

Chapter2 数据模型

2.1 数据模型的基本概念

什么是数据模型

□什么是数据模型？

- 数据是对于现实世界的符号抽象，而数据模型则是对数据特征的抽象，为数据库系统的信息表示和操作提供一个抽象框架，是数据库系统的核心与基础

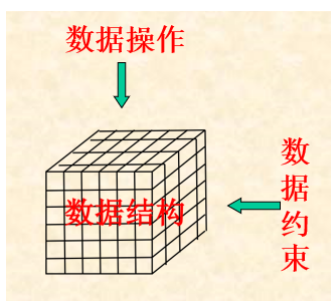
—数据模型应该能比较真实地模拟现实世界、易于人理解、便于在计算机上实现



数据模型

data model。定义：描述数据的结构，定义在该数据结构上可以执行的操作以及数据之间必须满足的约束条件

数据模型的组成：数据结构，数据操作，数据约束



数据结构：描述数据的类型、内容、性质以及数据间的联系。数据结构是一个数据模型的基础，数据操作与数据约束均是建立在相应的数据结构上的。这也是不同类型数据模型的划分依据

数据操作：在相应数据结构上可以执行的操作类型与操作方式。在不同的数据结构上可以提供不同的操作方式与操作类型。

数据约束：主要描述数据结构内数据间的相互关系，包括：数据间的语法/语义联系；数据间的制约与依存关系；数据（间）的动态变化规则，其目的是确保数据的正确、有效与相容。

□我们用建立在不同抽象层次上的‘数据模型’来表示每一步转化的结果：

—概念数据模型 (conceptual data model)

- 又简称为 ‘概念模型’

—逻辑数据模型 (logic data model)

- 又简称为 ‘数据模型’

—物理数据模型 (physical data model)

- 又简称为 ‘物理模型’

概念数据模型

侧重于对客观世界中复杂事物的结构描述及它们之间的内在联系的刻画，不涉及具体的描述细节和物理实现因素

E-R模型，EE-R模型；面向对象模型；谓词模型

逻辑数据模型

着重于数据模型在数据库系统一级的实现，即利用具体的DBMS所提供的工具（DDL）来定义的数据模型

是一种面向数据库系统的模型，概念数据模型只有在转换成逻辑数据模型后才能在数据库中得以表示

需要描述每个客观事物及其相互关系在选定的DBMS中的实现结构；即根据选定的DBMS来定义客观事物及其相互关系的实现结构。

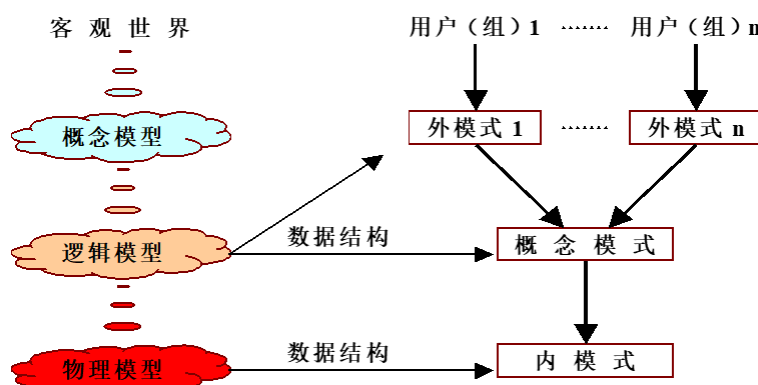
层次模型、网状模型；关系模型、面向对象模型、谓词模型；对象关系模型

物理数据模型

三种‘数据模型’与‘三级模式’之间的关系

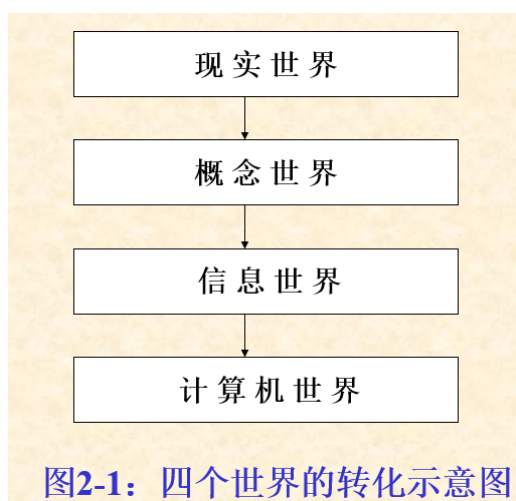
给出了数据模型在计算机内部的真正物理结构，是一种面向计算机物理实现的模型

一个概念数据模型将首先被转化为某一种逻辑数据模型，并通过所选择的DBMS将其进一步转化为具体的物理数据模型，才能使其在计算机中得以物理实现



2.2 数据模型的四个世界

现实世界，概念世界，信息世界，计算机世界

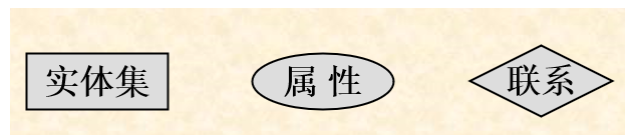


2.3 概念世界与概念模型

实体-联系(E-R)模型

E-R模型的图示法：E-R图（主要学会画E-R图）

基本概念的表达



每个实体集（联系）可以有多个属性，但每个属性只能隶属于一个实体集（联系）

为了刻画实体间的函数对应关系，必须在线段边上用1:1（一对一），1:n（一对多），n:m（多对多）等注明

关键字用下划线标记

扩充的实体-联系(EE-R)模型

IS-A联系

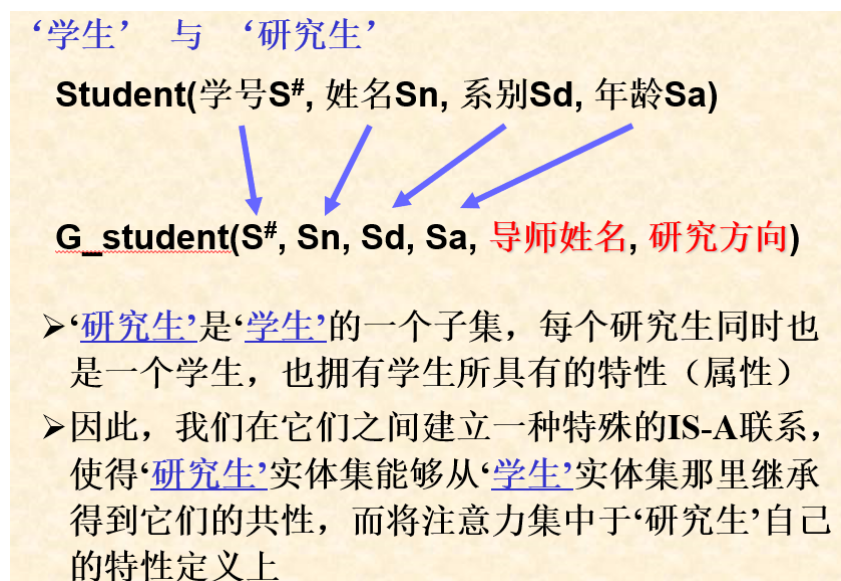
如果实体集B是实体集A的一个子集，且具有比实体集A更多的属性，则我们称在实体集A与实体集B之间存在着一种特殊的‘IS-A联系’。其中：

- 实体集A被称为实体集B的 超(实体)集
- 实体集B被称为实体集A的 子(实体)集



子集B可以通过IS-A联系继承超集A中的所有属性

例如：



符号：

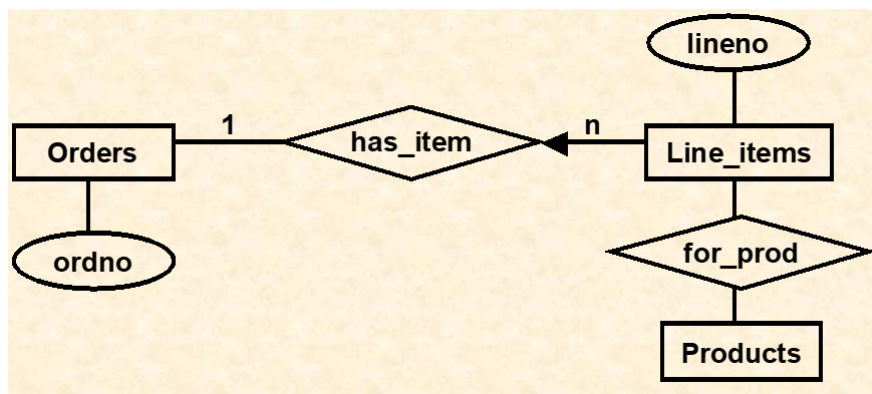
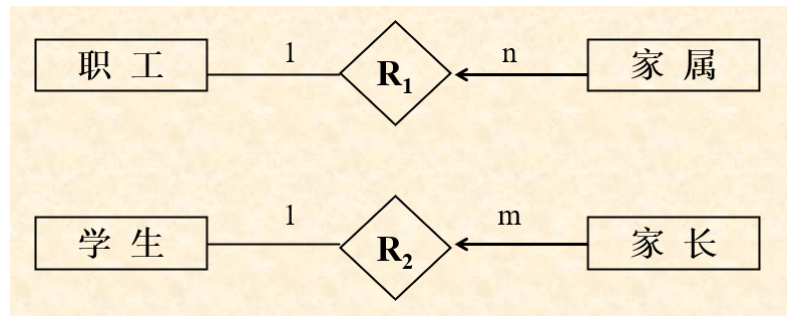
弱实体 (Weak Entity)

如果一个实体A的存在需要依赖于其他实体集中的某个实体的存在，那么实体A被称为弱实体。例如：

职工 vs 家属；学生 vs 家长

弱实体(集)与所依赖的实体(集)之间的函数对应关系应该是“多对一”的关系

例子：从弱实体到联系使用有向箭头



面向对象模型

我觉得不太可能考，看看课件过一遍算了，考到了算我倒霉。

类与类之间可以建立两种关系：‘IS-A’和‘IS-PART-OF’

统一建模语言UML

两个类之间构建一个UML的关联只需简单地通过在这两个类之间划一条线，并给这条线一个名字

每个与其他关联的类在连接的 对象的数量上有一定约束。这种约束通过在每个连接线的末端用一个m..n 标签形式来表明：这一端至少有m个对象，至多有n个对象与另外一端的对象连接。

用*代替n，表示“无限”，单独的*表示区间0..*

谓词模型

我觉得不太可能考，看看课件过一遍算了，考到了算我倒霉。

2.4 信息世界与逻辑模型

信息世界是数据库的世界，它着重于数据模型在数据库系统一级的构造与操作

信息世界用逻辑数据模型来进行描述

在数据库技术的发展历程中，具有比较重要的影响和历史地位的逻辑模型有：

层次模型和网状模型；关系模型和对象关系模型；面向对象模型；谓词模型。

概念模型与逻辑模型对应关系表

概念模型	E-R模型			EE-R模型	面向对象模型	谓词模型
逻辑模型	层次模型	网状模型	关系模型	对象关系模型	面向对象模型	谓词模型

关系模型

Relational model，是完全不同于层次模型和网状模型的一种新的逻辑模型。

关系模型的基本数据结构：二维表，简称‘表’ (Table)

关系模型的数据操纵：是建立在二维表上的操作，它包括对一张表及多张表间的查询，以及对一张表的删除，插入及修改等操作。

二维表 (Table)：二维表由表框架与元组所组成，表框架由若干个属性组成；存放于框架内的每‘一行数据’都被称为‘一个元组’ (Tuple)，或称‘行’ (Row)；一张二维表是由一个有n个属性的框架及m个元组组成。

关系

由行和列组成的二维表格

关系的约束

- 1)同一表中的属性名各不相同；2)表中的属性与属性的排放次序无关；3)表中的元组均不相同
- 4)表中的元组与元组的排列次序无关；5)表中的每一分量必须是一个不可分割的基本数据项

关系中的基本概念

关系模式：一个关系的属性名及其属性名的集合构成该关系的关系模式

关系数据库模式：该关系数据库中所有关系的关系模式的集合

元组：关系中的每一行

关键字：关系中的一个属性集的值能唯一标识关系中的一个元组，且又不含多余的属性值，则称该属性集为该关系的关键字。每一个关系都有关键字。一个关系也可以有多个关键字，所以关键字也被称为‘候选关键字’。

- 主关键字
可以从关系的候选关键字中选取一个作为该关系的主关键字
- 外关键字
设关系R中的属性集F，其取值来自于关系S中的主关键字K，则称属性集F是关系R的外关键字
关系R和关系S可以是同一个关系

关系模型上的数据操作

关系模型数据操作的对象是‘关系’，关系模型数据操作的结果也是一个‘关系’。

关系模型的五种基本操作：属性指定，元组选择，关系的合并，元组插入，元组删除。

2.5 计算机世界与物理模型

物理模型是面向计算机的模型，它构成数据库系统的物理实现

主要涉及操作系统级文件组织，有时还会涉及到硬件级数据组织

提高文件读写操作效率的方法

索引 (Index)

- 将文件中的记录与其物理地址（即磁盘块）间建立一张对应关系表以便于快速查找，这就是索引
- 索引一般也是一个文件。当数据文件中的记录数很大时，索引文件本身也还需要建立索引，这叫二级索引
- 依此类推，可以建立多级索引

Hash法

- 一种函数转换法，其主要思想是：通过一个hash函数将要查找的记录转换成该记录所在的物理地址，然后可以直接进行记录的定位读取操作

集簇 (Cluster)

- 在记录查找中往往需要按某项的项值查找，将具有相同或相邻项值的记录聚集在相同磁盘块内或圆柱体内以减少读盘次数，提高查找速度，这被称为集簇

本章小结

- 本章讨论数据模型，它是数据库系统的核心
- 数据模型
 - 基本概念
 - 三个抽象层次上的数据模型
- 概念数据模型
 - 四种概念模型
 - E-R模型，E-R图
 - EE-R模型：IS-A联系，弱实体
 - 面向对象模型
- 逻辑数据模型
 - 关系模型

