# 数据库管理系统

## 数据库系统概述

1.1 数据库技术的发展历史阶段

|  |  |
| --- | --- |
| 阶段名称 | 特点 |
| 人工管理阶段 | 1. 数据基本不保存 2. 没有对数据进行管理的软件系统 3. 没有文件的概念 4. 数据不具有独立性 |
| 文件系统阶段 | 1. 数据可以长期保存 2. 由文件系统管理数据 3. 文件的形式已经多样化 4. 数据具有一定的独立性 |
| 数据库系统阶段 | 1. 采用复杂的结构化的数据结构 2. 较高的数据独立性   **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。 数据的逻辑结构改变了，用户程序也可以不变。   1. 最低的冗余度 2. 数据控制功能 |

1.2 数据库系统的基本概念

数据 ------ 数据库中存储的基本对象。

数据库 ------ 存放数据的仓库，这个仓库就是计算机存储设备。

数据库管理系统（DBMS DataBase Management System）------- 专门用于管理数据库的计算机系统软件，介于应用程序与操作系统之间，是数据管理软件。其能够为数据库提供数据的定义、建立、维护、查询和统计等操作功能，并具有对数据完整性、安全性进行控制的功能。

数据库系统（DataBase System，DBS）------- 指在计算机系统中引入了数据库后的系统，由计算机硬件、数据库、DBMS、应用程序（Application）和用户（User）构成，即由计算机硬件、软件和使用人员构成。

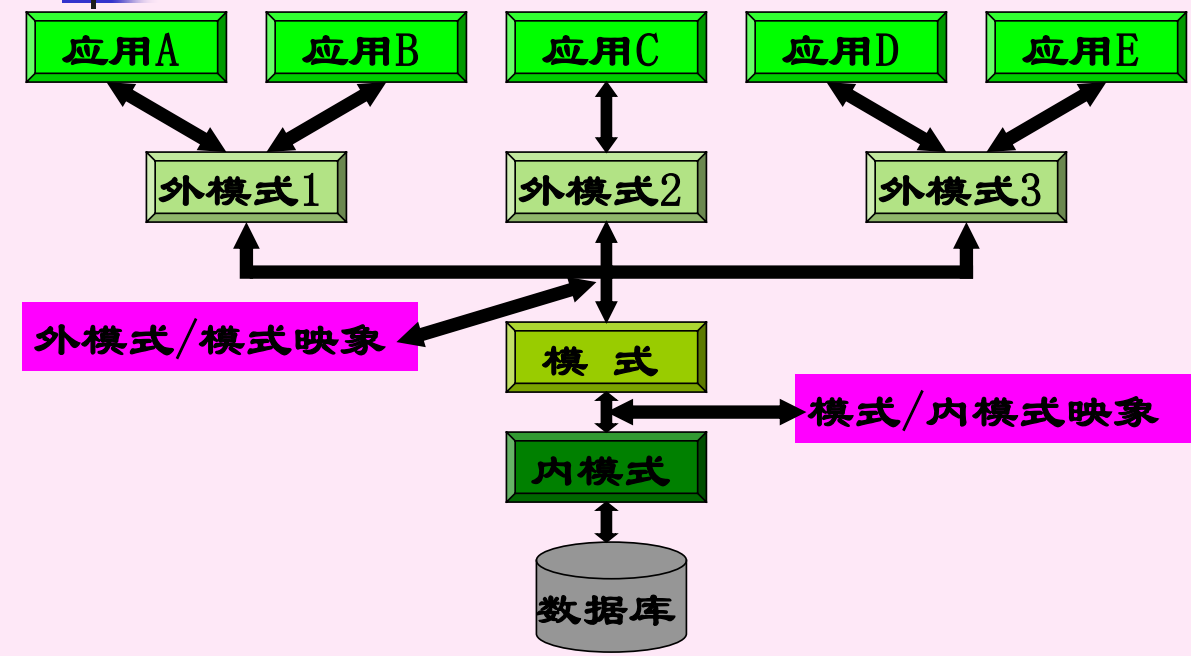
补充：

DBMS（数据库管理系统）提供的数据控制功能  
 (1)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。  
 (2)数据的完整性（ Integrity）检查：  
 将数据控制在有效的范围内，或保证数据之间满足一定的关系。  
 (3)并发（ Concurrency）控制 ：  
 对多用户的并发操作加以控制和协调，防止相互干扰而得到错  
误的结果。  
 (4)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。

1.3 数据库系统的三级模式结构

外模式（ External Schema）  
模式（ Schema ）  
内模式（ Internal Schema）



**模式**也成为逻辑模式，也是数据库中的全体的逻辑结构和特征的描述，是所有用户的公共数据视图。 数据库系统提供数据模式描述语言（Data Description Language DDL）来严格地表示这些内容。

**外模式**也称为子模式或用户模式，是数据库用户看到的数据视图，是与一应用有关的数据的逻辑表示。 数据库系统提供**外模式描述语言**来描述**用户数据视图**。

**内模式**也称为存储模式，是数据在数据库系统内部的表示或底层描述，即对数据库物理结构和存储方式的描述。**一个数据库只能有一个内模式**。

数据库系统提供**内模式描述语言**来描述数据库的**物理存储**。

**数据库系统的二级映射**

|  |  |  |
| --- | --- | --- |
| 名称 | 描述 | 用途 |
| **外模式／模式映象** | 定义外模式与模式之间的对应关系 每一个外模式都对应一个外模式／模式映象 映象定义通常包含在各自外模式的描述中 | 保证数据的逻辑独立性 |
| **模式/内模式映像** | 模式／内模式映象定义了数据全局逻辑结构与存储结构之间的对应关系。 例如，说明逻辑记录和字段在内部是如何表示的  **数据库中模式／内模式映象是唯一的** 该映象定义通常包含在模式描述中 | 保证数据的物理独立性 |

# 数据模型

**现实世界**

指**存在于人脑之外**的客观世界，泛指客观存在的事物及其相互间的联系。

**一个**实际存在并可以识别的**事物**称为**个体**，每个**个体**都有自己的**特征**，用以区别其他个体

实体：现实世界中存在的可以相互区分的客观事物或概念。

实体特征

实体集及实体集间的联系

具有相同特征或能用同样特征描述的实体的集合称为实体集。

**信息世界**

现实世界中的事物反映到人们的**头脑里**，经过认识、选择、命名、分类等综合分析而形成了概念和认识，这就是信息，即进入了信息世界

在信息世界中，每一个**被认识的个体**称为**实体**，个体的**特征**在头脑里形成的知识称为**属性**。

一个实体是由它所有的属性来表示的。

实例：实体通过其特征的表示。

属性：实体的特征在人们思想意识中形成的知识。

对象及对象间联系

对象：同类实例的集合，即实体集中的实体用属性表示得出的信息集合。

**计算机世界** 又称数据世界，是将信息世界中的信息经过抽象和组织，按照特定的数据结构，即数据模型，将数据存储在计算机中。

字段：用来标记实体的一个属性。

记录：有一定逻辑关系的字段的组合。

文件：同一类记录的集合。

文件及：若干文件的集合，即由计算机操作系统通过文件系统来组织和管理。

--------------------------------------------------------------------------------------------------------------------------------

**客观对象的抽象过程**

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；

概念模型 也称信息模型，它是按用户的观点来对数据和信息建模。

把概念模型转换为某一DBMS支持的数据模型。

数据模型 主要包括网状模型、层次模型、关系模型等，它是按计算机系统的观点对数据建模。

--------------------------------------------------------------------------------------------------------------------------------

概念模型相関概念

概念模型:用于信息世界的建模，是对现实世界的抽象和概括。

属性：实体所具有的某一特征。

码（key）：能在一个实体集中唯一表示一个实体的属性。

域（Domain）：某个（些）属性的取指范围。

实体型（Entity Type）：具有相同属性的实体具有共同的特征和性质。

实体集：同类型的实体集合。

**联系**：

两个实体之间的联系有三种：一对一联系、一对多联系、多对多联系。

**数据模型**的**组成要素：**数据结构、数据操作、数据的完整性约束条件**。**

数据结构是对系统静态特性的描述。

数据操作：对数据库中各种对象的实例允许执行的操作及有关的操作规则。

操作的类型---检索和更新（包括插入、删除、修改）。

数据操作是对系统动态特性的描述。

数据的完整性约束条件---- 一组完整性规则的集合。

数据库设计中广泛使用的**概念模型**当属**E-R数据模型**

1. R数据模型（Entity-Relationship data model），即实体联系数据模型，是P.P.S. Chen于1976年提出的一种语义数据模型。
2. R模型 三要素： 实体、联系、属性。

客观存在并可相互区别的事物称为实体。

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型内部的联系和实体型之间的联系。

联系

两个实体型间 一对一联系（1:1）

三个或三个以上实体型间 一对多联系（1:n）

同一个实体型内 多对多联系（m:n）

常用的数据模型： **层次模型、网状模型**、关系模型、面向对象模型。

非关系模型 关系模型

关系模型的术语：

关系 ---- 一个关系对应通常所说的一张二维表。

元组 ---- 表中的一行称为一个元组。

属性 ---- 表中的一列即为一个属性，给每一个属性起一个名称即属性名。

主码 ---- 表中的某个属性或属性组，它们的值可以唯一地确定一个元组。

域 ---- 属性的取值范围称为域。

分量 ---- 元组中的一个属性值，即行和列的交叉。

关系模式 ---- 关系的型称为关系模式，关系模式是对关系的描述。

关系模式的一般表示是：

关系名（属性1，属性2，..., 属性n）

联系 ---- 在关系模型中，实体以及实体之间的联系都是用关系来表示。

实体及实体间的**联系**的**表示方法**

实体型：直接用关系（表）表示。

属性：用属性名表示。

联系：用键表示。

在关系模型中，记录之间的联系通过**键**来实现

关系模型的操作语言是SQL。

关系必须是规范化的，满足一定的规范条件

最基本的规范条件：关系的**每一个分量必须是一个不可分的数据项**,

不允许表中还有表。

关系模型的组成： 关系数据结构、关系操作集合、关系完整性约束

**关系模型的优点：**

1）建立在严格的数学概念的基础上

2）概念单一

实体和各类联系都用关系来表示

对数据的检索结果也是关系

3）关系模型的存取路径对用户隐蔽

具有更高的数据独立性，更好的安全保密性

简化了程序员的工作和数据库开发建立的工作

**缺点**

1）查询效率往往不如非关系数据模型

2）为提高性能，必须对用户的查询请求进行优化

增加了开发DBMS的难度

**关系具有以下性质**：（**规范化关系**）

1. 列是同质的，即每一列中的分量是同一类型的数据，来自同一个域。
2. 不同的列可以出自同一个域，每一列称为一个属性。在同一关系中，属性名不能相同。
3. 列的顺序无关紧要，即列的顺序也可以任意转换。
4. 任意两个元组（行或记录）不能完全相同。
5. 行的顺序也无关紧要，即行的顺序也可以任意转换。
6. 行列的交集称为分量，每个分量的取值必须是原子值，即分量不能再分。

**2.4.3 关系代数**

关系代数 是一种抽象的查询语句，是关系数据操作语言的一种传统表达方式，它是对关系的运算来表达查询的。

关系代数的运算按运算符性质的不同分为两种情况：

1. 传统的集合运算 2）专门的关系运算

第三章数据库设计

**关系数据库的规范化理论**主要包括三个方面的内容：

**函数信赖（依赖）**

**范式（Normal Form）**

**模式设计**

其中，函数信赖起着核心的作用，是模式分解和模式设计的基础，范式是模式分解的标准。

**关系模式设计不当造成的问题有哪些**？

1. 数据冗余 2）插入异常 3）删除异常 4）更新异常

3.1 范式种类