## 1.数据库系统概述

* 数据库技术产生于六十年代末，是数据管理的最新技术，是计算机科学的重要分支。
* 数据库技术是信息系统的核心和基础，它的出现极大地促进了计算机应用向各行各业的渗透。
* 数据库的建设规模、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志。

### 1.1四个基本概念

数据（Data）

数据是数据库中存储的基本对象。

定义：描述事物的符号记录。

种类：文本、图形、图像、音频、视频、学生的档案记录、货物的运输情况等。

特点：数据与其语义是不可分的。

数据的含义称为数据的语义，数据与其语义是不可分的。

数据库（Database）

定义：数据库（Database，DB）是长期存储在计算机内、有组织的、可共享的大量数据的集合。

基本特征：

* 数据库按一定的数据模型组织、描述和存储。
* 可为各种用户共享
* 冗余度较小
* 数据独立性较高
* 易扩展

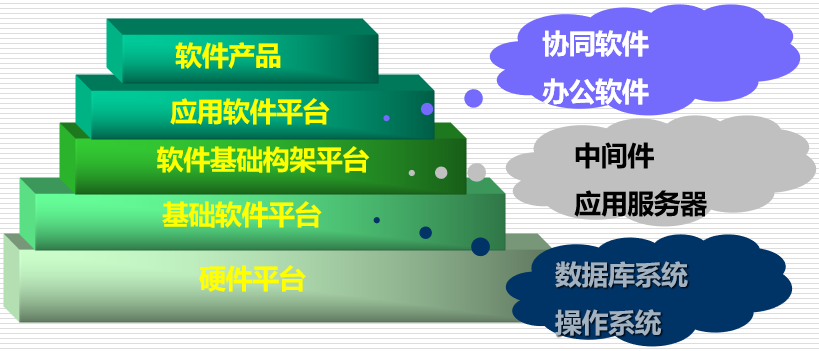
数据库管理系统（DBMS）

定义：位于用户与操作系统之间的一层数据管理软件。是基础软件，是一个大型复杂的软件系统。

用途：科学地组织和存储数据、高效得获取和维护数据。

主要功能：

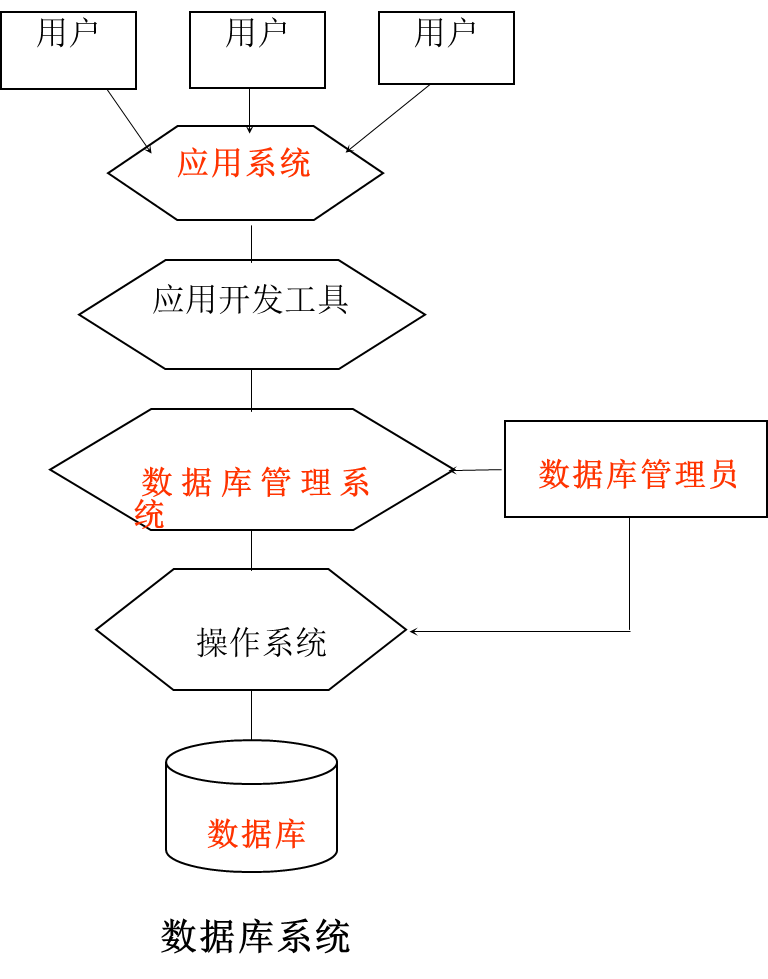
* 数据定义功能
  + 提供数据定义语言（DDL）
  + 定义数据库中的数据对象
* 数据组织、存储和管理
  + 分类组织、存储和管理各种数据
  + 确定组织数据的文件结构和存储方式
  + 实现数据之间的联系
  + 提供多种存储方法提高存取效率
* 数据操纵功能
  + 提供数据操纵语言（DML）
  + 实现对数据库的基本操作（查询、插入、删除和修改）
* 数据库的事务管理和运行管理
  + 数据库在建立、运行和维护时由DBMS统一管理和控制
  + 保证数据的安全性、完整性、多用户对数据的并发使用发生故障后的系统恢复
* 数据库的建立和维护功能（实用程序）
  + 数据库初始数据装载转换
  + 数据库转储
  + 介质故障恢复
  + 数据库的重组织
  + 性能监视分析等
* 其它功能：DBMS与网络中其它软件系统的通信、两个DBMS系统的数据转换、异构数据库之间的互访和互操作。



数据库系统（DBS）

Database System：由数据库及其管理软件组成的系统。

构成：数据库、数据库管理系统（及其开发工具）、应用系统、数据库管理员



### 1.2数据管理技术的产生和发展

什么是数据管理：

对数据进行分类、组织、编码、存储、检索和维护。

数据处理的中心问题。

数据库管理技术的发展过程：

人工管理阶段（20世纪40年代末——50年代中）

产生背景：

应用需求（科学计算）

硬件水平（无直接存取存储设备）

软件水平（没有操作系统）

处理方式（批处理）

特点：

数据的管理者：用户（程序员），数据不保存

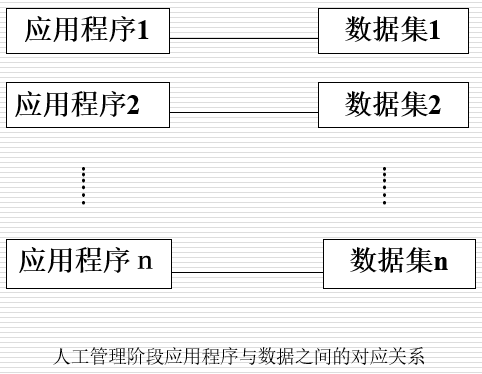
数据面向的对象：某一应用程序

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



文件系统阶段（20世纪50年代末——60年代中）

产生背景：

应用需求（科学计算、管理）

硬件水平（磁盘、磁鼓）

软件水平（有文件系统）

处理方式（联机实时处理、批处理）

特点：

数据的管理者：文件系统，数据可长期保存

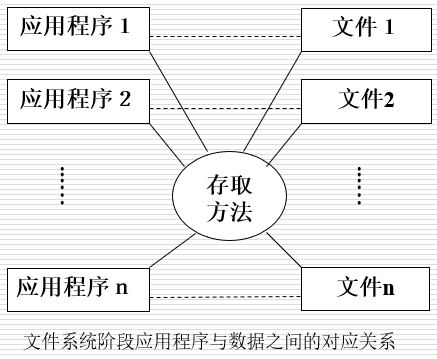
数据面向的对象：某一应用程序

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，数据的逻辑结构改变必须修改应用程序

数据控制能力：应用程序自己控制



文件系统中数据的结构：

记录内有结构

数据的结构是靠程序定义和解释的

数据只能是定长的（可以间接实现数据变长要求，但访问相应数据的应用程序复杂了）

文件间是独立的，因此数据整体无结构（可以间接实现数据整体的有结构，但必须在应用程序中对描述数据间的联系）

数据的最小存取单位是记录

数据库系统阶段（20世纪60年代末——现在）

产生的背景：

应用背景（大规模管理）

硬件背景（大容量磁盘、磁盘阵列）

软件背景（有数据库管理系统）

处理方式（联机实时处理，分布处理，批处理）

特点：

1. 数据结构化

整体数据的结构化是数据库的主要特征之一。

整体结构化：不再仅仅针对某一个应用，而是面向全组织；不仅仅数据内部结构化，整体是结构化的，数据之间具有联系。

数据库中实现的是数据的真正结构化：数据的结构用数据模型描述，无需程序定义和解释；数据可以变长；数据的最小存取单位是数据项。

1. 数据的共享性高，冗余度低，易扩充

数据系统从整体角度看待和描述数据，数据面向整个系统，可被多个用户、多个应用共享使用。

数据共享的好处：减少数据冗余，节约存储空间；避免数据之间的不相容性与不一致性；使系统易于扩充。

1. 数据独立性高

物理独立性：指用户的应用程序与存储在磁盘上的数据库中数据是相互独立的。当数据的物理存储改变了，应用程序不用改变。

逻辑独立性：指用户的应用程序与数据库的逻辑结构是相互独立的。数据逻辑结构改变了，用户程序也可以不变。

数据独立性是由DBMS的二级映像功能来保证的。

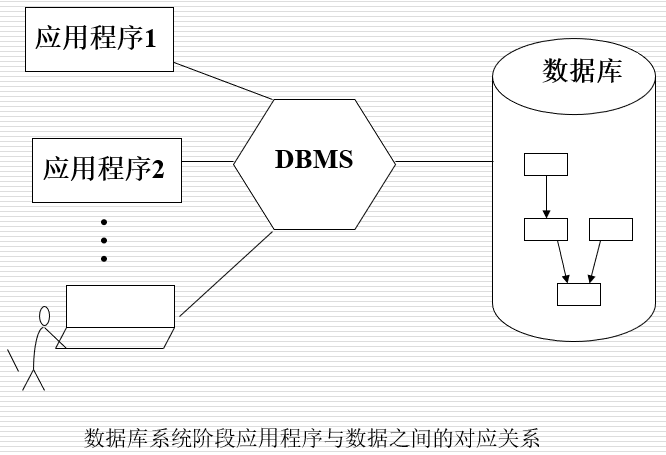
1. 数据由DBMS统一管理和控制

DBMS提供的数据控制功能：数据的安全性（Security）保护，保护数据，以防止不合法的使用造成的数据的泄密和破坏。

数据的完整性（Integrity）检查：将数据控制在有效范围内，或保证数据之间满足一定的关系。

并发（Concurrency）控制：对多用户的并发操作甲乙控制和协调，防止相互干扰而得到错误的结果。

数据库恢复（Recovery）：将数据库从错误状态恢复到某一已知的正确状态。



数据库管理技术的发展动力：

应用需求的推动

计算机硬件的发展

计算机软件的发展

## 2.数据模型

在数据库中用数据模型这个工具来抽象、表示和处理现实世界中的数据和信息。

通俗地讲数据模型就是现实世界的模拟。

数据模型应满足三方面要求：

能比较真实地模拟现实世界；容易为人所理解；便于在计算机上实现。

### 2.1两大类数据模型

1）概念模型（+信息模型）

它是按用户的观点来对数据和信息建模，用于数据库设计。

2）逻辑模型和物理模型

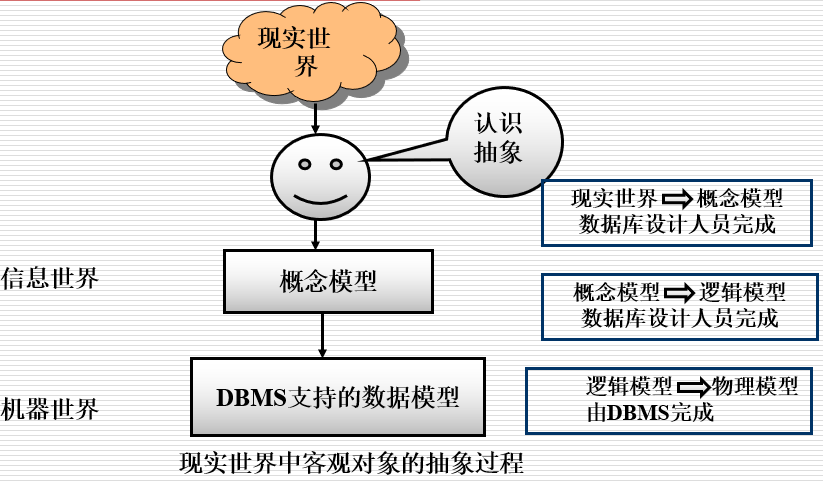
按计算机系统的观点对数据建模，用于DBMS实现。

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。

客观对象的抽象过程——两步抽象：

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。



### 2.2数据模型的组成要素

数据结构、数据操作、完整性约束条件

1）数据结构

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）数据操作

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。

数据操作的类型：查询、更新（包括插入、删除、修改）

数据模型对操作的定义：操作的确切含义，操作符号，操作规则（如优先级），实现操作的语言。

数据操作是对系统动态特性的描述。

3）完整性约束条件

一组完整性规则的集合。

完整性规则：给定的数据模型中数据及其联系所具有的的制约和存储规则。

用以限定符合数据模型的数据库状态及其状态的变化，以保证数据的正确、有效、相容。

数据模型对完整性约束条件的定义：

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。例如在关系模型中，任何关系必须满足实体完整性和参照完整性两个条件。

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概念模型

概念模型的用途：

概念模型用于信息世界的建模；

是显示世界到机器世界的一个中间层次；

是数据库设计的有力工具；

数据库设计人员和用户之间进行交流的语言。

对概念模型的基本要求：

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；

简单、清晰、易于用户理解。

1）信息世界中的基本概念

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、或抽象的概念。

属性（Atttibute）：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。

码（key）：唯一标识实体的属性集称为码。

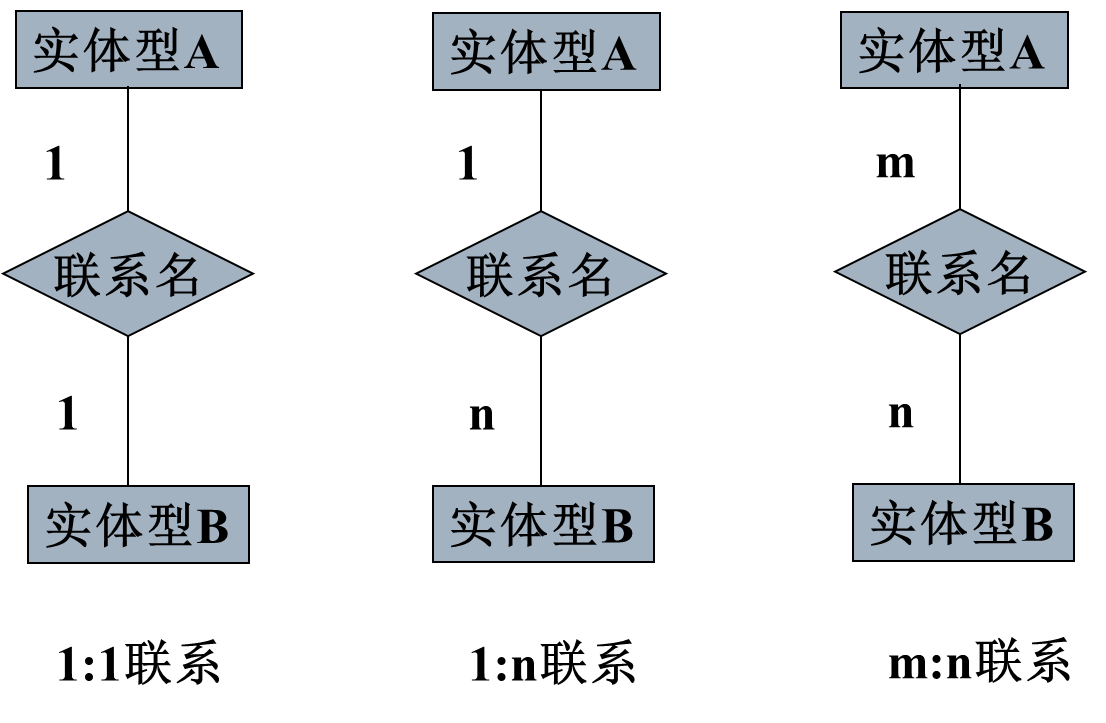
域（Domain）：属性的取值范围称为该属性的域。

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。

实体集（Entity Set）：同一类型实体的集合称为实体集。

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。

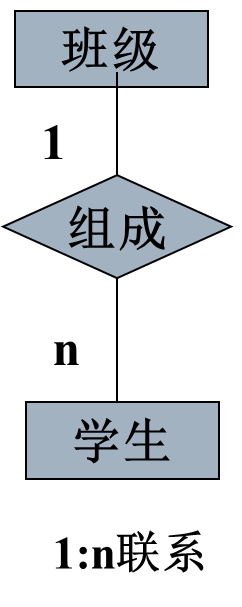
2）两个实体型之间的联系



一对一联系（1:1）：如果对于实体集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。



一对多联系（1：n）：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>=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。

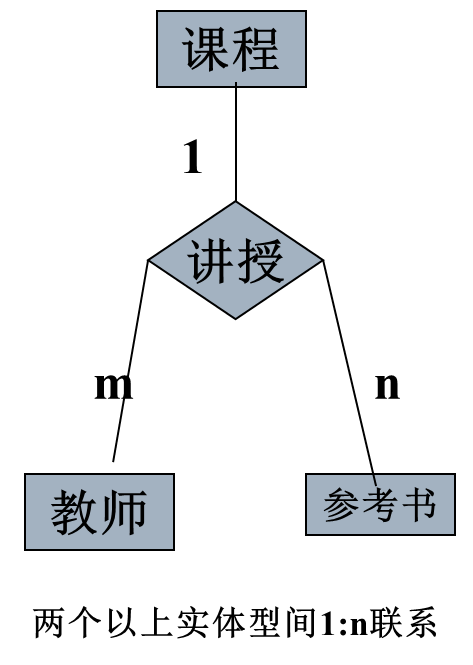


多对多联系（m：n）：如果对于实体集A中的每一个实体，实体集B中有n个实体（n>=0）与之联系。反之，对于实体集B中的每一个实体，实体集A中也有m个实体（m>=0）与之联系，则称实体集A与实体B具有多对多联系。



3）两个以上实体型之间的联系

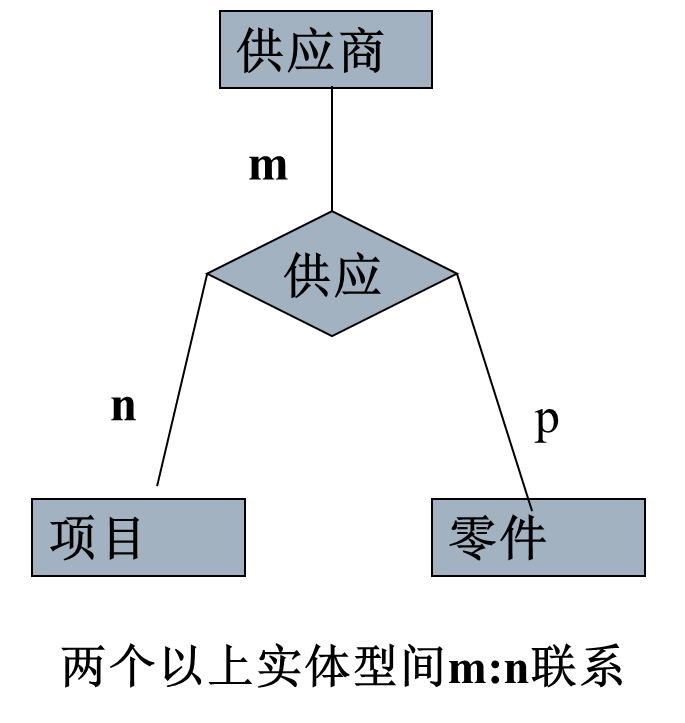
两个以上实体型之间一对多联系：若实体集E1，E2，...，En存在联系，对于实体集Ej（j=1，2，...，i-1，i+1，...，n）中的给定实体，最多只和Ei中的一个实体相联系，则我们说Ei与E1，E2，...，Ei-1，Ei+1，...，En之间的联系是一对多的。



（一门课程可以有若干个教师教授，使用若干本书，每一个教师只讲授有一门课程，每一本参考书只供一门课程使用。）

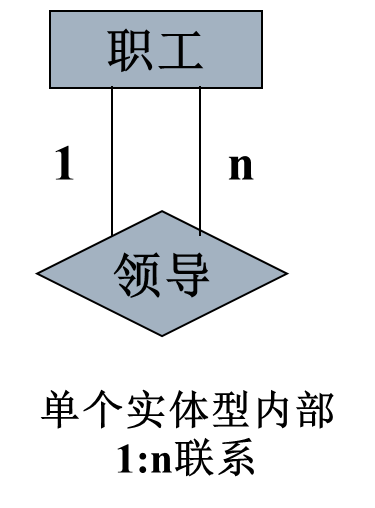
多个实体型间的一对一联系。

两个以上实体型间的多对多联系：

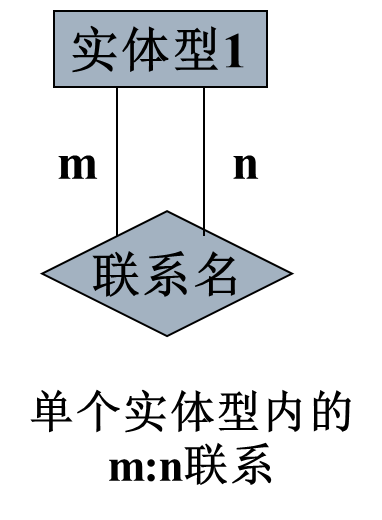


4）单个实体型内的联系

一对多联系：职工实体型内部具有领导被领导的联系。



多对多联系：



5）概念模型的一种表示方法

实体—联系方法（E-R方法）

用E-R图来描述现实世界的概念模型，E-R方法也称为E-R模型。

E-R图

* 实体型：用矩形表示 ，矩形框内写明实体名。
* 属性：用椭圆型表示，并用无向边将其与相应的实体连接起来。

6）一个实例

## 3.数据库系统结构

## 4.数据库系统的组成

## 5.小结