

第2讲 关系数据模型

赵英良
西安交通大学
2020



1

一、从现实到数据

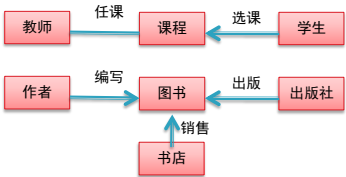
- 数据是信息和事物的符号表示。
- 事物的客观存在是**现实世界**。
- 事物的性质、状态的描述称为**信息世界**。
- 客观事物在信息世界中称为**实体**。
- 事物的性质在信息世界中称为**属性**。
- 信息世界中，实体可有若干属性来描述。如：
校名：西安交通大学，城市：西安，地址：咸宁西路28号。
- 信息数据化，称为数据世界。
- 10698，西安交通大学，西安，咸宁西路28号
- 每个实体的一组属性称为一条记录，每个属性称为一个数据项。



2

二、实体之间的联系

- 现实世界事物之间有联系



3



4

信息世界两个实体集之间的联系

- **一对一联系 (1:1)**
对于实体集A中的每一个实体，实体集B中**至多**有一个（也可以没有）实体与之联系，反之亦然，则称**实体集A与实体集B具有一对一联系**，记为1:1
- 实例
一个班级只有一个正班长
一个班长只在一个班中任职

班长	班级
李浩	化学71
王微	化学72

不是一一映射



5

两个实体集之间的联系（续）

- **一对多联系 (1:n)**
 - 对于实体集A中的每一个实体，实体集B中有n个实体（ $n \geq 0$ ）与之联系，反之，对于实体集B中的每一个实体，实体集A中**至多**只有一个实体与之联系，则称**实体集A与实体集B有一对多联系**，记为1:n
 - 实例
一个班级中有若干名学生，
每个学生只在一个班级中学习

班级	学生
化学71	李浩
化学72	王微
	张明



6

两个实体集之间的联系（续）

■ 多对多联系（m:n）

- 定义：如果对于实体集A中的每一个实体，实体集B中有n个实体（ $n \geq 0$ ）与之联系，反之，对于实体集B中的每一个实体，实体集A中也有m个实体（ $m \geq 0$ ）与之联系，则称实体集A与实体B具有多对多联系，记为m:n

■ 实例

课程与学生之间的联系：

一门课程同时有若干学生选修

一个学生可以同时选修多门课程

课程	学生
高等数学	101
C++	102
数据库	103

7

单个实体集内的联系

■ 实例

- 职工
- 职工中有经理、部门经理、组长
- 某一职工（干部）“领导”若干名职工
- 一个职工仅被另外一个职工直接领导
- 这是一对多的联系

单个实体集内也可以有一对一、一对多、多对多联系。

- 一对一是一对多的特例
- 一对多是多对多的特例

姓名	部门	职务
钱蒙	生产	部长
赵瀚	生产	组长
孙潇涵	生产	组长
王文博	生产	组长
马振华	开发	部长
吴舒	开发	组长
马雨龙	开发	组长

8

多个实体集之间的联系

- 若实体集E1, E2, ..., En存在联系，对于实体集Ej (j=1, 2, ..., i-1, i+1, ..., n) 中的给定实体，最多只和Ei中的一个实体相联系，则我们说Ei与E1, E2, ..., Ei-1, Ei+1, ..., En之间的联系是一对多的。
- 例如，一个学生可以申请多个项目，但必须只能指定一名指导教师。



9

三、数据模型的概念

■ 模型

- 依照原物或计划中的事物的形式做成的物品

■ model

- A model of an object is a physical representation that shows what it looks like or how it works.
- The model is often smaller than the object it represents.

■ 缩影

■ 数据模型

- 实体及实体间联系的描述，就是数据模型。
- 对数据、数据间关系及相关操作的描述。

10

模型的要求

- 数据模型应满足三方面的要求
- (1) 能比较真实地描述现实世界。
- (2) 容易被人理解。
- (3) 便于在计算机上实现。

11

■ 数据模型描述的内容包括：

- 数据结构：
 - 数据的内容、类型、性质及数据间的联系
- 数据操作：
 - 插入、删除、修改、排序、查找
- 完整性约束条件：
 - 语法、语义联系，制约、依存关系；变化规则；正确、有效、相容规则
- 数据模型的三要素

12

数据模型的层次分类

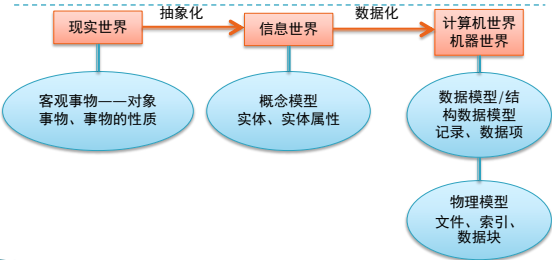
- 数据管理的不同阶段，采用不同的数据模型
- 在现实世界的抽象阶段，采用概念模型
- 在计算机逻辑实现阶段，采用逻辑数据模型
- 在计算机逻辑实现阶段，采用物理数据模型

数据模型的层次分类

- 概念数据模型——概念模型、信息模型
 - 信息世界中事物的描述；
 - 描述客观世界中的事物性质、结构及其内在联系
 - 将用于数据库设计。
 - 按用户的观点来对数据和信息建模
 - 与DBMS无关；与计算机平台无关；
 - E-R模型，扩充E-R模型，面向对象模型等。

- 逻辑数据模型——数据模型
 - 面向数据库系统的模型；是概念模型的进一步抽象；
 - 是实体在计算机世界的逻辑描述；
 - 按计算机系统的观点对数据建模，用于DBMS实现。
 - 考虑在计算机中如何表示和组织
 - 但仍与具体的数据库管理系统无关；是一般的逻辑描述；类似于算法。
 - 网状模型、层次模型、关系模型
- 物理数据模型——物理模型
 - 是对数据最底层的抽象，描述数据在系统内部的表示方式和存取方法，在磁盘或磁带上的存储方式和存取方法。
 - 是实现方式的描述

模型转换



四、概念数据模型

- 概念模型是现实世界中的事物及事物之间的联系在信息世界的表示。
- 是数据库设计人员的用于信息世界建模的工具。
- 是数据库设计人员和用户之间进行交流的语言
- 1976年P.P.S.Chen(陈品山)提出E-R(Entity-Relation)方法用实体和实体之间的联系描述事物和事物之间的联系



E-R方法

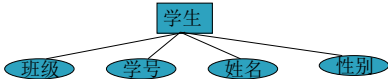
- 用图形化的方法表示实体、属性及实体之间的联系
- 用E-R图来描述现实世界的概念模型
- E-R图也称为E-R模型，实体-关系模型，实体-联系模型

E-R图

- 用**矩形**表示实体（**型**），矩形框内写明实体集的名字。



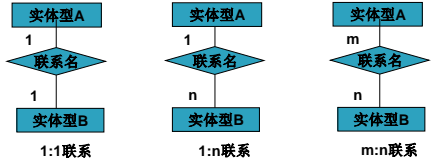
- 属性
用**椭圆形**表示，并用**无向边**将其与相应的实体连接起来
属性不能脱离实体，属性是相对实体而言的。



19

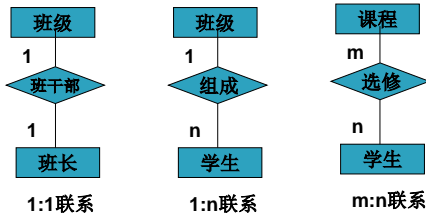
E-R图(续)

- 联系
用**菱形**表示，菱形内写明**联系名**，并用无向边分别与有关实体连接起来，同时在无向边旁标上联系的类型（1:1、1:n、n:1或m:n）



20

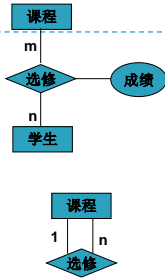
联系的表示方法示例



21

联系的属性

- ❖ **联系的属性：**
 - 联系也可以有属性。
 - 如果一个联系具有属性，则这些属性也要用无向边与该联系连接起来
- ❖ 一个实体集内的联系
- ❖ 一个课程是另一些课程的先修课程



22

一个实例

- 某设备公司物资管理
- 问题描述
采购物资在仓库中管理
(1) 一个仓库可以存放多种零件，一种零件可以存放在多个仓库中。
(2) 一个仓库有多个仓库保管员，一个职工只能在一个仓库工作。
(3) 仓库主任领导若干保管员。
(4) 一个供应商可以供应多种零件。
(5) 一个项目使用多种零件，一个零件用于多个项目中。
- 涉及的实体
仓库、职工（管理员、仓库主任）、零件、供应商、工程项目

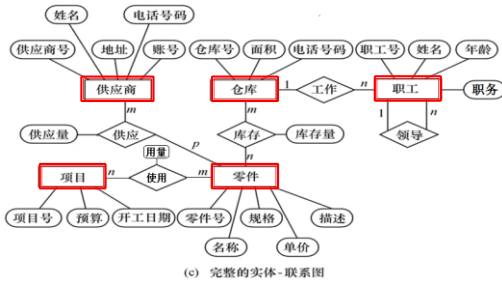
23

一个实例

- 实体的属性
 - 仓库：仓库号、面积、电话号码
 - 零件：零件号、名称、规格、单价、描述
 - 职工：职工号、姓名、年龄、职务
 - 供应商：供应商号、姓名、地址、电话号码、帐号
 - 项目：项目号、预算、开工日期

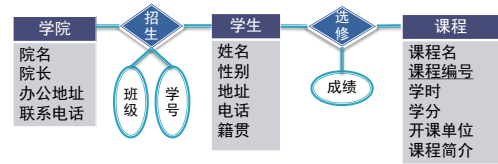
24

一个实例



25

E-R图的另一种画法



联系的类型?

UML(Unified Modeling Language),统一建模语言

26

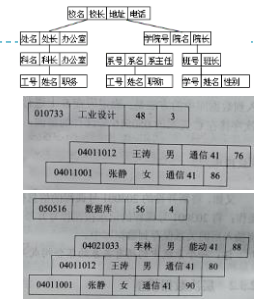
五、逻辑数据模型

- 概念模型
 - E-R 图, 用户角度描述现实, 现实世界->信息世界。
- 逻辑数据模型
 - 将概念模型转换为适合在计算机中表示的结构
 - 层次模型、网状模型、关系模型

27

层次模型

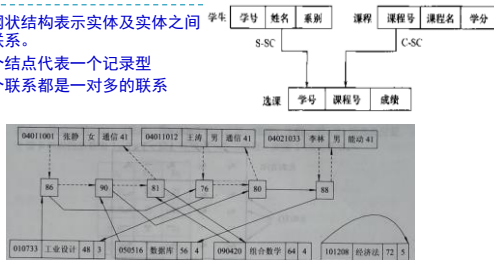
- 用树型结构表示实体及实体之间的联系。
- 每个结点是一个实体(型), 根无父结点
- 适合表示1:n的联系
- 信息的查找只能按树型路径查找



28

网状模型

- 用网状结构表示实体及实体之间的联系。
- 每个结点代表一个记录型
- 每个联系都是一对多的联系



29

关系模型

- 将数据组织成由若干行, 每行由若干列组成的表格形式。
- 这种表格称为关系。
- 每个实体集是一个关系表示
- 每个联系也是一个关系表示。

关系 S (选课表)					关系 SC (选课表)		
学号	姓名	性别	班级	课程号	课程名	学时	学分
04011001	张静	女	通信 41	010733	工业设计	48	3
04011012	王涛	男	通信 41	050516	数据库	56	4
04021033	李林	男	能动 41	090420	组合数学	64	4
				101208	经济法	72	5

30

六、关系和关系的性质

关系模型用**关系**表示实体集、实体集之间的联系

■ 一个关系是一个行列结构的二维表格

- 表有唯一名字→**关系名、表名**
- 每列称为一个**属性、字段**，
- 列有唯一名字：**属性名或字段名**。**属性和个数称为关系的度**；
- 第1行是各列的名字
- 以下每行称为一个**元组**，或**记录**
- 一个元组是由具体的一组属性值构成的，表示一个实体；一元组、二元组...。**属性的取值范围称为属性的域**。
- 元组的数量称为**关系的基数**

一个关系至少有一个属性，但可以没有元组

学生信息表

学号	姓名	年龄	性别	系名	年级
2015003	王明	19	女	社会学	2015
2015012	黄鹏	20	男	英语	2015
2015017	张文	18	女	法学	2015
...

关系的性质

- ①每列都是不可再分的基本字段
- ②列名是唯一的（列名不能相同）
- ③列是同质的。类型相同、值域、约束相同
- ④不同的列可以取值同一个域
- ⑤列的次序可任意交换，不会改变关系的意义
- ⑥行的次序可任意交换，不会改变关系的意义
- ⑦不允许有相同的元组

31

32

不可再分的基本字段？

职工号	姓名	职称	工资			扣除		实发
			基本	津贴	职务	房租	水电	
86051	陈平	讲师	1305	1200	50	160	112	2283

职工号	姓名	职称	基本	津贴	职务	房租	水电	实发
86051	陈平	讲师	1305	1200	50	160	112	2283

七、键

- 键（key）
 - 可以唯一地决定一个元组的**最小属性集**——
 - 键（key）、码、**候选码**、**候选键**（candidate key）
- 主属性：候选码中的属性
- 非主属性：不包含在任何候选码中的属性
- 超键（supper key），超码；全键；
- 主键（primary key），简称主键、主码
- 在数据库中，每个元组的主键的值不能为空，也不能与其他元组的主键值相同，这一条件称为**实体完整性约束**！！
- 键、码，与数据的逻辑意义有关，在DBMS中由用户设定！
- 主码的选择应该是那些**从不或极少变化的属性**

关系C(课程表)

课程号	课程名	学时	学分
010733	工业设计	48	3
030516	数据库	56	4
090420	组合数学	64	4
101208	经济法	72	5

33

34

- 外键(foreign key)
 - 设X是关系R的一个属性组，它并非R的主键，但却是另一个关系S的主键，X称为R关于S的**外键**

关系S(学生表)

学号	姓名	性别	班级
04011001	张静	女	通信41
04011012	王涛	男	通信41
04021033	李林	男	能动41

关系SC(选课表)

学号	课程号	成绩
04011001	010733	86
04011001	030516	90
04011001	090420	81
04011012	010733	76
04011012	030516	80
04021033	030516	88

- 学生表S中的主键为“学号”；
- 选课表SC中主键为“学号+课程号”；学号是外键；

作用？

八、关系模式

- 用关系名及写在括号中用逗号隔开的属性名表示关系的结构，这种表示称为**关系模式**。
- 关系模式是关系数据结构的描述。
- 学生（学号,姓名,性别,班级）
- 选课表（学号,课程号,成绩）
- 主码用下划线表示，且列在前面
- 格式为：
 - R(A1,A2,A3,...,An) 或 R(U)
 - 其中：R为关系名，Ai为属性名，U为属性名集合
 - 每个属性都有一个域
 - 关系模式和关系的关系？

35

36

记号

- 关系模式 $R(A_1, A_2, A_3, A_4, \dots, A_n)$
- 关系 R
- 元组 $t, t \in R$
- $t[A_i]$ —— t 的第 i 个分量, $t.A_i$, 元组在属性 A_i 的值
- $A = \{A_1, A_2, \dots, A_k\}$, A_1, \dots, A_k 是 A_1, \dots, A_n 的一部分, A 称为 **属性列** 或 **域列**
- $t[A]$ 是 t 在属性列上的取值
- $A' = U - A$

37

- 相互联系的若干关系组成关系数据库
- 关系模式及域的定义等是 **关系数据库模式**
- 在数据库管理系统中创建数据库,
 - 创建数据库文件(数据库名、文件名)
 - 关系模式(各种表的格式)
 - 定义各种约束条件

38

38

九、关系完整性约束

- 数据(元组、字段)在语义上的限制, 统称为 **完整性约束**
- 1. **实体完整性**
 - 每个关系都有一个主键, 每个元组的主键值是唯一的。即不空, 也不重复。
- 2. **引用完整性/参照完整性**
 - 外键的取值要么为空,
 - 要么是主表中相应属性的现有值之一。

关系 S (学生表)

学号	姓名	性别	班级
04011001	张静	女	通信 41
04011012	王涛	男	通信 41
04021033	李林	男	能动 41

关系 SC (选课表)

学号	课程号	成绩
04011001	010733	86
04011001	030516	90
04011001	090420	81
04011012	010733	76
04011012	030516	80
04021033	030516	88

39

- 3. **用户定义的完整性**
- 针对具体的列、元组的取值要求设置的, 如:
 - 年龄不超过 150,
 - 性别限定在{男, 女}中,
 - 电话号码限定 11 位数字, 数字+'-'+数字;
- 完整性约束应先设计好, 在定义数据库、数据表时由用户设定。
- 输入数据或数据改变时系统自动检查, 如果不符合条件, 不能输入数据, 给出提示信息。

40

40



谢谢

41

41