

第2章 信息与数据模型

人工智能学院
古 晶



第2章 信息与数据模型

- 2.1 信息的三种世界及描述
- 2.2 数据模型
- 2.3 概念模型
- 2.4 逻辑模型
- 2.5 概念模型向逻辑模型的转换
- 2.6 知识点小结
- 课堂测试



第2章 信息与数据模型

- 数据库系统的**核心**是数据模型。要为一个数据库建立数据模型，需要经过以下过程：
 - 1) 要深入到现实世界中进行系统需求分析。
 - 2) 用概念模型真实地、全面地描述现实世界中的管理对象及联系。
 - 3) 通过一定的方法将概念模型转换为数据模型。
- 本章主要介绍**信息的三种世界**以及彼此之间的联系；**概念模型**及相关概念；**概念模型转化为逻辑模型规则**；**数据模型**及作用、要素、优缺点。



第2章 信息与数据模型

□2.1 信息的三种世界及描述

□2.2 数据模型

□2.3 概念模型

□2.4 逻辑模型

□2.5 概念模型向逻辑模型的转换

□2.6 知识点小结

□课堂测试



2.1 信息的三种世界及描述

□信息的三种世界

- 1) 现实世界
- 2) 信息世界
- 3) 计算机世界（数据世界）

□ **转化流程**：**现实世界**存在的客观事物及其联系，经过人们大脑的认识、分析和抽象后，用物理符号、图形等表述出来，即得到**信息世界**的信息，再将信息世界的信息进一步 具体描述、规范并转换为计算机所能接受的形式，则成为**机器世界**的数据表示。



三个世界术语的对应关系

□ 现实世界、信息世界和计算机世界这3个领域是由**客观**到认识、由**认识**到使用**管理**的3个不同层次，后一领域是前一领域的抽象描述。

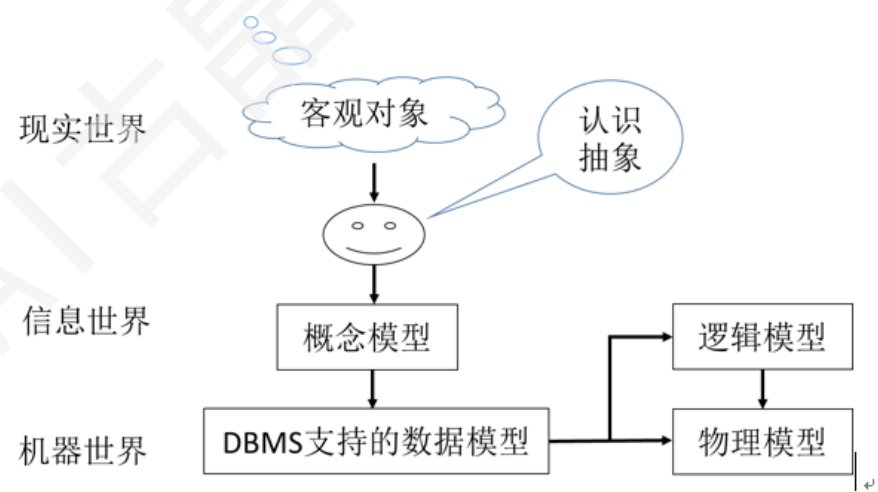
□ 对应关系：

现实世界	信息世界	计算机世界
实体	实例	记录
特征	属性	数据项
实体集	对象	数据或文件
实体间的联系	对象间的联系	数据间的联系
	概念（信息）模型	数据模型



三个世界术语的对应关系

- 现实世界的事物及联系，通过**需求分析**转换成为信息世界的**概念模型**（数据库设计人员）；
- 然后再把概念模型转换为计算机上某个DBMS所支持的**逻辑模型**（数据库设计人员和数据库设计工具DBMS）；
- 最后逻辑模型再转换为最底层的**物理模型**，从而进行最终实现（DBMS）。



将客观对象抽象为数据模型的完整过程



第2章 信息与数据模型

□2.1 信息的三种世界及描述

□2.2 数据模型

□2.3 概念模型

□2.4 逻辑模型

□2.5 概念模型向逻辑模型的转换

□2.6 知识点小结

□课堂测试



2.2.1 数据模型的概念

- **定义**：数据模型（Data Model）也是一种模型，它是对现实世界中**数据特征**及**数据之间联系**的抽象。
- 是实现**数据抽象**的主要工具，也是数据库系统中用于**信息表示**和**提供操作手段**的形式化工具。
- 是数据库系统的**核心**和**基础**。
- **数据模型满足的三个要求**：
 - 能比较**真实地模拟**现实世界；
 - 容易**理解**；
 - 便于在计算机上**实现**。



2.2.2 数据处理三层抽象描述

- 一种数据模型要很好地满足上述提到的三个方面的要求在目前尚很困难，在数据库系统中针对不同的使用对象和应用目的，通常采用**逐步抽象**的方法，在不同层次采用不同的数据模型。
- 一般可分为：
 - **概念层**
 - **逻辑层**
 - **物理层**



概念层

- **定义**：数据抽象级别的**最高层**，其目的是按用户的观点来对现实世界建模。
- 概念层的数据模型称为概念数据模型，简称**概念模型**。
- 概念模型**独立于**任何DBMS，但容易向DBMS所支持的逻辑模型转换。
- 常用的概念模型有**实体--联系模型**（ Entity-Relationship Model，简称**E-R模型**）。



- **定义**：是数据抽象的**中间层**，描述数据库数据整体的逻辑结构。
- 逻辑层的数据抽象称为逻辑数据模型，简称**数据模型**。
- 它是用户通过DBMS看到的现实世界，是基于计算机系统的观点来对数据进行建模和表示。
- **常见的数据模型**有：**层次模型**（ Hierarchical Model ）、**网状模型**（ Network Model ）、**关系模型**（ Relation Model ）和**面向对象模型**（ Object Oriented Model ）。



- **定义**：物理层是数据抽象的最底层，用来描述数据**物理存储结构**和**存储方法**。
- 这一层的数据抽象称为物理数据模型，它不但由DBMS的设计决定，而且与操作系统、计算机硬件密切相关。
- 物理数据结构一般都向用户屏蔽，用户不必了解其细节。



2.2.3 数据模型的组成要素

- 数据模型是严格定义的一组概念的集合。
- 这些概念精确地描述了系统的**静态特征**、**动态特征**和**完整性约束条件**。
- 因此**数据模型**通常由**数据结构**、**数据操作**和**完整性约束**三部分组成。



2.2.3 数据模型的组成要素

□ 数据结构

- **定义**：描述数据库的**组成对象**，以及对象之间的**联系**。
- **描述的内容**
 - 与数据**类型**、**内容**、**性质**有关的对象；
 - 与数据之间**联系**有关的对象。
- 数据结构是对系统**静态特性**的描述。



2.2.3 数据模型的组成要素

□ 数据操作

- **定义**：对数据库中各种对象(型)的实例(值)允许执行的操作及有关的**操作规则**。
- **类型**：**检索**和**更新**（包括插入、删除、修改）。
- **对操作的定义**：这些操作的确切含义、操作符号、操作规则（如优先级）以及实现操作的语言。
- 数据操作是对系统**动态特性**的描述。



2.2.3 数据模型的组成要素

□ 完整性约束条件

- 完整性约束条件的**定义**：
 - 是一组完整性规则。
 - 完整性规则主要描述数据结构中的数据之间的**语义联系**、数据之间的**制约**和**依存关系**，以及数据动态变化规则。
 - 数据约束主要用于保证数据的**完整性**、**有效性**和**相容性**。
- **数据模型对完整性约束条件的定义**
 - 数据模型应该反映和规定本数据模型**必须遵守的基本的通用的完整性约束条件**。
 - 提供定义**完整性约束条件**的机制，以反映具体应用所涉及的数据必须遵守的特定的语义约束条件。



2.3.4 数据模型与数据模式的区别

□ 主要区别：

- **数据模型**是描述现实世界数据的手段和工具。
- **数据模式**是利用这个手段和工具对相互间的关系所进行的描述，是关于型的描述。它与DBMS和OS硬件无关。



数据模型与数据模式的区别（续）

数据模型和数据模式都分了三个层次，其对应关系如下：

- ① **概念模式**：是用逻辑数据模型对**一个单位的数据**的描述。
- ② **外模式**：外模式也称子模式或用户模式，是与应用程序对应的数据库视图，是数据库的一个子集，是用逻辑模型对用户所用到的那**部分数据**的描述。
- ③ **内模式**：是**数据物理结构**和**存储方式**的描述，是数据的数据库内部表示方式。内模式也称存储模式。



第2章 信息与数据模型

- 2.1 信息的三种世界及描述
- 2.2 数据模型
- 2.3 概念模型**
- 2.4 逻辑模型
- 2.5 概念模型向逻辑模型的转换
- 2.6 知识点小结
- 课堂测试



2.3 概念模型

- 概念模型用于信息世界的建模，是现实世界到信息世界的**第一层抽象**。
- 是数据库设计人员进行**数据库设计**的有力工具。
- 是**数据库设计人员**和**用户**之间进行交流的**语言**：
 - 概念模型一方面应该具有较强的**语义表达**能力，能够方便、直接地表达应用中的各种语义知识。
 - 另一方面它还应该简单、清晰、易于用户理解。



2.3.1 概念模型的基本概念

(1) **实体** (Entity)

- 客观存在并可相互区别的事物称为实体。
- 可以是具体的人、事、物或抽象的概念。

(2) **属性** (Attribute)

- 实体所具有的某一**特性**称为属性。
- 一个实体可以由**若干个**属性来刻画。

(3) **码** (Key) : **唯一**标识实体的属性集称为码。

(4) **域** (Domain) : 属性的**取值范围**称为该属性的域。



2.3.1 概念模型的基本概念

(5) **实体型** (Entity Type) : 用**实体名**及其**属性名**集合来抽象和刻画同类实体称为实体型。

(6) **实体集** (Entity Set) : 同一类型实体的**集合**称为实体集。

(7) **联系** (Relationship)

- 现实世界中事物内部以及事物之间的联系在信息世界中反映为**实体内部的联系**和**实体之间的联系**。
- 实体内部的联系通常是指组成实体的**各属性之间**的联系。
- 实体之间的联系通常是指**不同实体集之间**的联系。



2.3.2 E-R模型

- **常用的概念层数据模型**有实体-联系（ Entity-Relationship , E-R ）模型、语义对象模型。
- E-R图由**实体**、**属性**和**联系**三个要素构成。
 - **实体**
 - 客观存在并可相互区别的事物称为实体。
 - E-R图中的实体用于表示现实世界具有相同属性描述的事物的集合，它不是某一个具体事物，而是某一种类别所有事物的统称。
 - 表示：**矩形框**，把实体名写在框内。
 - 每个实体**由实体名唯一标记**。
 - 每个实体对应于数据库中的一张**数据库表**，每个实体的具体取值对应于数据库表中的一条**记录**。



2.3.2 E-R模型

➤ 属性

- 用于表示实体的某种特征或者表示实体间关系的特征
- 用一个实体通常包含多个属性，每个属性由属性名唯一标记，画在**椭圆**内。
- E-R图中实体的属性对应于数据库表的**字段**。
- 在E-R图中，属性是一个**不可再分**的最小单元。
- 如果属性能够再分，则可以考虑将该属性进行细分，或者可以考虑将该属性“升格”为另一个实体。



2.3.2 E-R模型

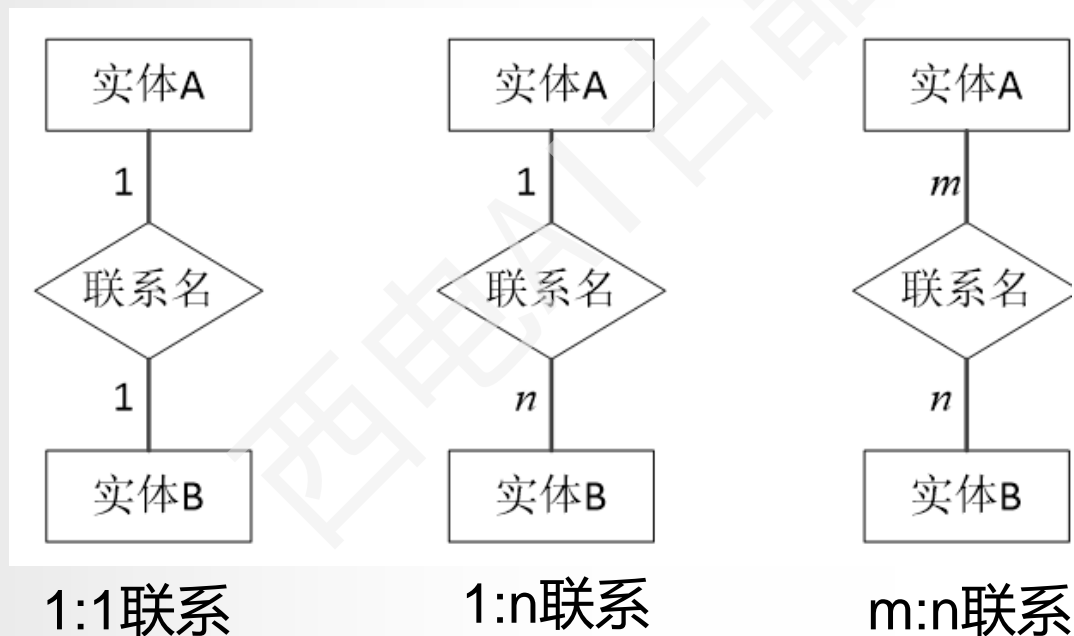
➤ 联系

- 数据之间的关联集合，是客观存在的应用语义链。
- 实体内部的联系通常是指组成实体的各属性之间的联系；。
- 实体之间的联系通常是指不同实体集之间的联系。
- 在E-R图中联系用菱形表示，框内写上联系名，并用连线将联系框与它所关联的实体连接起来。
- 基数：表示一个实体到另一个实体之间关联的数目。
- 从基数的角度将关系分为**一对一(1:1)**、**一对多(1:n)**、**多对多(m:n)关系**。



两个实体之间的三类联系

实体之间**三类联系**的图形表示：

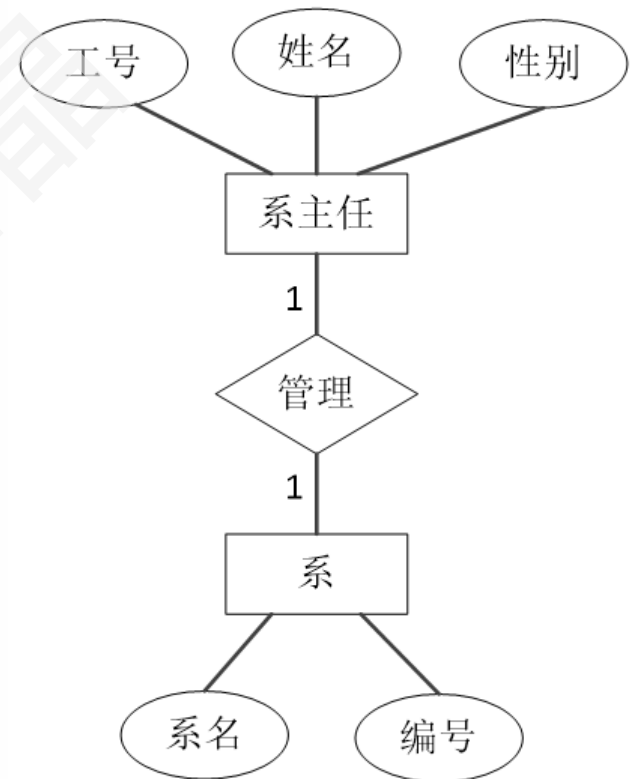




一对一联系 (1:1)

□ **定义**：如果实体集A中的每个实体，在实体集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有一对一联系，记为1:1。

□ **实例**：学校里一个系和正系主任（假设一个系只有一个正主任，一个人只能担任一个系的正系主任），则系和正系主任是一对一联系。

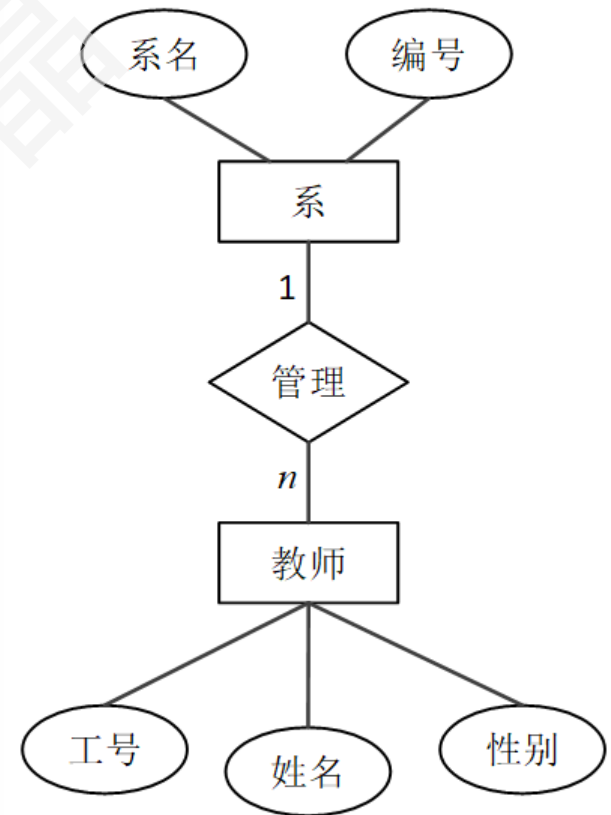




一对多联系 (1:n)

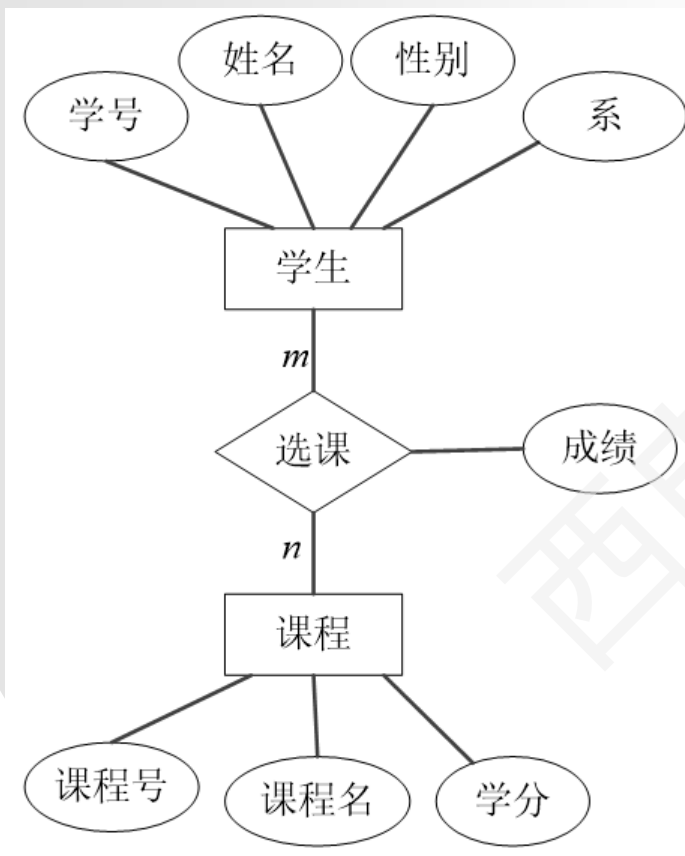
□ **定义**：如果对于实体集A中的每一个实体，实体集B中有 n 个实体 ($n \geq 0$) 与之联系，反之，对于实体集B中的每一个实体，实体集A中**至多**只有一个实体与之联系，则称实体集A与实体集B有一对多的联系，记为1:n。

□ **实例**：一个系有多名教师。而每个教师只能在一个系工作，则系和教师之间是一对多联系。





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中有m个实体 ($m \geq 0$) 与之联系，则称实体集A与实体集B有多对多的联系，记为m:n。

□ 实例：一门课程同时有若干个学生选修，而一个学生可以同时选修多门课程，则课程与学生之间具有多对多联系。



E-R模型设计原则与设计步骤

□ ER模型设计原则：

- 属性应该存在于且只存在于某一个地方（实体或者关联）。——避免**数据冗余**。
- 实体是一个单独的个体，不能存在于另一个实体中成为其属性。——避免“**表中套表**”。
- 同一个实体在同一个E-R图内**仅出现一次**。



E-R模型设计原则与设计步骤

□ E-R模型设计步骤：

- ①划分和确定**实体**。
- ②划分和确定**联系**。
- ③确定**属性**。作为属性的“事物”与实体之间的联系，必须是一对多的关系，作为属性的“事物”不能再有需要描述的性质或与其他事物具有联系。
- ④画出**E-R模型**。重复过程①～③，以找出所有实体集、关系集、属性和属值集，然后绘制E-R图。设计E-R分图，即用户视图的设计，在此基础上综合各E-R分图，**形成E-R总图**。
- ⑤**优化**E-R模型。利用**数据流程图**，对E-R总图进行优化，消除数据实体间**冗余**的联系及属性，形成基本的E-R模型。



第2章 信息与数据模型

- 2.1 信息的三种世界及描述
- 2.2 数据模型
- 2.3 概念模型
- 2.4 逻辑模型
- 2.5 概念模型向逻辑模型的转换
- 2.6 知识点小结
- 课堂测试



2.4 逻辑模型

- 数据模型是按其**数据结构**而命名的，根本区别在于数据之间联系的表示方式不同，即数据记录之间的**联系方式**不同。
- 最常用的**逻辑数据模型**及其**数据结构**：

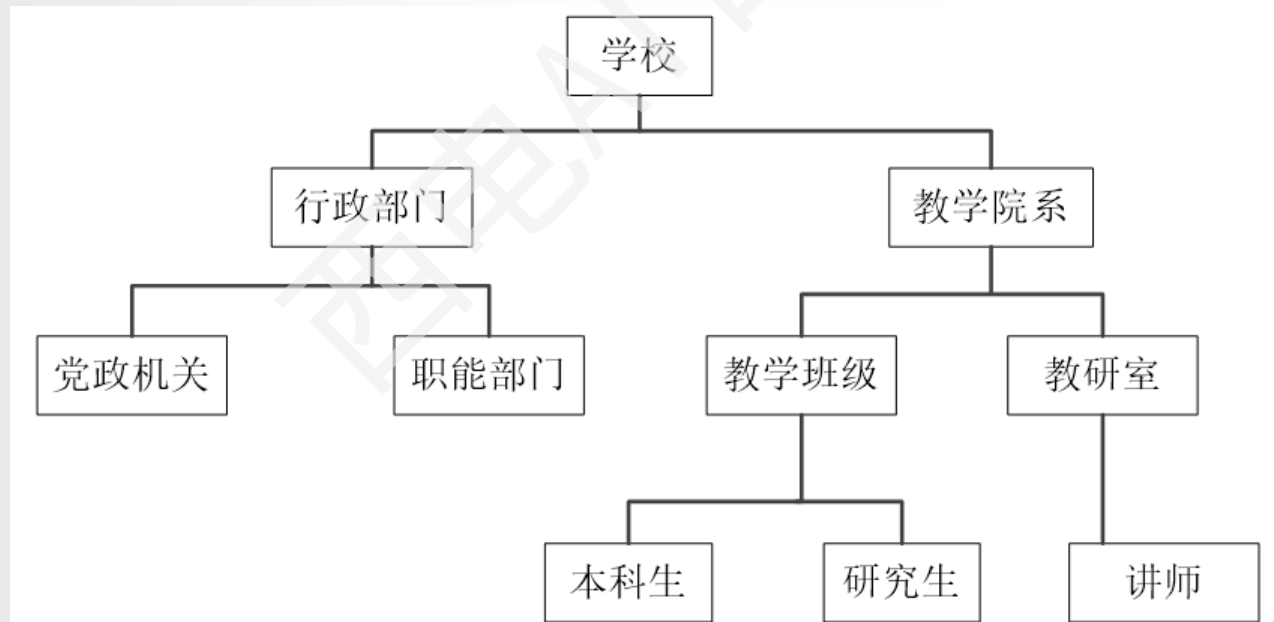
数据模型	对应的数据结构
层次模型	树结构
网状模型	图结构
关系模型	二维表（关系）
面向对象模型	引用类型



2.4.1 层次模型

□ 层次数据模型是数据库系统中最早出现的数据模型，它用**树形结构**表示各类实体以及实体间的联系。

□ 实例：





2.4.1 层次模型

- 层次模型对父子实体集间具有一**对多**的层次关系的描述非常自然、直观、容易理解。
- **突出问题**：
 - 在层次模型中具有一定的**存取路径**，需按路径查看给定记录的值。
 - 层次模型比较适合于表示数据记录类型之间的一对多联系，而对于**多对多**的联系难以直接表示，需进行转换，将其**分解**成若干个一对多联系。



2.4.1 层次模型

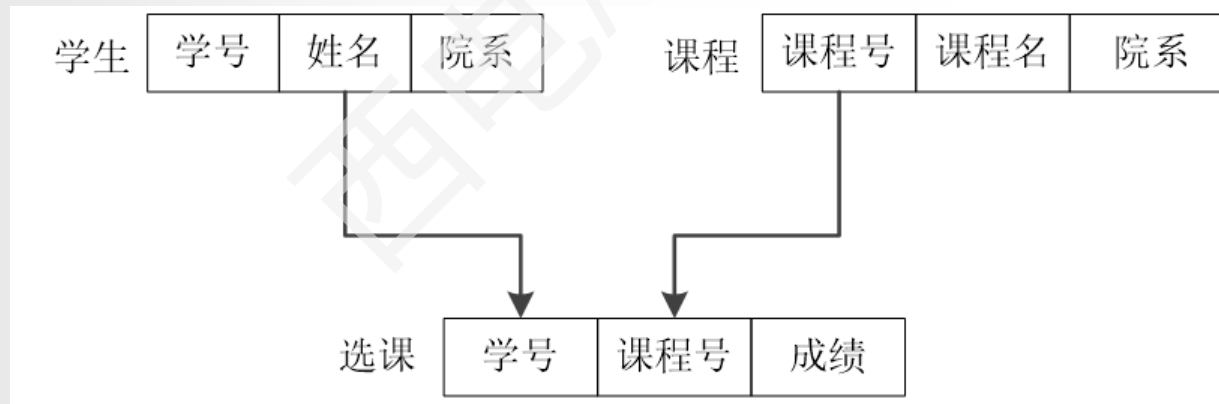
□ 主要优缺点：

- ①数据结构较**简单**；查询**效率高**。
- ②提供良好的**完整性支持**。
- ③**不易**表示多对多的联系。
- ④数据操作**限制多**、**独立性较差**。



2.4.2 网状模型

- 网状模型是一个**图结构**，它是由字段（属性）、记录类型（实体型）和系(set)等对象组成的网状结构的模型。
- 是一个不加任何条件的**有向图**。
- **实例：**





2.4.2 网状模型

- 网状模型是用图结构来表示各类实体集以及实体集间的系。
- 网状模型与层次模型的**根本区别**是：一个子结点可以有**多个**父结点；在两个结点之间可以有**多种**联系。
- 同样，网状模型对于多对多的联系难以直接表示，需进行**转换**，将其分解成若干个一对多联系。



2.4.2 网状模型

□ 主要优缺点：

- ① 较为直接地描述现实世界。
- ② **存取效率较高。**
- ③ 结构较**复杂、不易**使用。
- ④ **数据独立性较差。**



2.4.3 关系模型

□ **关系**就是一张**二维表**，它由行和列组成。**关系模型**将数据模型组织成表格的形式，这种表格在数学上称为关系。表中存放数据。在关系模型中实体以及实体之间的联系都用关系也就是二维表来表示。

□ **是最重要也是应用最广泛的一种基本模型。**

□ **实例：**

学号 ↴	姓名 ↴	性别 ↴	出生年月 ↴	所在系 ↴	入学年份 ↴
20170621 ↴	金小可 ↴	女 ↴	1999 ↴	计算机 ↴	2017 ↴
20170522 ↴	王大斌 ↴	男 ↴	2000 ↴	大数据系 ↴	2017 ↴
20170854 ↴	李一明 ↴	女 ↴	2001 ↴	会计系 ↴	2017 ↴
... ↴	... ↴	... ↴	... ↴	... ↴	... ↴



2.4.3 关系模型

□ 主要优缺点：

- ① 有坚实的**理论基础**；
- ② 结构**简单、易用**；
- ③ 数据**独立性及安全性**；
- ④ 查询效率**较低**。



2.4.4 面向对象模型

- 它是用**面向对象**的观点来描述现实世界中的事物（对象）的逻辑结构和对象间的联系等的数据库模型。
- **对象**：是对现实世界中的事物的高度抽象，每个对象是**状态**和**行为**的封装。
 - 对象的状态是属性的集合。
 - 行为是在该对象上操作方法的集合。
- 不仅可以处理各种复杂多样的数据结构，而且具有**数据**和**行为相结合**的特点。



2.4.4 面向对象模型

□ 优点：

- ①适合处理各种各样的数据类型。
- ②**面向对象程序设计**与数据库技术相结合。
- ③提高开发效率。
- ④改善数据访问。

□ 缺点：

- ①没有准确的定义。
- ②维护困难。
- ③不适合所有的应用。



第2章 信息与数据模型

- 2.1 信息的三种世界及描述
- 2.2 数据模型
- 2.3 概念模型
- 2.4 逻辑模型
- 2.5 概念模型向逻辑模型的转换**
- 2.6 知识点小结
- 课堂测试



2.5 概念模型向逻辑模型的转换

□ 要解决的问题：

- 如何将实体型和实体间的**联系**转换为关系模式。
- 如何确定这些关系模式的**属性**和**码**。

□ E-R图转换为关系模型转换的实质：

- 将实体、实体的属性和实体之间的联系转换为关系模式。



实体的转换

□ 原则：

- 一个实体对应一个关系模型；
- 实体的名称即是关系模型的名称；
- 实体的属性就是关系模型的属性；
- 实体的码就是关系模型的码。

□ 注意的问题：

- **属性域**的问题。
- **非原子属性**的问题。



联系的转换

□ 1:1联系转换

- **方法一**：将1:1联系转换为一个**独立的**关系模式，与该联系**相连**的各实体的**码**以及**联系本身的属性**均转换为关系模式的属性，**每个实体的码均是该关系模式的码**。
- **实例**：它描述的是实体学生和校园卡之间的联系，这里假设：一个学生只能办理一张校园卡，一张校园卡只能属于一个学生，因此，联系的类型是1:1。转换情况为：
 - 实体转换：学生（**学号**，姓名），校园卡（**卡号**，余额）；
 - 联系办卡的转换：办卡（**学号**，**卡号**，办卡日期）。

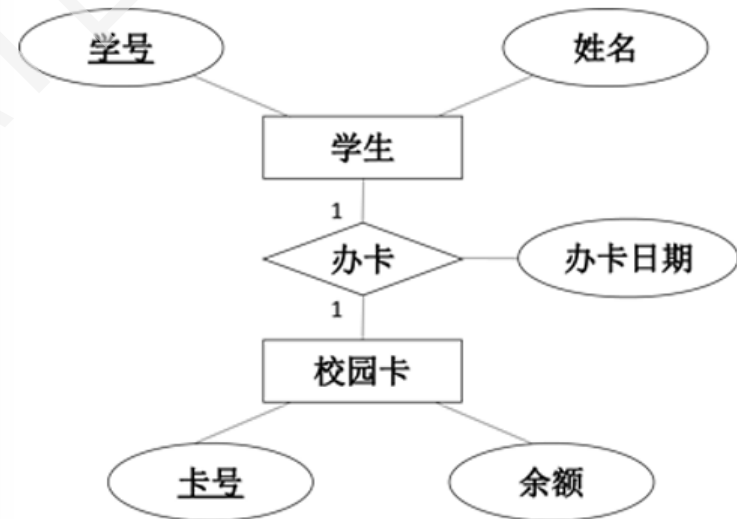


联系的转换

- **方法二**：与**任意一端**对应的关系模式合并。合并时，需要在该关系模式的属性中加入**另一个**关系模式的**码**和**联系本身的属性**。

- **转换情况**为：

- 学生（学号，卡号，姓名，办卡日期）
- 校园卡（卡号，学号，余额，办卡日期）





联系的转换

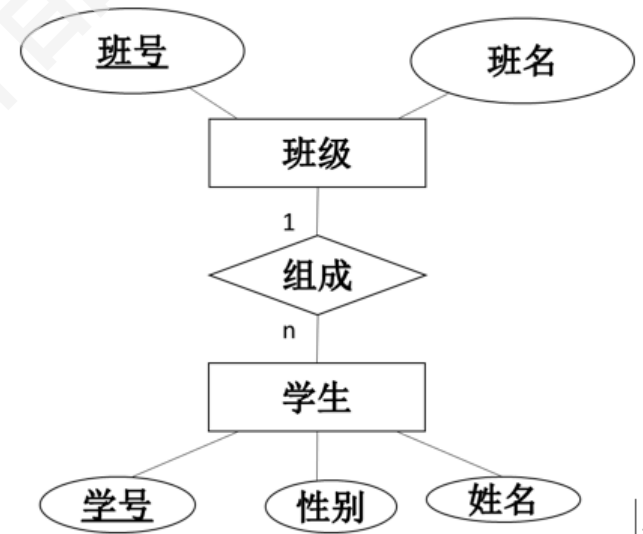
□ 1:n联系转换

- **方法一**：转换为一个**独立的**关系模式，与该联系相连的各实体的码以及联系本身的属性均转换为关系模式的属性，而**关系模式的码为n端实体的码**。
- **实例**：它描述的是实体学生和班级之间的联系，这里假设：一个学生只能在一个班级学习，一个班级包含多个学生。因此，联系的类型是1:n。
- **转换情况**为：
 - 实体转换：学生（学号，性别，姓名），班级（班号，班名）；
 - 联系组成的转换:组成（学号，班号）。



联系的转换

- **方法二**：与**n端**对应的关系模式合并，在该关系模式中加入**1端**实体的码和联系本身的属性。
- **转换情况**为：
 - 实体转换：学生（**学号**，性别，姓名），班级（**班号**，班名）。
 - 联系与学生一端合并，则关系模型学生变为：学生（**学号**，**班号**，性别，姓名）。





联系的转换

□ m:n联系转换

- m:n联系不能由一个实体的码唯一标识，必须由所关联实体的码**共同标识**。
- 需要将联系单独转换为一个**独立的**关系模式，与该联系相连的**各实体的码**以及**联系本身的属性**均转换为关系模式的属性，每个实体的码组成关系模式的码或关系模式的码的一部分。

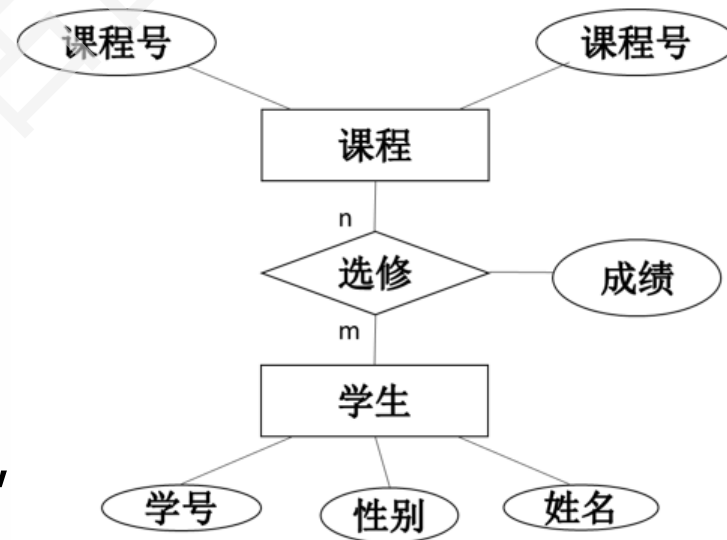


联系的转换

- **实例**：它描述的是实体学生和课程之间的联系，这里假设：一个学生可以选修多门课程，一门课程可以由多个学生选修。因此，联系的类型是m:n。

➤ **转换情况**为：

- 实体转换：学生（学号，性别，姓名）和课程（课程号，课程名）；
- 联系选修的转换：选修（学号，课程号，成绩）。





第2章 信息与数据模型

- 2.1 信息的三种世界及描述
- 2.2 数据模型
- 2.3 概念模型
- 2.4 逻辑模型
- 2.5 概念模型向逻辑模型的转换
- 2.6 知识点小结**
- 课堂测试



知识点小结

本章知识小结：

- 数据模型及其三要素
- 概念层模型
- 关系层数据库模型（逻辑模型）
- E-R模型
- E-R模型向关系模型的转换



第2章 信息与数据模型

- 2.1 信息的三种世界及描述
- 2.2 数据模型
- 2.3 概念模型
- 2.4 逻辑模型
- 2.5 概念模型向逻辑模型的转换
- 2.6 知识点小结
- 课堂测试



本章实验

□ 课堂测试：

见教材第41页27-29任选一题，画出概念模型（E-R图），将其转换为关系模式，并指出关系模式的码。



Thanks !

See you