dbADS 2.0

数据库设计与自动生成系统

王　昀　黄立农
王晓敏　敬　喜　编著

北京科海电子出版社

需要全本请访问 www.ertongbook.com

内 容 简 介

无论您是信息系统技术及相关专业的研究生、本科生还是专科生，抑或信息系统领域的开发人员，本书都将使您如获至宝。

全书共分上、下两篇。上篇主要介绍数据库的基本知识和原理，并给出了一个数据库设计样例；下篇则对 dbADS 2.0 数据库自动生成系统从最初的设计到实现再到系统的运用一一作了详解，并着重突出数据库自动设计的原理和算法。

光盘内附作者自行研发的 dbADS 2.0 数据库自动生成系统软件，与本书相辅，为您轻松愉快地实现一个最优化的数据库奉献了一套完整的开发工具。

dbADS 2.0 数据库设计与自动生成系统

封面设计：林陶
产品负责人：王华红

*

出　　品：北京科海电子出版社
社　　址：北京市海淀区上地七街国际创业园 2 号楼 14 层
邮　　编：100085
发　　行：北京科海电子出版社　www.khp.com.cn
销售电话：（010）82896443　82896438
客户服务：（010）82896445/46 转 8407
经　　销：各地新华书店　音像部门

*

印　　数：1－2000
出品日期：2005 年 9 月
ISBN：7-900420-76-2

定　　价：68.00 元（1CD+1 配书）
（版权所有　翻印必究　印装有误　负责调换）

无论您是信息系统技术及相关专业的研究生、本科生还是专科生，抑或信息系统领域的开发人员，本书都将使您如获至宝。

全书共分上、下两篇。上篇主要介绍数据库的基本知识和原理，并给出了一个数据库设计样例；下篇则对 dbADS 2.0 数据库自动生成系统从最初的设计到实现再到系统的运用一一作了详解，并着重突出数据库自动设计的原理和算法。

光盘内附作者自行研发的 dbADS 2.0 数据库自动生成系统软件，与本书相辅，为您轻松愉快地实现一个最优化的数据库奉献了一套完整的开发工具。

目　　录

上篇　数据库基本原理

第 1 章　绪论 1

1.1　三层次数据模型 1
1.2　数据库模型 3
1.3　数据库类型 4
1.4　数据库设计目标 5
1.5　数据库设计过程 6
1.6　数据库逻辑结构设计的实施过程 6
1.6.1　工作活动性质鉴别阶段 7
1.6.2　定义阶段 8
1.6.3　概念化阶段 10
1.6.4　计算机效率优化阶段 11
1.7　关于数据库设计自动化 12

第 2 章　关系模型 13

2.1　引言 13
2.2　关系模型的基本结构 13
2.2.1　关系 13
2.2.2　关系的属性及属性值 15
2.2.3　关系的特性 15
2.3　函数依赖 16
2.4　完全函数依赖 18
2.5　多值依赖 19
2.6　复合函数与传递依赖 19
2.7　关系的规范化 22
2.7.1　非规范化关系和 1NF 关系 23
2.7.2　1NF 关系和 4NF 关系 26
2.7.3　最佳 4NF 关系 28

第 3 章 关系的设计 31
3.1 引言 31
3.2 确定不可分解的基本单元 32
3.2.1 不可分解性的标准 32
3.2.2 关系的分解 33
3.2.3 普遍性分解准则 35
3.2.4 分解关系的步骤 39
第 4 章 数据库的实体-联系模型 41
4.1 引言 41
4.2 实体-联系模型（E-R 模型） 41
4.2.1 实体、联系和属性 41
4.2.2 标识符 42
4.2.3 E-R 模型的图形表示 42
4.2.4 实体和属性的抉择 44
4.2.5 实体、联系及关系 45
4.2.6 二元联系及 n 元联系 45
4.2.7 递归联系 48
4.3 设计 E-R 模型的步骤 48
4.4 模型转化为关系的准则 54
4.5 作用模型和类型模型 55
4.5.1 作用及其结构 56
4.5.2 类型模型及其结构 64
4.6 聚合模型 72
第 5 章 概念体设计 76
5.1 概念体 76
5.1.1 实体集合 76
5.1.2 关联集合 76
5.1.3 定义域 76
5.1.4 实体属性和关联属性 77
5.2 实体键与概念体的表示 77
5.2.1 实体键 77
5.2.2 实体集合的表示方法 78
5.2.3 关联集合的表示方法 79
5.3 概念体与关系模型 80

需要全本请访问 www.ertongbook.com

5.3.1 实体集合用关系表示 80
5.3.2 用关系表示属性间关联的关联集合 81
5.3.3 用关系表示在实体集合之间关联的关联集合 81
5.4 确定概念体 81

第 6 章 数据库说明 83

6.1 数据库说明的组成 83
6.2 关系模型转化为逻辑记录结构 83
6.3 存取需求说明 89
6.4 量化数据说明 90

第 7 章 数据库设计过程 92

7.1 引言 92
7.2 数据库概念结构设计实施的各个阶段 94
7.2.1 工作活动性质鉴别阶段 94
7.2.2 定义阶段 94
7.2.3 概念化阶段 97
7.2.4 计算机效率优化阶段 101
7.3 数据库设计中的一般步骤和方法 101
7.4 数据库设计过程例子 115
7.4.1 工作性质鉴别 115
7.4.2 定义概念体 117
7.4.3 建立 E-R 模型 119
7.4.4 E-R 模型转化为关系模型 121
7.4.5 构造 SRS 数据库的总体逻辑模型 123
7.4.6 数据库说明 128
7.4.7 物理实现 130

第 8 章 数据库设计自动化 134

8.1 引言 134
8.2 设计的自动化原理 134
8.2.1 有向图与基本关系 134
8.2.2 基本关系的距离 139
8.2.3 建立复合关系 A^p 矩阵及其性质 139
8.2.4 建立基本关系的距离矩阵 D 及其性质 140
8.2.5 复合关系的距离及其权数 141

8.2.6 最小覆盖 142
8.2.7 最佳 4NF 的总体逻辑模型 142
8.3 设计的自动化算法 142
8.3.1 确定传递闭包的算法 dbADS-TC 143
8.3.2 确定传递闭包程序 dbADS-TC 执行过程 143
8.3.3 确定基本关系的距离的算法 dbADS-DT 145
8.3.4 复合关系及其权数和距离 147
8.3.5 确定最小覆盖算法 dbADS-OV 149
8.3.6 确定最小覆盖的一般步骤 149
8.3.7 删除关系的原则 150
8.3.8 删除算法 151
8.3.9 设计总体逻辑结构算法 dbADS-LM 157
8.4 自动化算法应用举例 158
8.4.1 dbADS 系统 158
8.4.2 例子 158
第 9 章 数据库设计样本 166
9.1 课题与设计指导 166
9.1.1 课题 166
9.1.2 设计指导 166
9.2 阶段 1：选择某个系作为考察对象 170
9.3 阶段 2：建立 E-R 模型 173
9.3.1 建立 E-R 模型 173
9.3.2 标明属性 176
9.3.3 E-R 模型转化为关系模型 178
9.4 阶段 3：总体逻辑结构设计 180
9.4.1 分解关系：确定 TMDB 的不可分解单元 180
9.4.2 总体逻辑结构 188
9.5 编写数据库说明 190
9.5.1 组织模型 190
9.5.2 关系模型转化为逻辑记录结构 190
9.5.3 存取需求说明 192
9.5.4 量化数据说明 192
9.6 阶段 4：物理设计（Ⅰ）——构造表 196
9.6.1 计表结构 196
9.6.2 用 DBD 构造表 199

9.7 阶段 4：物理设计（II）——链接表 200
9.7.1 主索引和次级索引 201
9.7.2 建立 STUDENT.db 表和 RESALTS.db 表的链接 201
9.7.3 建立 COURSE.db 表和 RESALTS.db 表的链接 207
9.7.4 给表录入数据 215
9.7.5 运行程序 218
9.8 EDS 数据库设计课题 222
9.8.1 阶段 1：概念化阶段 222
9.8.2 阶段 2：分析与分解关系 226
9.8.3 阶段 3：求传递闭包 228
9.8.4 阶段 4：求最小覆盖 230
9.8.5 阶段 5：求总体逻辑结构 230
9.9 结束语 231

下篇　数据库自动生成系统 dbADS

第 10 章　数据库自动生成系统 dbADS 操作说明书 232
10.1 系统概述 232
10.1.1 dbADS 系统的组成 232
10.1.2 “dbADS 系统”的运行环境 232
10.1.3 dbADS 系统安装 233
10.1.4 启动 dbADS 系统 236
10.2 dbADS 系统菜单 236
10.2.1 文件 237
10.2.2 数据准备 237
10.2.3 执行算法 240
10.2.4 单步执行 241
10.2.5 保存结果 243
10.2.6 帮助 244
10.3 操作示例 244
10.3.1 原始数据 244
10.3.2 输入属性 245
10.3.3 构造关系和分解关系 245
10.3.4 确定传递闭包、关系距离和最小覆盖 248

10.3.5 总体逻辑结构 250
第 11 章 数据库设计自动化的研究报告 253
11.1 前言 253
11.2 原理 253
11.2.1 有向图与基本关系 254
11.2.2 基本关系的距离 258
11.2.3 建立 A^p 矩阵及其性质 261
11.2.4 建立基本关系的距离矩阵 D 及其性质 263
11.2.5 复合关系的距离及其权数 265
11.2.6 最小覆盖 267
11.2.7 关系分解准则 268
11.3 算法 271
11.3.1 关系分解算法 dbADS-RD 271
11.3.2 传递闭包 271
11.3.3 关系的距离 274
11.3.4 最小覆盖 277
11.3.5 总体逻辑结构 288
11.4 dbADS 系统及其测试 289
11.4.1 dbADS 系统 289
11.4.2 测试 290
第 12 章 关系数据库自动生成系统（dbADS）设计详解书 294
12.1 课题与设计指导 294
12.1.1 课题 294
12.1.2 设计指导 294
12.2 演示数据库自动生成的方法和过程 298
12.2.1 阶段 1：真实世界的概念化阶段 299
12.2.2 阶段 2：确定 TMDB 的不可分解单元（即基本关系） 306
12.2.3 阶段 3：确定传递闭包并消去没有意义的复合 314
12.2.4 阶段 4：确定关系的距离和权数 327
12.2.5 阶段 5：确定最小覆盖 339
12.2.6 阶段 6：总体逻辑结构设计 347
参考文献 353

前　言

自从 1970 年 Codd 等人提出关系设计理论以来，关系型数据库一直被公认为是最有效率的组织数据方式。关系数据库模型是针对解决有效组织数据问题而提出的一套以数学理论为基础的数据模型。目前已经出现了多种支持关系模型的数据库管理系统（RDBMS），例如 dBASE、FoxPro、SyBase、InterBase、Paradox、Informix、ORACLE 等。但是，在关系数据库（或关系模型）设计中存在的一个问题始终没能得到很好解决，这就是在关系数据库设计中，为了得到一个最佳设计需要处理数以百计，乃至数以千计的关系。这样一来，像以往那样用手工操作就会有许多困难，甚至无法实现一种最佳设计。而如果一个数据库不是最佳的，那么这个数据库就会存在很大问题，例如存在存储操作异常的问题。

由于手工设计操作有困难，必然要寻求借助于计算机使其自动完成手工所不能完成的任务。为此，本书作者经过研究，开发了一种 Windows 操作系统下的数据库自动生成工具，即 dbADS（Database Automatic Design System）系统。该系统在 Windows 窗口环境下以菜单操作方式，通过一系列算法实现数据库的自动生成，使数据库设计真正地实现了自动化和最佳化。该系统目前的版本为 2.0，系统载体为本书所带 CD-ROM。

本书把注意力集中于关系型数据库的设计过程和具体设计方法上，旨在帮助读者快速掌握开发数据库的方法和技巧。本书在讲解上采用循序渐进、理论联系实际的方式，对涉及到的重要概念尽可能配合示例说明，并注意讲解在学习中可能出现的问题。全书的可操作性和实用性良好，详细的步骤演示，实用的操作技巧，图文并茂的解释，易于理解的概念阐述，通过一个完整的实例——数据库设计样本，介绍开发数据库的全过程，这些都会给你的学习和开发数据库工作带来方便。本书另一个特点是，立足于数据库的基础理论和设计方法，着眼于最佳化和自动化两个方面，详细介绍了数据库自动化设计的理论和算法、dbADS 系统的开发方法及其应用。读者将会发现，本书其实正在实现轻松愉快地学习开发一个最佳数据库的一种思想。

无论您是信息系统技术及相关专业的研究生、本科生还是专科生，抑或信息系统领域的开发人员，本书都将使您如获至宝。

版权声明

光盘中所载系统为数据库自动生成系统 V2.0，**简称** dbADS 2.0 **系统，英文名为** dbADS V2.0 for Windows。**本系统及其配套书，禁止复制。版权所有，侵权必究。**

本系统开发者联系方式为：jzyy0961@vip.sina.com。

编　者
2005 年 6 月

8.3.9　**设计总体逻辑结构算法** dbADS-LM 157
8.4　**自动化算法应用举例** 158
8.4.1　dbADS **系统** 158
8.4.2　**例子** 158

8.3.9　设计总体逻辑结构算法 dbADS-LM

算法 dbADS-LM

需要全本请访问 www.ertongbook.com

上篇　数据库基本原理

第 1 章　绪　　论

在一个单位或部门的工作中，经常需要使用数据或用数据进行辅助决策，所以数据是该单位或部门工作过程的支柱。数据库就是组织数据的仓库，所以在一个单位或部门的工作中，数据库发挥着核心作用。用户在决策过程中，通过访问数据库获得数据，并可以将决策结果保存在数据库中。数据库的布局和存取功能对工作效率影响极大，如果用户在所有地点都能很容易地存取各种数据，就能快速响应决策人的要求，否则就会影响及时地作出决策。另外，如果获取数据很难，或者获取的数据不完整，都会给决策带来负面影响，甚至不能正常工作。数据库必须具有灵活的结构，这样，即使对数据的需求发生变化或者有新的任务要求进行新的决策时，数据库也能存取新的数据以及数据间新的联系，以支持对新的数据或新的决策的需求。

本书的主题是：为了满足信息的需求如何科学地设计出最佳的数据库。为此，讨论了数据库设计方法的一系列问题，诸如数据模型、关系的范式及最佳四范式的实现方法、数据库设计方法(步骤和过程)、数据库自动设计原理和算法,以及数据库自动生成系统 dbADS 的设计方法与应用等。本章就涉及到的数据库设计方法做一个综述，以帮助读者理解和学习。

1.1　三层次数据模型

数据库技术是计算机应用、信息处理、经济、工商、经营管理等专业的一门重要的课程，它是一门综合性的软件技术，是使用计算机进行各种信息管理的必备知识。

设计一个数据库，概括起来说有以下要求：

(1) 要为多种多样的应用项目服务；

(2) 要能很容易增加新的数据类型和新的应用项目；

（3）要做到随着用途和使用方式的演变，存在磁盘上的数据结构也要能加以改变，而又不影响应用程序；

（4）要选择支持数据库在某个计算机上运行的数据库管理系统（Database Management System，DBMS）。

一般来说，数据库是庞大而复杂的实体，且与环境有着密切的关系。设计时，就是要把这些原则和具体情况有机地结合起来。在某种意义上来说，分析与设计的目的就是鉴别出对于一个部门来说是长期不变的特点及根据上面 4 条要求所得到的结论（或原则）来设计数据库。这样的要求和原则对所有工作部门都是适用的。

1975 年美国国家标准协（学）会 ANSI/SPARC 提出三层次数据库管理系统结构的建议报告。这种三个层次组成的 DBMS 标准结构是：

（1）表征用户视图的外部层次；
（2）表征组织数据整体的概念层次；
（3）表征数据库总体逻辑结构的内部层次（或称存储层次）。

通常是：

- 外部层次由几个外部数据模型来代表，外部数据模型是用户看得到的简化模型，因此又称之为用户视图，外部数据模型总可以从内部数据模型推导出来，但不一定是内部模型的子集；
- 概念层次由一个概念数据模型来代表；
- 内部层次由一个描述已存数据结构的内部数据模型来代表。

事实上，现有的 DBMS 很少是支持概念数据模型的。尽管如此，一个独立于 DBMS 的概念数据模型也是很有用的。因为它有下列用途：

- 可以在一个适当的抽象程度上建立一个概念数据模型，然后转换成为该 DBMS 支持的数据模型，即内部数据模型，又称之为总体逻辑模型。内部层次是用来存储概念层次客体的记录结构的，这个层次也包括索引、链表或进行记录存取的其他物理路径。
- 它可以作为编写和理解一个单位或部门的数据的书面资料。
- 可用来定义使用现有 DBMS 来实现的数据库。
- 为控制数据使用提供依据。

图 1.1 说明了这种三层次数据模型。在做这个数据库时是从概念数据模型设计开始的。概念数据模型是推导出来的 4 个基本关系 R1(E#,E_name),R2(E#,D#),R3(M3,M_name)和 R4(M#,D#)。如果现有的 DBMS 是关系型数据库管理系统，那么就可以将这 4 个关系直接

作为内部数据模型，不过这不是最佳的，因为要存储 4 个关系。但经过最佳化处理后可以用两个基本关系 ER1(E#,E_name,D#)和 ER2(M#,M_name,D#)来取代。外部数据模型则是根据用户的需求经过运算得到的，如图 1.1 所示的用户视图是通过如下连接和投影运算而得到：

```
ER(E#,M#,D#)←ER1 * ER2[E#,M#,D#]
```

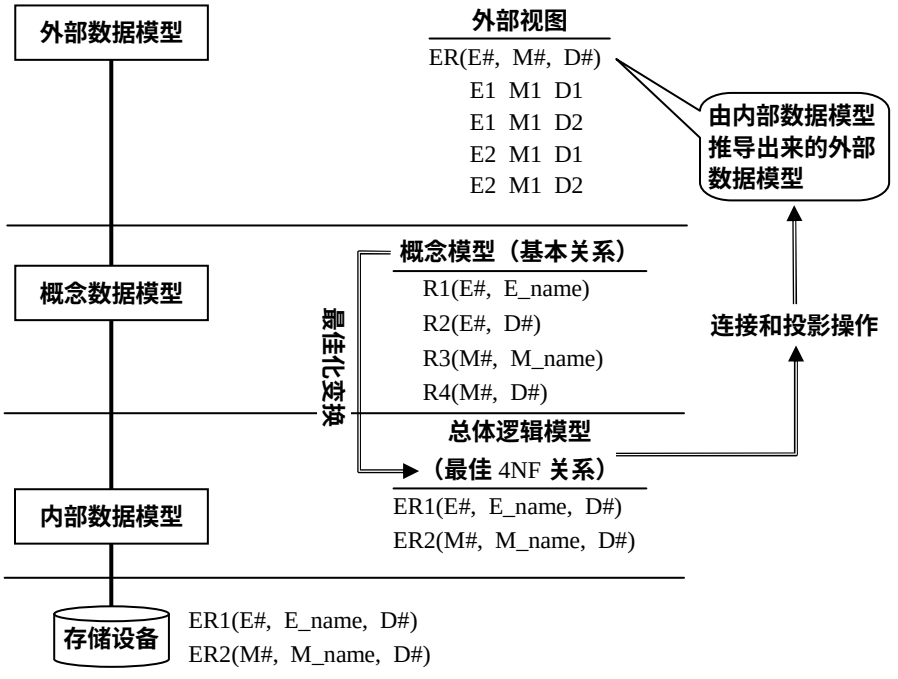


图 1.1　数据模型的三层次结构及关系型数据库的示例

1.2　数据库模型

实体-联系模型（Entity-Relationship Model，简称 E-R 模型）是 P.Pchen 于 1976 年提出的，其在数据库设计中已得到广泛采用。E-R 模型为设计人员提供了三个主要语义概念，即实体 、联系和属性，并使用这三个概念描述一个部门或组织。实体是描述部门或组织中独立的客体，联系是客体间的有意义的相互作用，属性是对实体和联系的描述，每个属性值都关联于一个值集，属性的值取自这个值集。联系也可以看作为抽象的客体。联系用连线连接到参与该联系的客体，并用不同名字来标记不同的连线，由此，引入 E-R 模型的图形表示方法。建立数据库的 E-R 模型的过程是：首先确定实体集，然后确定联系集，最后

给出实体集和联系集的属性。但是，由于设计人员对该部门或组织中客体重要程度的理解不同，不同的设计人员可能设计出不同的 E-R 模型。因此，只有实体和联系这两种抽象还不足以刻画一个部门或组织的某些特性，还需要引入新的抽象，即作用模型和类型模型，作为 E-R 模型的补充，以对 E-R 模型不能表达的联系进行模型化。本书将作用模型和类型模型看作是 E-R 模型的组成部分。有关 E-R 模型、作用模型及类型模型的具体设计的问题，将在第 4 章中详细讨论。

当具体实现数据模型时，就要涉及到现有的 DBMS 所支持的数据库类型是关系型，还是网状型，抑或是层次型。不同的数据库类型，E-R 模型的转换方法也不同。本书只就关系型数据库进行讨论，并且将关系模型设计实现了自动化[①]。

1.3 数据库类型

大多数的商用 DBMS 只支持一种数据模型，不同的 DBMS 支持的数据模型可能各不相同。数据库类型通常分为 3 类：

- 关系型数据库
- 网状型数据库
- 层次型数据库

本书讨论关系型数据库。关系型数据库中，数据被组织成表，提供给用户的是表格界面，由于表格已为人们所熟悉，所以用表格构造组织模型是被广泛接受的。而且关系型数据库一直被公认为是最有效率的数据组织方式。关系型数据库的模型是针对解决有效组织数据问题而提出的一套以数学理论为基础的数据模型。依据关系理论作为数据库的一种设计理论，不只是适用于关系型数据库，而且也适用于网状型数据库和层次型数据库。

对同一个数据库，DBMS 必须能够保持不同用户视图的相对独立性、保持每个用户视图与数据库的物理存储结构的独立性，通常用数据独立性来描述 DBMS 的这一特性。为此，数据模型必须独立于诸如索引、排序方式及物理存储路径之类和计算机有关的结构。也就是说，根据数据模型由 DBMS 提供用户界面，该界面着重于用户问题的逻辑结构，而独立于计算机的物理结构。因此，数据库设计应保证数据的物理独立性，即逻辑结构和物理结构之间的独立性。数据的物理独立性的重要特征是，物理结构的改变不致影响逻辑结构。物理结构的改变可能是由于使用方式的变化而引起的。由于程序只与逻辑结构有关，所以物理结构的任何变化不会影响用户程序，即程序独立于物理结构，从而使程序易于维护。

① 本书不讨论向网状型和层次型数据模型的转换方法及其自动化算法，有兴趣的读者可以和作者联系。

数据的物理独立性对数据库设计有着重要影响，它将数据库设计过程划分为两个阶段：即逻辑设计和物理设计。如图 1.2 所示。在逻辑设计中，用户的组织（或部门）模型转化为数据库的逻辑结构，转化过程不必考虑最终的数据物理结构的细节。物理结构是在物理设计中选择，选择后再加入到在逻辑设计阶段选定的数据库逻辑结构中去。

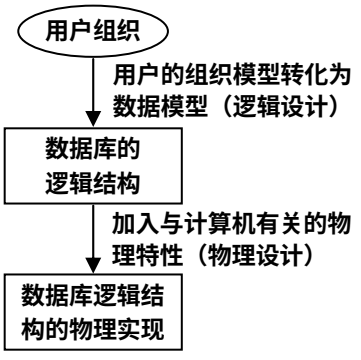


图 1.2　数据库的逻辑设计和物理设计

在关系型数据库的三层次结构中，概念数据模型是基本关系的集合，内部数据模型是最佳四范式（4NF）关系的集合，外部数据模型是由内部数据模型的重新编排而得到。如图 1.1 所示的那样，有为表示概念数据模型所需要的基本关系，以便于在物理上实现该概念数据模型的内部总体逻辑模型，以及从内部数据模型经过数据重新编排而得到的外部数据模型或外部视图。

总之，关系型数据库的实质是：

（1）外部模型是用各种关系来定义的。这样，每一个应用程序都是用诸如投影、连接、选取等运算对各种关系进行操作的。因此，用户是根据关系以及为得到新的关系而对关系进行的操作来考虑问题的。诸如投影 、连接 、选取等运算把关系作为操作数，并把产生新的关系作为运算的结果。这些运算叫做关系代数。实现这些运算的子过程都是一些简单的算法。用户只要规定所需要的关系是什么，而不用规定应该如何进行计算才能够得到运算的结果。

（2）内部的总体逻辑模型是用关系来表示的。

（3）物理模型的实现要使所要求的操作能够高效地进行。

1.4　数据库设计目标

数据库设计必须满足若干准则。这些准则分为两大类：结构准则和性能准则。

结构准则主要考虑的是如何保持数据的特性，主要有：

- 存储准则：保持数据的关系范式，避免数据的二义性和数据的重复。
- 连接准则：保持客体集之间连接的完整性，避免数据库的不一致。

性能准则主要考虑资源的合理使用和数据库存取的效率，主要要求有：

- 必须满足事物对响应时间的要求。
- 所用的存储量最少。
- 内外存之间传递的次数最少。

为了达到上述的数据库设计目标，通常有两种设计方法可供选择：基于数据的方法和基于存取的方法。这两种方法的区别在于是先考虑结构准则还是先考虑性能准则。

基于数据的方法强调的是结构准则。首先将组织或部门模型转化为数据模型，这一步不考虑存储需求，仅考虑结构准则。然后在逻辑结构上标出逻辑存取路径。

基于存取的方法强调的是性能准则。首先结合存取路径考虑逻辑记录结构，根据这些路径选择数据模型结构。

本书遵循的是基于数据的方法，强调的是结构准则。

1.5 数据库设计过程

数据库是庞大而复杂的实体集合，且与环境有着密切的关系。数据库设计要按照一定的过程来进行，整个过程包括以下 6 个方面：

（1）设计逻辑结构，设计出概念数据模型和内部数据模型。
（2）设计物理结构，即数据库的物理组织方法。
（3）要对设计做出评价和性能预测。
（4）物理实现。
（5）试运行。
（6）投入使用。

1.6 数据库逻辑结构设计的实施过程

一个好的数据库逻辑结构设计实施过程，应划分为以下几个阶段：

（1）工作活动性质鉴别阶段；

需要全本请访问 www.ertongbook.com