

全国高等院校“十五”计算机规划教材
计算机科学与技术教材系列 (6)

现代 数据库系统 教程

徐洁磐 编 著

11.13
02



北京希望电子出版社
Beijing Hope Electronic Press
www.bhep.com.cn

321

TP311.13
X7462

全国高等院校“十五”计算机规划教材

计算机科学与技术教材系列（6）

现代数据库系统教程

徐洁磐 编著



北京希望电子出版社

Beijing Hope Electronic Press

www.bhp.com.cn

内 容 简 介

这是一部关于现代数据库系统的基本原理与技术的教科书,其中包括网络数据库、Web 数据库、数据仓库以及安全数据库等最新技术内容。

本书由 3 部分 16 章构成,第 1 章是数据库的基本概念。第 2 章介绍数据模型。第 3 章介绍关系数据库系统。第 4 章介绍关系数据库系统标准语言 SQL。第 5 章讲解数据库的物理组织。第 6 章讲解关系数据库的查询优化。第 7 章介绍数据库的安全性 with 完整性。第 8 章讲述关系数据库规范化理论。第 9 章讨论数据库设计与数据库管理。第 10 章介绍事务处理、并发控制与故障恢复技术。第 11、12 章讲解面向对象数据库系统和知识库系统。第 13 章介绍分布式数据库系统。第 14 章讨论网络与数据库。第 15 章阐述决策支持系统与数据仓库。第 16 章讨论移动式数据库与嵌入式数据库。此外,本书各章均配有丰富的习题供读者练习、巩固所学知识。

作者从事一线的教学数十年,积累了丰富的教学经验,本书在整体结构安排、内容取舍以及整书的编写过程中,都充分考虑了教与学的特点,以及所面对的特定读者的具体需要。在内容上既注重讲述基础知识,又能反应当代数据库系统的新技术。本书结构清晰,概念准确,文字叙述简洁明了,可读性强,既便于教师课堂讲授,又便于自学者阅读。

本书可作为普通高校、职业学校计算机科学与技术专业本科学生的教材和教学参考书,部分内容还可作为研究生教材,同时也可以作为广大数据库应用开发人员的开发指导书。

系 列 书 名 : 全国高等院校“十五”计算机规划教材

计算机科学与技术教材系列(6)

书 名 : 现代数据库系统教程

总 策 划 : 北京希望电子出版社

文本著作者 : 徐洁磐 编著

责 任 编 辑 : 蒋湘群

出版、发行者 : 北京希望电子出版社

地 址 : 北京市海淀区知春路 63 号卫星大厦三层 100080

网址: www.bhp.com.cn

E-mail: lxr@bhp.com.cn

电 话 : 010-62520290,62521724,62528991,62630301,62524940,62521921,82610344

(发行) 010-82675588-202 (门市) 010-82675588-501,82675588-201 (编辑部)

经 销 : 各地新华书店、软件连锁店

排 版 : 希望图书输出中心 张月岭

文本印刷者 : 列电印刷厂印刷

开本 / 规格 : 787 毫米×1092 毫米 16 开本 17.25 印张 385 千字

版次 / 印次 : 2002 年 8 月第 1 版 2002 年 8 月第 1 次印刷

印 数 : 0001-5000 册

本 版 号 : ISBN 7-900101-58 -6

定 价 : 20.00 元

说明: 凡我社产品如有残缺,可执相关凭证与本社调换。

序

近年来计算机科学与技术的发展突飞猛进,其应用范围之广,对国民经济影响之大前所未有,特别是计算机网络、电子商务、多媒体技术等发展,正在彻底改变人类的工作方式与生活方式,同时也彻底改变了传统产业与传统的工作模式。目前,计算机科学与技术是高新技术的主要标志,是先进生产力的重要支柱,因此,发展计算机事业是摆在我们面前的重要任务。有鉴于此,我们特组织编辑了以大学本科学生为对象的《计算机科学与技术教材》丛书,为我国信息化培养人才作一份贡献。

当前,在我国计算机教学及教材建设中普遍存在着一些弊病与不足,主要有如下几种:

1. 在计算机教材特别是基础性教材中严重存在知识陈旧、落后,跟不上计算机科学与技术发展的步伐。由于计算机技术的飞速发展,内容更新要求极快,一般三五年就需作重大调整而一二年需作必要调整,而现有教材大都不能适应此种变化速度,这种现象在基础性教材中尤为突出。

2. 现有教材很多以国外教材为蓝本,存在着脱离我国具体应用实际的弊病。如何根据我国国情并参考国外先进技术编写教材是当务之急。

3. 现有的教材大都适应面窄,多数仅适应计算机本科专业而尤其仅适应少数重点院校与重要院校。而目前,由于计算机教育发展迅速,各种与计算机相近与相关的专业蓬勃发展,在计算机本科专业中也出现了不同层次的要求,特别是以应用型人才为主的培养要求。因此,迫切需要有一套适应面较广的基础性教材以满足多种层次的要求。

根据以上分析,本丛书编写原则是

1. 适应计算机科学与技术飞跃发展的需要,本教材丛书具有先进性与时代特性,并且每隔一二年作一次小的调整,每隔四五年重新修订出版。

2. 本教材丛书具有适应面广,基础性强的特点,能满足多种层次、多种类型的计算机专业本科学生的需要,特别是满足计算机应用型人才培养的需要。

3. 本教材丛书具有密切结合我国应用实际、反映我国计算机事业发展需求的基本特点,并且能在实际应用中发挥作用。

4. 教材是有别于一般书籍的一种特殊读物,它要求基本概念清楚,基本理论扎实,知识量大与实际应用联系紧密等特点,它还要求教材的内容逻辑性强,可读性好,深入浅出,并附有习题与参考资料等内容,本教材丛书将突出体现这些特点使其适合于教材需要。

5. 本教材丛书选题是根据我国目前实际并参考国际最新动态而制订的。本教材丛书第一批 8 本都是具有普遍性与基础性的教材,在不久我们将分别推出第二批、第三批教材满足不同的需要。

6. 本教材丛书聘请有深厚理论基础与应用实践经验且长期在教学、科研第一线工作的高校教授编写,特别是有专长的中、青年教授是本丛书教材编写的主要力量。

我们感谢丛书编委会各位委员为本丛书出版所作出的贡献,我们也感谢北京希望电子出版社为丛书的立项、编审所作出的努力,最后我们感谢丛书的各位作者为丛书编写、发

行与发展所作出的创造性的和富有成效的工作。

我们还期待广大读者为丛书提出宝贵意见与建议，我们将通过修订，努力把本丛书办得更好。

计算机科学与技术教材编委会

2002年3月于南京大学

前 言

数据库技术在我国日渐普及,数据库应用也日益广泛,据不完全统计,它在计算机应用中占有的百分比已超过 60%。数据库不仅在传统的商业领域、事务处理领域中发挥着重要作用,而且在非传统应用中也起到越来越大的作用。例如,在工程领域、数据集成领域、统计领域中以及在图形、图像、声音等多媒体领域中都起着重要的数据支撑作用。此外,数据库与网络的结合已成为目前发展的趋势。

随着应用的需要以及技术的成熟,数据库本身也发生着显著的变化。关系数据库技术已成为当前数据库发展的主流,层次与网状数据库已退出舞台,成为历史产物。与此同时,面向对象数据库以及知识库等新型数据库系统正在兴起,与网络相关的技术正在蓬勃发展,如分布式数据库、Web 数据库、数据仓库以及数据库安全等新技术正在成为主流技术。

鉴于以上情况,有必要编写一本与数据库新技术有关的图书以适应数据库应用的发展,适应面向 21 世纪的计算机应用的需要。本书正是为此目的而写,它包括数据库系统的最新内容,主要有如下几点:

- (1) 突出以关系数据库技术为主要内容,删减与压缩层次与网状数据库内容。
- (2) 增加面向对象数据库以及知识库系统内容。
- (3) 增加分布式数据库、网络与数据库、数据仓库及数据库安全等与网络有关内容。
- (4) 增加数据库应用方面内容。

以上述几点为基础,再通过结构上的调整从而构成一本能适应最新数据库技术发展的教科书。

本书可作为高等院校计算机相关专业的大学本科及专科的数据库课程教材,也适用于数据库应用开发人员参考,部分内容还可作为研究生教材之用。本书重点在于阐述数据库系统的基本原理与基本技术,同时介绍最新技术。

从内容组织上看,本书由 3 部分构成:

(1) 数据库的基本原理 它由第 1 章和第 2 章组成。主要讲述数据库的最基本概念与数据库模型,它是数据库的基础部分。

(2) 关系数据库系统的原理与技术 它由第 3 章至第 10 章组成。主要介绍关系数据库系统基本理论、原理、技术以及应用,是本书核心部分。

(3) 数据库新技术 由第 11 章至第 16 章组成。主要讨论几种新的数据库系统、数据库网络应用以及数据库微小化的应用等。

在本教材使用方面也可分为 3 个层次:

(1) 本科生层次 数据库基本原理部分与关系数据库系统部分是必须讲授的,而新技术部分则可选取一部分讲授。

(2) 研究生层次 数据库基本原理部分与新技术部分是必须讲授的,而关系数据库系统部分则可选取一部分讲授。

(3) 大专生层次 数据库基本原理部分是必须讲授的,而关系数据库系统部分则可选取其主要部分讲授,新技术部分则可选取少量部分讲授。

本书内容深入浅出，文字浅显易懂，组织结构严密，这是本书的一大特色。
值本书付梓之时，作者感谢山东大学董继润教授对本书的审阅以及他所付出的劳动。
本书在编写过程中得到陈巧珍老师的帮助，在此特表示感谢。
由于作者水平有限，书中难免有缺点和错误，殷切希望读者批评指正。

作者

2002年2月于 南京

计算机科学与技术教材系列

编委会成员名单

主任委员：徐洁磐

副主任委员：秦小麟 宋方敏

委员：（按姓氏笔画排序）

王元元 教授 解放军理工大学

宋方敏 教授 南京大学

张桂芸 副教授 天津师范大学

张新荣 教授 天津大学

柏家求 教授 天津大学

秦小麟 教授 南京航空航天大学

徐永森 教授 南京大学

徐洁磐 教授 南京大学

殷新春 副教授 扬州大学

蔡庆生 教授 中国科学技术大学

目 录

第 1 章 概论	1
1.1 基本概念	1
1.2 数据库系统的发展	5
1.3 数据库系统的基本特点	8
1.4 数据库系统的内部结构体系	9
1.4.1 数据库系统的三级模式	10
1.4.2 数据库系统的两级映射	10
习题 1	11
第 2 章 数据模型	12
2.1 数据模型的基本概念	12
2.2 数据模型的四个世界	12
2.3 概念世界与概念模型	13
2.3.1 E-R 模型	13
2.3.2 扩充的 E-R 模型、EE-R 模型	19
2.3.3 面向对象模型	21
2.3.4 谓词模型	23
2.4 信息世界与逻辑模型	24
2.4.1 概述	24
2.4.2 层次模型与层次模型数据库系统	25
2.4.3 网状模型与网状模型数据库系统	27
2.4.4 关系模型与关系模型数据库系统	31
2.5 计算机世界与物理模型	32
2.5.1 文件系统	32
2.5.2 关系模型的存储结构	33
习题 2	34
第 3 章 关系数据库系统	35
3.1 关系数据库系统概述	35
3.2 关系数据库系统的衡量准则	36
3.3 关系模型	37
3.3.1 关系数据结构	38
3.3.2 关系操纵	40
3.3.3 关系中的数据约束	41
3.4 关系代数——关系模型数学理论之一	41
3.4.1 关系的表示	41
3.4.2 关系操纵的表示	42
3.4.3 关系模型与关系代数	44

3.4.4	关系代数中的扩充运算	44
3.4.5	关系代数实例	48
3.5	关系演算——关系模型数学理论之二	49
3.5.1	关系的表示	49
3.5.2	关系操纵的表示	50
3.5.3	关系演算的例子	51
3.5.4	关系演算的安全限制	52
3.5.5	关系代数与关系演算	52
习题 3		52
第 4 章	关系数据库系统标准语言 SQL	54
4.1	SQL 概貌	54
4.2	SQL 数据定义功能	55
4.2.1	SQL 基本数据类型	55
4.2.2	基表的定义、删除与修改	56
4.2.3	索引的建立与删除	56
4.3	SQL 数据操纵功能	57
4.3.1	SQL 的查询功能	58
4.3.2	SQL 计算、统计、分类的功能	62
4.3.3	SQL 的更新功能	64
4.4	视图	65
4.5	嵌入式 SQL	67
4.6	动态 SQL	68
4.7	SQL 的特点	69
习题 4		69
第 5 章	数据库的物理组织	71
5.1	概论	71
5.2	数据库的物理存储介质	71
5.3	磁盘存储器及其结构	72
5.4	文件组织	74
5.4.1	文件记录与磁盘块	74
5.4.2	文件的定长记录与变长记录	75
5.5	文件记录组织	76
5.6	索引技术与散列技术	77
5.6.1	索引技术	77
5.6.2	索引技术中的 B ⁺ 树	80
5.6.3	散列技术	82
5.7	数据库与文件	83
5.7.1	数据库中数据分类	83
5.7.2	数据库存储空间组织	84

习题 5	85
第 6 章 关系数据库的查询优化	86
6.1 概述	86
6.2 查询优化的可能性	86
6.3 关系代数等价变换规则	88
6.4 查询优化策略与算法	90
习题 6	91
第 7 章 数据库的安全性与完整性	92
7.1 数据库的安全性	92
7.1.1 数据库的安全与安全数据库	92
7.1.2 数据库的安全标准	93
7.1.3 SQL 对数据库安全的支持	97
7.2 数据库的完整性	100
7.2.1 完整性的三个内容	100
7.2.2 完整性规则的组成	101
7.2.3 ORACLE 中的数据库完整性	101
7.2.4 触发器	103
习题 7	104
第 8 章 关系数据库规范化理论	105
8.1 概述	105
8.2 规范化理论	107
8.2.1 函数依赖	108
8.2.2 与函数依赖有关的范式	110
8.2.3 多值依赖与第四范式	114
8.2.4 小结	116
8.3 规范化所引起的一些问题	117
习题 8	118
第 9 章 数据库设计与数据库管理	119
9.1 数据库设计概述	119
9.2 数据库设计的需求分析	120
9.2.1 需求调查	120
9.2.2 需求分析	120
9.2.3 数据需求分析说明书	121
9.3 数据库的概念设计	121
9.3.1 数据库概念设计概述	121
9.3.2 数据库概念设计的过程	122
9.4 数据库的逻辑设计	127
9.4.1 逻辑设计基本方法	127
9.4.2 关系视图设计	130

9.5	数据库的物理设计	130
9.5.1	索引设计	130
9.5.2	集簇设计	131
9.5.3	分区设计	131
9.6	数据库管理	131
9.6.1	数据库的建立	132
9.6.2	数据库的调整	132
9.6.3	数据库的重组	132
9.6.4	数据库安全性控制与完整性控制	132
9.6.5	数据库的故障校复	133
9.6.6	数据库监控	133
9.7	DBA	133
	习题 9	133
第 10 章	事务处理、并发控制与故障恢复技术	134
10.1	事务处理	134
10.1.1	事务	134
10.1.2	事务的性质	135
10.1.3	事务活动	135
10.1.4	有关事务的语句	136
10.2	并发控制技术	136
10.2.1	事务的并发执行	136
10.2.2	封锁	139
10.2.3	封锁协议	139
10.2.4	两阶段封锁协议	141
10.2.5	封锁粒度	142
10.2.6	活锁与死锁	142
10.3	数据库恢复技术	143
10.3.1	概述	143
10.3.2	数据库故障分类	143
10.3.3	数据库故障恢复三大技术	144
10.3.4	恢复策略	145
10.3.5	数据库镜像	146
	习题 10	146
第 11 章	面向对象数据库系统	147
11.1	面向对象概念与数据库系统	147
11.2	面向对象方法的基本思想	148
11.2.1	概述	148
11.2.2	简单回顾	150
11.3	面向对象方法的基本概念	150

11.4	面向对象数据模型	155
11.5	面向对象数据库管理系统	156
11.5.1	面向对象数据库管理系统的内容	157
11.5.2	面向对象数据库语言	161
11.6	面向对象数据库的应用	167
11.6.1	概述	167
11.6.2	面向对象数据库在图形中的应用	168
11.7	面向对象数据库管理系统产品介绍	175
11.7.1	面向对象数据库管理系统的发展阶段	175
11.7.2	面向对象数据库管理系统产品的结构流派	175
11.7.3	典型 OODBMS 产品介绍	176
11.8	对象关系数据库系统	177
	习题 11	178
第 12 章	知识库系统	180
12.1	概 述	180
12.1.1	数据库与人工智能	180
12.1.2	数据库的智能化和智能化数据库	180
12.1.3	知识库系统及其发展	181
12.1.4	知识与知识表示	182
12.1.5	知识库、知识库管理系统与知识库系统	184
12.1.6	知识库语言	185
12.2	知识库系统原理	186
12.2.1	知识库系统的一阶谓词逻辑表示	186
12.2.2	数据库、演绎数据库及知识库	189
12.2.3	基于证明论的知识库系统	192
12.2.4	基于模型论的知识库系统	193
12.3	几个典型的知识库系统	196
12.3.1	NAIL! 系统	196
12.3.2	LDL 系统	197
12.4	知识库系统的应用	198
	习题 12	205
第 13 章	分布式数据库系统	206
13.1	概述	206
13.1.1	何谓分布式数据库	206
13.1.2	分布式数据库系统的特点	207
13.1.3	分布式数据库系统的形成和发展	208
13.2	分布式数据库系统的分类与结构	208
13.2.1	分布式数据库系统的分类	208
13.2.2	分布式数据库系统的一般结构	209

13.2.3	分布式数据库管理系统中的抽象级	210
13.2.4	分布式数据库管理系统	211
13.3	数据分布策略	213
13.3.1	数据的分布方式	213
13.3.2	关系的分割	214
13.3.3	数据分布对数据管理的影响	216
13.4	查询分解和优化	216
13.4.1	全局查询树的变换	217
13.4.2	复本的选择	219
13.4.3	查询树的分解	220
13.4.4	全局查询优化中考虑的问题	221
13.4.5	半联接在分布式联接中的应用	222
13.4.6	用直接联接进行分布式联接	224
13.4.7	多复本的更新	225
习题 13		225
第 14 章	网络与数据库	226
14.1	网络上数据的分布与组织	226
14.1.1	网上的应用结点与数据结点	226
14.1.2	两种典型的网上数据分布结构	227
14.2	网上异构多数据源的访问	229
14.2.1	SQL/CLI 的基本原理和基本内容	229
14.2.2	ODBC 的基本结构与工作流程	232
14.2.3	JDBC 接口	232
14.2.4	CORBA	233
14.3	Web 与 Web 数据库	235
14.3.1	互联网与 Web	235
14.3.2	Web 与数据库接口	235
14.3.3	XML 与 XML 数据库	236
14.3.4	Web 数据库	237
习题 14		238
第 15 章	决策支持系统与数据仓库	239
15.1	决策支持系统	239
15.2	数据仓库的基本原理	240
15.2.1	概论	240
15.2.2	数据仓库特点	241
15.3	数据仓库的基本结构	243
15.3.1	数据源	243
15.3.2	数据仓库管理层	244
15.3.3	数据集市层	244

15.4 数据仓库的设计	245
15.4.1 数据仓库设计的原则	245
15.4.2 数据仓库设计步骤	245
15.5 OLAP 与 OLAP 服务器	247
15.5.1 OLTP 与 OLAP	247
15.5.2 OLAP 的基本内容	247
15.6 数 据 挖 掘	252
15.6.1 数据挖掘的方法	253
15.6.2 数据挖掘的步骤	254
15.6.3 数据挖掘的应用	254
15.7 DSS 中的建模与展示	255
15.8 DSS 整体结构	255
习题 15	256
第 16 章 移动式数据库与嵌入式数据库	257
16.1 概述	257
16.2 移动式数据库与嵌入式数据库的特点	257
16.3 移动式数据库与嵌入式数据库的基本功能	258
16.4 移动式数据库与嵌入式数据库的前景	259
习题 16	259
参考文献	260

第1章

概 论

本章主要介绍数据库系统的基本概况,包括基本概念、特点与其基本内容,本章对全书内容具有提纲契领的作用。

1.1 基本概念

计算机科学与技术的发展,计算机应用的深入与拓展,使得数据库在计算机应用中的地位与作用日益重要,它在商业中、事务处理中占有主导地位,近年来在统计领域、多媒体领域以及智能化应用领域中的地位与作用也变得十分重要。随着网络应用的普及,它在网络中的应用也日渐重要,因此,数据库已成为构成一个计算机应用系统的重要支持。本书以数据库为核心对其基本原理、应用作全面的介绍。本节将对其六个基本概念作介绍,它们是数据、数据库、数据库管理系统、数据库管理员、数据库系统及数据库应用系统。

1. 数据 (data)

数据是计算机领域中的专门术语,属软件范畴。数据是一种有结构的符号串,它是软件中程序加工的原料与结果。在计算机中一个软件只有有了程序与数据才构成完整的运行实体,因此软件即是由程序与数据这两部分组成。

数据有很多特性与分类。

■ 数据表现的多样性

数据表现的形式很多,除常用的数字、文字、时间等表示形式外,还包括图像、图形、语言、视频等多媒体数据以及表示知识、规则、数学符号及推理等抽象数据,数据表现多样性为数据的应用提供了有力的基础。

■ 数据的可构造性

在软件中的数据是有一定结构的,首先,数据有型 (type) 与值 (value) 之分,数据的型给出了数据表示的类型如整型、实型、字符型等,而数据的值给出了符合给定型的数值。随着应用需求的扩大,数据的型有了进一步的扩大,它包括了将多种相关数据以一定结构方式组合构成特定的数据框架称为数据结构 (data structure),在特定条件下可称为数据模式 (data schema)。

■ 数据的挥发性和持久性

计算机中数据一般分为两部分,其中一部分与程序仅有短时间的交互关系,随着程序的结束而消亡,它们称为临时性数据或挥发性 (transient) 数据。这类数据一般存放于计算机内存中,而另一部分数据则对系统起着长期持久的作用,它们称为持久性 (persistent) 数据,这类数据一般存放于计算机中的次级存储器内 (如磁盘)。

■ 数据的私有性与共享性

数据按其使用对象可分为私有性与共享性两种。为特定应用 (程序) 服务的数据称私有数据 (private),而为多个应用 (程序) 服务的数据则称为共享 (share) 数据。

■ 数据“量”的表示

数据有量的区别，一般讲数据的量大致分为小量、大量及海量三种。数据的量是衡量与区别数据的重要标志，这主要是由于数据“量”的变化可能会引起数据“质”的变化。

随着技术的进步与应用的扩大，数据的“质”、“量”与“地位”都在发生变化，这些变化主要表现为

- (1) 数据的量由小量到大量、由大量到海量。
- (2) 数据的结构由简单到复杂。
- (3) 数据的服务范围由私有到共享。
- (4) 最根本性的变化是数据在软件中的地位由附属地位到主导地位。

过去，在软件系统中是以程序为主体，而数据则以私有形式从属于程序。此时数据在系统中是分散、凌乱的，这也是造成了数据管理的混乱，如数据冗余度高、数据一致性差以及数据的安全性差等多种弊病。近 10 余年来，数据在软件系统中的地位产生了变化，在数据库系统及数据库应用系统中数据已占有主体地位，而程序已退居附属地位。在此种结构中需要对数据作集中、统一的管理并使其对应用（程序）共享，其关系结构如图 1.1 所示。

本书所介绍的数据库系统是以海量的、复杂结构的、持久的和共享的数据的统一管理为目标，本书如不作特别说明，所提数据即具此四种特性。

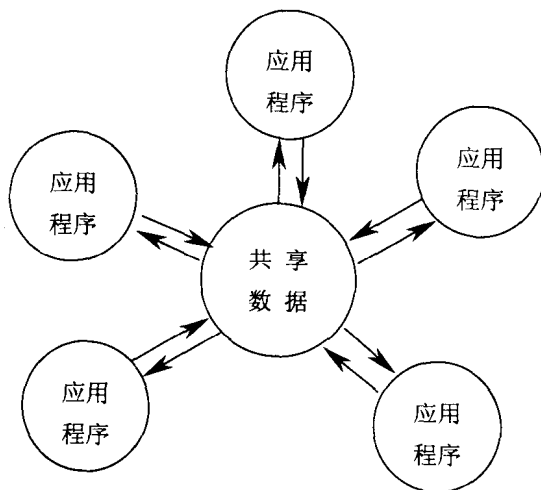


图 1.1 以数据为主体的软件系统

2. 数据库 (database, 简称 DB)

数据库是数据的集合，它具有统一的结构形式并存放于统一的存贮介质内，它由多种应用数据集成，并可被应用所共享。

数据库存放数据，数据按所提供的数据模式存放，它能构造复杂的数据结构以建立数据间内在联系与复杂关系，从而构成数据的全局结构模式。

数据库中的数据具有“集成”、“共享”的特点，亦即是数据库集中了各种应用的数据，并对其进行统一的构造与存储，而数据可为不同应用服务与使用。

3. 数据库管理系统 (database management system, 简称 DBMS)

数据库管理系统是管理数据库的机构，它是一种软件（属系统软件），它负责：