available'.

	id	name	keyword_id	parent_id	level	name_cn
▶	0	root	root	0	0	NULL
	1	计算机系统与高性能计算	计算机系统与高性能计算	0	1	NULL
	2	计算机网络	计算机网络	0	1	NULL
	3	网络与信息安全	网络与信息安全	0	1	NULL
	4	软件工程、系统软件与程序设计语言	软件工程、系统软件与程序设计语言	0	1	NULL
	5	数据库、数据挖掘与内容检索	数据库、数据挖掘与内容检索	0	1	NULL
	6	计算机科学理论	计算机科学理论	0	1	NULL
	7	计算机图形学与多媒体	计算机图形学与多媒体	0	1	NULL
	8	人工智能与模式识别	人工智能与模式识别	0	1	NULL
	9	人机交互与普适计算	人机交互与普适计算	0	1	NULL
	10	前沿、交叉与综合	前沿、交叉与综合	0	1	NULL

方法2：利用开发工具设计界面、处理数据，调用ODBC等接口访问数据库，  
如：ASP，JSP，VC++，PHP，PowerBuilder，Delphi



浙江大學

专业培养计划查询

必修课教学计划

院系选修课

限定性选修课

辅修课教学计划

学院

计算机科学与技术学院

专业

计算机科学与技术

年级

2002

学期

全部

课程代码	课程名称	学分	周学时	考核方式	课程性
02110010	思想道德修养	2.0	1.0-2.0	考查	必修课
02110020	法律基础	1.5	1.0-1.0	考查	必修课
02110032	毛泽东思想概论(乙)	1.5	1.0-1.0	考试	必修课
03110030	体育 I	1.0	0.0-2.0	考查	必修课
05110010	大学英语 I	3.0	2.0-2.0	考试	必修课
06110042	微积分(甲) I	4.5	4.0-1.0	考试	必修课
06110091	线性代数(甲)	3.0	3.0-0.0	考试	必修课
08110012	工程图学(乙)	2.5	2.0-1.0	考试	必修课
31110010	计算机文化	0.5	0.0-1.0		必修课
03110010	军事理论	1.5	1.0-1.0	考查	必修课
03110040	体育 II	1.0	0.0-2.0	考查	必修课
05110020	大学英语 II	3.0	2.0-2.0	考试	必修课
06110052	微积分(甲) II	4.5	4.0-1.0	考试	必修课
06110200	离散数学	4.0	4.0-0.0	考试	必修课

## □ 数据库（DB）

- 与企业相关的数据集合
- 具有完整性和持久性的数据集合。[R. Ramakrishnan, J. Gehrhe]
- 长期（常常多年）存在的信息集合。[Ullman]
- 长期存储在计算机内，有组织的，可共享的数据集合。[萨师煊，王珊]

## □ 数据库管理系统（DBMS）

- 数据库 + 一组用以访问、更新和管理这些数据的程序

# DBMS的主要特性

- ❑ 数据访问的高效和可扩展性
- ❑ 缩短应用开发时间
- ❑ 数据独立性（物理数据独立性 / 逻辑数据独立性）
- ❑ 数据完整性和安全性
- ❑ 并发访问和鲁棒性（恢复）

# DBMS的发展历史

- ❑ File processing system (1950s–1960s)
- ❑ Network and hierarchical DBMS (1960s–1970s)
  - 网状数据模型、层次数据模型 – 网状数据库、层次数据库（结构复杂、使用很困难）
- ❑ Relational database systems (RDBMS)
  - 关系模型 (1970, E. F. Codd)
  - RDBMS开始发展 (1970s)
  - RDBMS走向市场 (1980s)
  - RDBMS技术成熟 (1990s)

# DBMS的发展历史

- ❑ 面向对象数据库系统: Object-oriented database system (OODBMS)
- ❑ 对象关系数据库系统: Object-relational database systems (ORDBMS)
- ❑ 面向应用数据库系统: Application-oriented database systems
  - 空间、时间、多媒体、网络数据库
- ❑ 数据仓库 ( Data Warehousing )、联机分析处理 (Online Analytical Processing )、数据挖掘 ( Data Mining )

## □ 文件处理系统由传统操作系统所支持：

- 随着需求的增长，需要编写新的应用程序，并创建新的数据文件
- 但在相当长的时间内，数据文件可以是不同的格式。数据文件是相互独立的

## □ 在文件处理系统中存储组织信息的主要弊端：

- 数据冗余和不一致
  - 多种文件格式、信息重复存储
- 数据访问困难
  - 需要编写一个新的程序来完成每一个新的任务
- 数据孤立
  - 多文件多格式，检索、共享数据困难

## □ 在文件处理系统中存储组织信息的主要弊端：

### ■ 完整性问题

- 完整性约束（如账户余额 $>0$ ）成为程序代码的一部分
- 增加新的约束或更改现有的约束很困难

### ■ 原子性问题

- 在进行部分数据更新时，一旦发生故障，可能导致数据库处于不一致的状态
- 例如，从一个账户转移资金到另一个账户，此操作要么完成，要么根本不会发生

### ■ 并发访问异常

- 为了提高系统的总体性能，许多系统允许并发访问
- 不受控制的并发访问可能导致数据不一致
- 例如，两个用户读取同一账号余额，并在同一时间更新它



# Database systems VS File Processing Systems

- 在文件处理系统中存储组织信息的主要弊端：
  - 安全性问题
    - 并非所有用户都可以访问所有数据
- 数据库系统为以上所有问题，提供了解决方案
  - MySQL是LAMP的重要组成部分（Linux, Apache, MySQL, PHP/ Perl/ Python），