

引言（数据库系统的演变与发展）

四位图灵奖得主

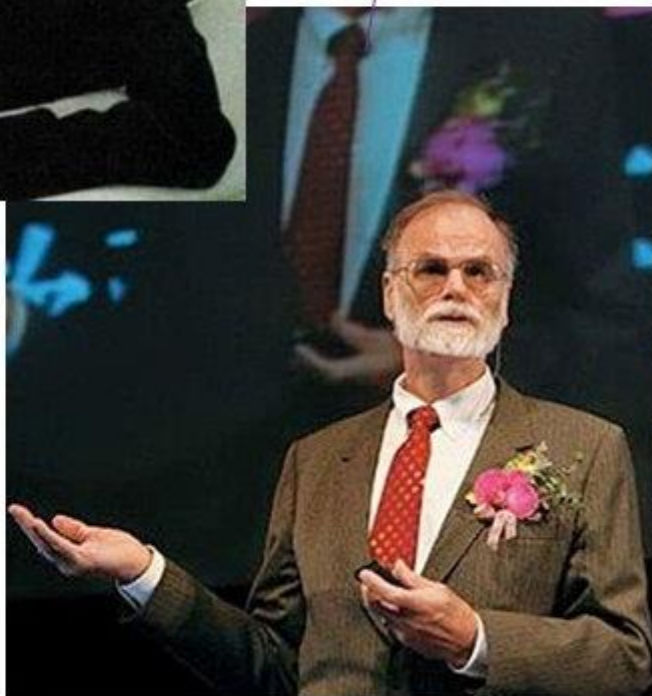


Charles W. Bachman
网状数据库，奠基者兼实践者



Jim Gray
事务处理，数据库系统实现

Michael Stonebraker
现代数据库概念与实践



E. F. Codd
关系数据库



数据库界的四位图灵奖得主

国内数据库研究做得比较好



- 李国良
- 清华大学
- AI4DB



- 李飞飞
- 阿里云
- 云原生



- 邹磊
- 北京大学
- 图数据库



你!

主要学习目标

- 数据管理的发展过程
- 概念模型
- 数据模型

前测小问题

数据库技术在哪些方面影响着我们的工作与日常生活？

一 引言

1.1 数据库技术如何影响我们的生活

数据库技术的重要性

(也包括任何一种技术或创新)

- 1) 提高效率
- 2) 改变或创新商业模式
- 3) 改变生活方式
- 4) 增加人类自由的维度

案例分析

- 以银行发展史为例：

固定地点存取（卡片/存折，开户与使用）



通存通兑



跨行跨国（包括自助银行）



信用卡（增加自由的维度）



微信支付（进一步增加自由的维度）



?

* 数据库技术的发展如何影响支付方式的发展？

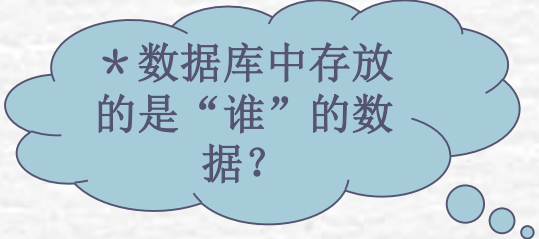
通过分析支付方式演变史，培养分析问题、批判思维和创新意识等能力

1.2 数据库如何描述客观世界

(一) 主要讨论问题

数据库中存放的是“谁”的数据？

- 谁: 一个应用系统（企业）需要的客观数据
- 涉及到数据对象的结构、范围和使用权限



* 数据库中存放的是“谁”的数据？

“谁”又是如何描述和表示的？

- 客观数据对象的**结构**即特征如何表示
- 客观数据对象的语义**约束**(有效性)如何表示
- 客观数据对象的使用**权限**如何表示



“谁”又是如何描述和表示的？

(二) “谁”的抽象描述方法

- 客观世界

是由**事**(注册, 选课)或**物**(学生, 课程)及相互联系构成

- 描述方法

通常地, 最直接和有效的方法是:

用**概念**来描述事或物及相互联系

- **事物**是所有概念的统称

概念: 是思维的逻辑单位

- 概念包含**内涵**和**外延**

内涵: 是概念的性质和**特征**

外延: 是满足上述性质及特征的所有客观个体的**集合**

(可在百度上查“事物”与“概念”等术语的解释)

如**学校**这样的应用系统，其客观数据对象(**事物**—**事**与**物**)及相互联系，可以这样来描述：

- **学生的内涵和外延：**（物）

学号，姓名，性别，籍贯，民族，…等**属性**（**内涵**）

属性也可以有许多分类，比如：

自然属性(性别，出生日期等)和社会属性(党员，国籍等)

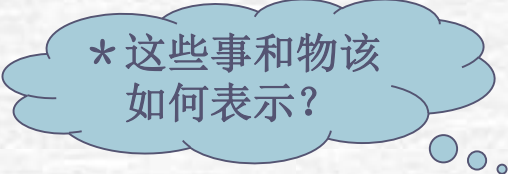
客观属性(住址，电话等)和主观属性(偶像，喜好，信仰等)

张三；李四；王麻子；…同类型的不同客观个体（**外延**）

- **课程的内涵和外延：**（物）

课程号，课程名，学分，学时数，…（**内涵**）

C语言，高等数学，大学物理，…（**外延**）



* 这些事和物该如何表示？

- **选课的内涵和外延：**（事）

简单有效的方式：

选课注册表：**学号**，**课程号**，选课日期，课程成绩，…（**内涵**）

注册表上的各个行…（**外延**）

其它可能的描述方式：学生课程之间有某种特殊联系

(四) 二维表与事物描述方法

- 在(关系)数据库中用二维表来描述概念:

内涵: 列(列即属性, 多寡取决应用需要)

外延: 行(满足上述内涵特征的所有个体的集合)

* 二维表是如何描述事物的?

教师编号	教师名	身份证号	护照号	医保卡
1001	张三	201202199009090371	G10221466	S231456756655656
1002	李四	201202199091212370	G34431244	S675697031277755

- 举例说明: 概念之间是如何相互区别和联系, 以及个体之间又是如何相互区别的?

(这正是下一步**主码**、**外码**概念的由来)

* 二维表又是如何联系和区分事物的?

- 数据库与客观世界的关系

数据库是数据化的客观世界

数据库设计是对客观世界的数据库化

学生: 学号, 姓名, 性别, 籍贯, 民族

课程: 课程号, 课程名, 学分, 学时数

选课: 学号, 课程号, 选课日期, 课程成绩

二 数据管理的发展过程

- 1) 手工阶段(早期—至今并存)
- 2) 程序阶段
- 3) 文件阶段(50年代末—60年代中)
- 4) 数据库阶段(60年代末—现在)

二 数据管理的发展过程

2.1 手工阶段

最早期的银行业务管理

（银行主要业务的手工管理过程）

1) 开户 或 销户 或 挂失

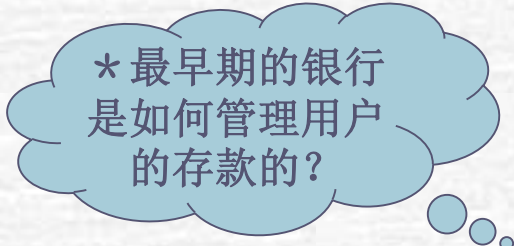
（每个人保留一张存折）

2) 存取款 或 转账

（每次变动在存折记录一行）

3) 账户余额查询

（查看自己的存折，未带在身边或遗失时到银行去寻问）#



* 最早期的银行
是如何管理用户
的存款的？

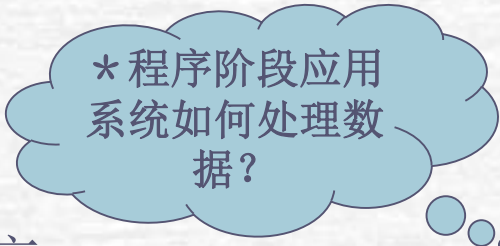
简而言之，就是**手动**地记录数据

2.2 程序阶段

早期的应用程序处理和管理数据

该阶段的基本特点：（主要任务是**完成自动计算**）

- 数据的管理者：应用程序，数据不保存。
- 数据面向的对象：某一应用程序
- 数据的共享程度：无共享、冗余度极大
- 数据的独立性：不独立，完全依赖于程序
- 数据的结构化：无结构
- 数据控制能力：应用程序自己控制 #



* 程序阶段应用系统如何处理数据？

简而言之，就是程序自己管理自己的数据，数据**不保存**，**无共享**

2.3 文件阶段

(50年代末--60年代中期)

* 文件阶段应用系统如何处理数据?

基于文件系统的数据库管理程序的数据管理的特点:

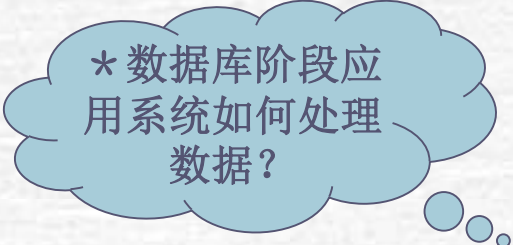
- 数据的管理者: 文件系统, 数据文件可长期保存
- 数据面向的对象: 某一(或极少数几个)应用程序
- 数据的共享程度: 共享性差、冗余度大
- 数据的结构化: 记录内有结构, 整体无结构
- 数据的独立性: 独立性差, 数据的存储方式以及逻辑结构改变必须大幅度修改应用程序, (牵一发而动全身)
- 数据控制能力: 应用程序自己控制 #

简而言之, 就是程序自己管理自己的数据, 数据保存, 少量共享

2.4 数据库阶段

(60年代末期—现在)

数据库管理阶段的基本特点：



* 数据库阶段应用系统如何处理数据？

- 数据的管理者：DBMS (与应用独立的专门管理软件)
- 数据面向的对象：现实世界(添加和删除方便易行)
- 数据的共享程度：共享性高
- 数据的独立性：高度的物理独立性和一定的逻辑独立性(随着不同发展时期，独立性程度在不断改进)
- 数据的结构化：整体结构化
- 数据控制能力：由DBMS统一管理和控制 #

简而言之，就是**独立于具体应用**，数据由DBMS统一管理

三 数据模型

3.1 什么是数据模型

* 什么是数据模型？

- 数据模型
 - 是一种用来**抽象、表示和处理**客观世界数据对象结构的描述方式
 - 是对客观世界的模拟(一种主观建模)
- 数据模型应满足如下要求
 - **形式化**(书面表示/书面语言)
 - 能够尽可能**真实**地反映客观世界
 - **容易**被人所理解
 - 便于在计算机上实现 #

* 数据模型应满足什么要求？

3.2 两类模型

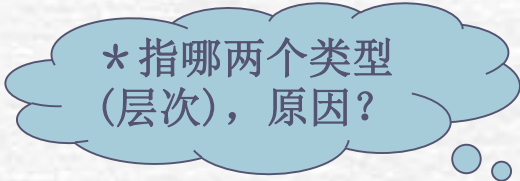
根据描述用途的不同，可分成两个层次：概念模型和数据模型

(1) **概念模型**：（也称信息模型）

是按用户的观点来对数据和信息建模；

几乎不涉及计算机专业技术知识。

（面向**客观**世界建模）



* 指哪两个类型
(层次)，原因？

(2) **数据模型**：

主要包括网状模型、层次模型、关系模型、对象模型等；

是按计算机系统的观点对数据建模。

（面向**计算机实现**建模）

3.3 数据库建模的两大阶段

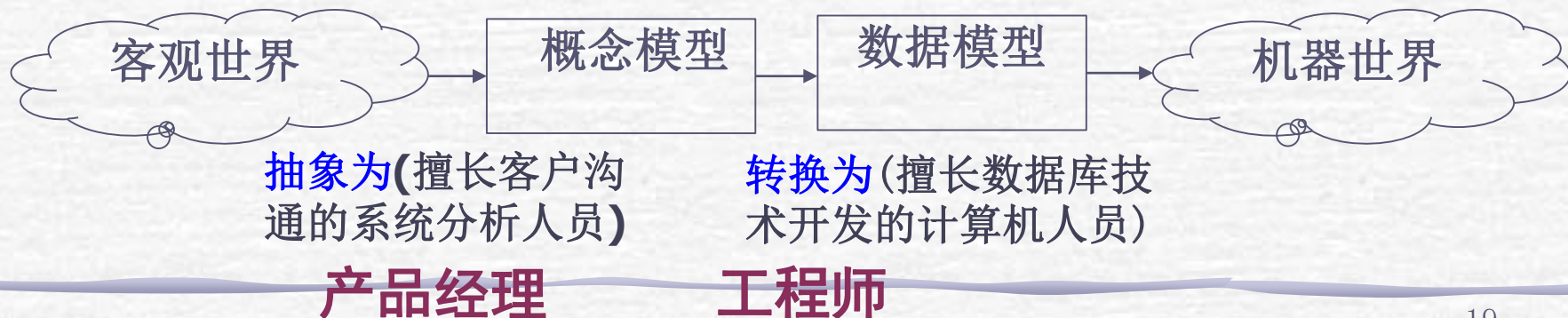
- 客观对象的抽象过程——**两步抽象**:

* 指哪两个抽象建模阶段?

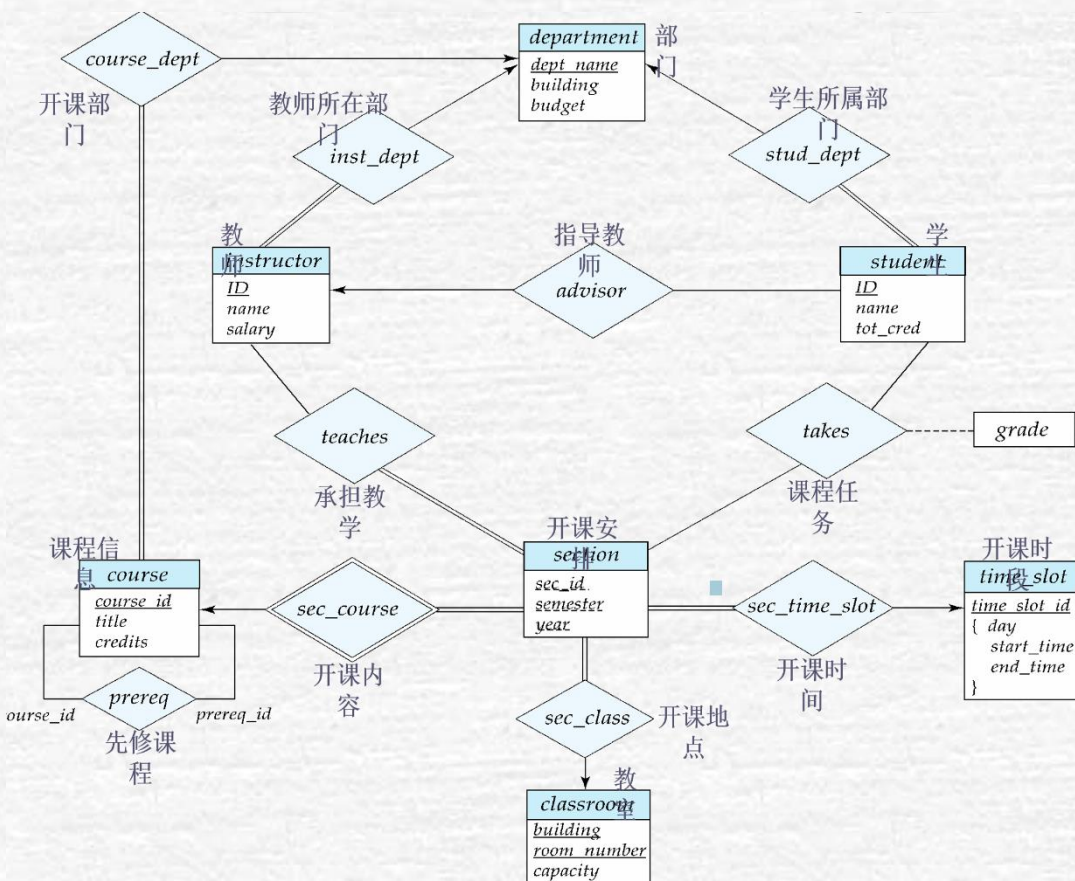
现实世界中的客观对象**抽象为**概念模型;
把概念模型**转换为**某一DBMS支持的数据模型。

- 概念模型的重要作用:
是现实世界到机器世界的一个中间层次;
可将业务模型与计算机实现工作隔离开;
复杂性减少, 便于分工, 成功性增大。

* 为何要分开为两个阶段?



数据库设计主要有哪些环节?



course_id	title	dept_name	credits
BIO-101	Intro. to Biology	Biology	4
BIO-301	Genetics	Biology	4
BIO-399	Computational Biology	Biology	3
CS-101	Intro. to Computer Science	Comp. Sci.	4
CS-190	Game Design	Comp. Sci.	4
CS-315	Robotics	Comp. Sci.	3
CS-319	Image Processing	Comp. Sci.	3
CS-347	Database System Concepts	Comp. Sci.	3
EE-181	Intro. to Digital Systems	Elec. Eng.	3
FIN-201	Investment Banking	Finance	3
HIS-351	World History	History	3
MU-199	Music Video Production	Music	3
PHY-101	Physical Principles	Physics	4

图 2-2 course 关系

ID	name	dept_name	salary
10101	Srinivasan	Comp. Sci.	65000
12121	Wu	Finance	90000
15151	Mozart	Music	40000
22222	Einstein	Physics	95000
32343	El Said	History	60000
33456	Gold	Physics	87000
45565	Katz	Comp. Sci.	75000
58583	Califieri	History	62000
76543	Singh	Finance	80000
76766	Crick	Biology	72000
83821	Brandt	Comp. Sci.	92000
98345	Kim	Elec. Eng.	80000

图 2-1 instructor 关系

随堂小测试

- 数据库逻辑设计后得到什么？
- 关系模型如何存储数据呢？

课后小结和作业安排

- 基本知识:
 - 事物和概念
 - 内涵和外延
 - 概念模型和数据模型
- 延展性学习:
 - 客观对象的抽象过程--两步抽象
- 作业

安装MySQL，快速学习MySQL基本知识：

<https://www.runoob.com/mysql/mysql-tutorial.html>