



“十二五”普通高等教育本科国家级规划教材

数据库系统概论

(第5版)

王珊 萨师焯 编著

高等教育出版社



“十二五”普通高等教育本科国家级规划教材

数据库系统概论

(第5版)

王珊 萨师焯 编著

高等教育出版社·北京

内容提要

本书第1版于1983年出版,至今已修订至第5版。第5版被列入“十二五”普通高等教育本科国家级规划教材。相应课程于2004年被评为北京市精品课程,2005年被评为国家精品课程,2014年被批准为国家精品课程资源共享课。

本书系统全面地阐述了数据库系统的基础理论、基本技术和基本方法。全书分为4篇16章。第一篇基础篇,包括绪论、关系数据库、关系数据库标准语言SQL、数据库安全性和数据库完整性,共5章;第二篇设计与应用开发篇,包括关系数据理论、数据库设计和数据库编程,共3章;第三篇系统篇,包括关系查询处理和查询优化、数据库恢复技术、并发控制和数据库管理系统,共4章;第四篇新技术篇,包括数据库技术发展概述、大数据管理、内存数据库系统和数据仓库与联机分析处理技术,共4章。

本书可以作为高等学校计算机类专业、信息管理与信息系统等相关专业数据库课程的教材。也可供从事数据库系统研究、开发和应用的工程技术人员和工程技术人员参考。

图书在版编目(CIP)数据

数据库系统概论 / 王珊, 萨师焯编著. --5版. --

北京: 高等教育出版社, 2018.5

ISBN 978-7-04-049683-3

I. ①数… II. ①王… ②萨… III. ①数据库系统-高等学校-教材 IV. ①TP311.13

中国版本图书馆CIP数据核字(2018)第071744号

策划编辑 倪文慧

责任编辑 倪文慧

版式设计 王艳红

责任校对 吕红颖

责任印制 赵义民

出版发行 高等教育出版社

社址 北京市西城区德外大街4号

邮政编码 100120

印刷 北京盛通印刷股份有限公司

开本 787mm×1092mm 1/16

印张 27.25

字数 560千字

购书热线 010-58581118

咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>

<http://www.hep.com.cn>

网上订购 <http://www.hepmall.com.cn>

<http://www.hepmall.com>

<http://www.hepmall.cn>

版 次 1983年4月第1版

2018年5月第5版

印 次 2018年5月第1次印刷

定 价 98.00元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换

版权所有 侵权必究

物料号 49683-00

作者简介



王珊，中国人民大学教授、博士生导师。现任数据工程与知识工程教育部重点实验室学术委员会副主任。曾任中国人民大学信息学院院长，中国计算机学会副理事长，中国计算机学会数据库专业委员会主任，教育部第五届科学技术委员会委员等。

自1981年以来一直从事数据库方向的教学、科研和系统开发工作。1984—1986年应邀赴美国马里兰大学工作，设计和开发可扩展的关系数据库管理系统XDB。回国后主持和承担了国家科技攻关项目、863高科技计划项目、北京市科技计划重大项目、

国家自然科学基金重点及重大项目、国家核高基科技重大专项等50多项科研项目。在国内外杂志发表论文300余篇，出版数据库方面的著作30多部。

30多年来培养博士硕士研究生130余名，主持研制和开发了多个具有自主知识产权的数据库管理系统。科研和教学成果曾获多项国家级和省部级科技进步奖；编著的教材曾获国家级优秀教材奖，相应课程被评为国家精品课程资源共享课；曾获全国五一劳动奖章以及全国优秀教师、国家教学名师等荣誉称号。



萨师煊（1922—2010），中国人民大学教授，我国最早从事数据库教学和科研的专家之一，中国数据库学科的奠基人。在国内率先设立了数据库系统课程，发表了多篇论文和多部著作。历任中国人民大学信息系主任、中国计算机学会常务理事及中国计算机学会数据库学组组长等职务。我国数据库学术活动的倡导者和领导者，对我国数据库技术的发展、应用和学术交流起了非常重要的推动作用，对

我国数据库技术跟踪国际前沿、缩短与发达国家的差距做出了重要贡献。

前 言

为了反映数据库学科的新成果和应用的新方向,适应数据库技术的进展,保持本书的先进性、科学性和实用性,我们对本书第4版进行了修订。本书第1版、第2版、第3版和第4版分别于1983年、1991年、2000年和2006年出版。第5版是“十二五”普通高等教育本科国家级规划教材。

本书分为4篇16章,如下表所示。第1至第11章(表中序号1~11)是本科专业的教程(书中有*号的部分除外),第12至第16章(表中序号12~16)是高级教程。

本科生必读	1. 绪论 2. 关系数据库 3. 关系数据库标准语言 SQL 4. 数据库安全性 5. 数据库完整性	基础篇	实验准备 实验 1. 数据库定义与操作语言 实验 2. 安全性语言 实验 3. 完整性语言 实验 4. 触发器	DBA 和数据库 应用系统 设计开发 人员
	6. 关系数据理论 7. 数据库设计 8. 数据库编程	设计与应用开发篇	实验 5. 数据库设计 实验 6. 存储过程 实验 7. 数据库应用开发 实验 8. 数据库设计与应用开发大作业	
	9. 关系查询处理和查询优化 10. 数据库恢复技术 11. 并发控制	系统篇	实验 9. 数据库监视与性能优化 实验 10. 数据库恢复技术 实验 11. 并发控制	
本科生、研究生选读	12. 数据库管理系统			
	13. 数据库技术发展概述 14. 大数据管理 15. 内存数据库系统 16. 数据仓库与联机分析处理技术	新技术篇		

第5版主要修改的内容包括:

(1) 在基础篇中保持重点讲解关系数据库系统的传统,对SQL的内容根据标准的发展做了相应更新。随着数据的安全性、完整性越来越重要,对数据库安全性和完整性的内容进行了补充修改。

(2) 在设计与应用开发篇中把原来第4版第1章讲解的E-R图设计移到了第7章7.3概念结构设计一节中,成为概念结构设计的重要知识点。作为选读,增加了扩展E-R图的内容。修改补充了第8章数据库编程中PL/SQL、存储过程和函数、ODBC、OLE DB、JDBC等概念和方法。

(3) 在系统篇的第11章并发控制中增加了三级封锁协议的内容,并在11.8其他并发控制机制一节(作为选读内容)概要介绍了多版本并发控制(MVCC)技术。

(4) 在新技术篇中修改了第13章数据库技术发展概述和第16章数据仓库与联机分析处理技术的内容,增加了反映数据管理最新发展的重要技术,如大数据管理、内存数据库系统等章节,限于篇幅删去了第4版中分布式数据库系统、对象关系数据库系统和XML数据库。

(5) 提供了作者编写的配套教辅用书《数据库系统概论(第5版)实验指导与习题解析》(高等教育出版社出版)。其中根据教材章节内容安排了必修实验和选修实验,进一步加强了实验教学环节;对本书各章习题做了解析,还增加了补充习题。

在高等教育出版社易课程网站 <http://abook.hep.com.cn/187532> 和精品课程教学网站 <http://www.chinadb.ruc.edu.cn/> 上均给出了本书配套的教学资源,包括电子教案、教学视频、实验要求及部分实验报告示例、补充习题及参考答案等,供读者学习参考。这些实验均使用国产金仓数据库管理系统 Kingbase ES 作为实验平台。该系统可以从北京人大金仓信息技术股份有限公司的网站 <http://www.kingbase.com.cn> 免费下载。

本书内容全面丰富,教师可以针对不同专业和不同类别的学生挑选书中不同章节的内容进行讲解。

全书由王珊教授执笔。

大连海事大学张俊教授详细审阅了书稿,提出了许多有益的意见。中国人民大学陈红教授和杜小勇教授根据讲授本书的实际体会,对内容和实验提出了许多中肯有益的修改建议并审阅了书稿。陈红教授、杜小勇教授、张孝副教授、文继荣教授、李翠平教授、中国调查与数据中心(中国人民大学)张延松博士、教育部数据工程与知识工程重点实验室(中国人民大学)陈跃国博士等协助修改和撰写了部分内容,在此向他们表示衷心感谢。

北京人大金仓信息技术股份有限公司任永杰博士、冯玉博士、李海华博士和冷建全高工等提供了金仓数据库产品的技术资料,研究生周宁南和孟庆钟等参与了部分资料的收集

工作，在此向他们表示衷心感谢。

还要感谢广大读者、教师和学生在使用本书时通过邮件、课程网站对本书内容和习题提出的问题和建议，这是本书修订中非常宝贵的参考资料。

在本书的修订过程中，作者阅读参考了大量国内外教材、专著、论文和资料，努力跟踪数据库学科的新发展、新技术，有选择地把它们纳入到教材中来，但因学科发展太快，书中必有许多不足之处，希望学术同仁不吝赐教。

王 珊

2014 马年春节

目 录

第一篇 基础篇

第 1 章 绪论	3	2.1.3 关系数据库	43
1.1 数据库系统概述	3	2.1.4 关系模型的存储结构	43
1.1.1 数据库的 4 个基本概念	3	2.2 关系操作	43
1.1.2 数据管理技术的产生和发展	6	2.2.1 基本的关系操作	43
1.1.3 数据库系统的特点	9	2.2.2 关系数据语言的分类	44
1.2 数据模型	14	2.3 关系的完整性	45
1.2.1 两类数据模型	15	2.3.1 实体完整性	45
1.2.2 概念模型	16	2.3.2 参照完整性	45
1.2.3 数据模型的组成要素	17	2.3.3 用户定义的完整性	48
1.2.4 常用的数据模型	18	2.4 关系代数	48
1.2.5 层次模型	19	2.4.1 传统的集合运算	49
1.2.6 网状模型	22	2.4.2 专门的关系运算	50
1.2.7 关系模型	25	*2.5 关系演算	57
1.3 数据库系统的结构	27	*2.5.1 元组关系演算语言 ALPHA	57
1.3.1 数据库系统模式的概念	27	*2.5.2 元组关系演算	62
1.3.2 数据库系统的三级模式结构	28	*2.5.3 域关系演算语言 QBE	64
1.3.3 数据库的二级映像功能与 数据独立性	29	2.6 小结	70
1.4 数据库系统的组成	31	习题	70
1.5 小结	33	实验	72
习题	34	本章参考文献	72
本章参考文献	35	第 3 章 关系数据库标准语言 SQL	75
第 2 章 关系数据库	37	3.1 SQL 概述	75
2.1 关系数据结构及形式化定义	38	3.1.1 SQL 的产生与发展	75
2.1.1 关系	38	3.1.2 SQL 的特点	76
2.1.2 关系模式	42	3.1.3 SQL 的基本概念	78
		3.2 学生-课程数据库	79

3.3 数据定义	80	4.2.3 自主存取控制方法	140
3.3.1 模式的定义与删除	81	4.2.4 授权: 授予与收回	141
3.3.2 基本表的定义、删除与修改	82	4.2.5 数据库角色	145
3.3.3 索引的建立与删除	87	4.2.6 强制存取控制方法	147
3.3.4 数据字典	89	4.3 视图机制	149
3.4 数据查询	89	4.4 审计	149
3.4.1 单表查询	90	4.5 数据加密	151
3.4.2 连接查询	99	4.6 其他安全性保护	153
3.4.3 嵌套查询	103	4.7 小结	154
3.4.4 集合查询	111	习题	154
3.4.5 基于派生表的查询	113	实验	155
3.4.6 SELECT 语句的一般格式	114	本章参考文献	156
3.5 数据更新	115	第 5 章 数据库完整性	157
3.5.1 插入数据	115	5.1 实体完整性	158
3.5.2 修改数据	117	5.1.1 定义实体完整性	158
3.5.3 删除数据	118	5.1.2 实体完整性检查和违约处理	159
3.6 空值的处理	119	5.2 参照完整性	160
3.7 视图	121	5.2.1 定义参照完整性	160
3.7.1 定义视图	121	5.2.2 参照完整性检查和违约处理	160
3.7.2 查询视图	124	5.3 用户定义的完整性	163
3.7.3 更新视图	126	5.3.1 属性上的约束条件	163
3.7.4 视图的作用	128	5.3.2 元组上的约束条件	164
3.8 小结	129	5.4 完整性约束命名子句	165
习题	130	*5.5 域中的完整性限制	166
实验	131	5.6 断言	167
本章参考文献	131	5.7 触发器	168
第 4 章 数据库安全性	133	5.7.1 定义触发器	169
4.1 数据库安全性概述	133	5.7.2 激活触发器	171
4.1.1 数据库的不安全因素	133	5.7.3 删除触发器	172
4.1.2 安全标准简介	134	5.8 小结	173
4.2 数据库安全性控制	137	习题	173
4.2.1 用户身份鉴别	138	实验	173
4.2.2 存取控制	140	本章参考文献	173

第二篇 设计与应用开发篇

第6章 关系数据理论	177	7.3 概念结构设计	215
6.1 问题的提出	177	7.3.1 概念模型	215
6.2 规范化	180	7.3.2 E-R模型	215
6.2.1 函数依赖	180	*7.3.3 扩展的E-R模型	218
6.2.2 码	181	*7.3.4 UML	223
6.2.3 范式	182	7.3.5 概念结构设计	224
6.2.4 2NF	182	7.4 逻辑结构设计	231
6.2.5 3NF	184	7.4.1 E-R图向关系模型的转换	231
6.2.6 BCNF	184	7.4.2 数据模型的优化	233
6.2.7 多值依赖	185	7.4.3 设计用户子模式	234
6.2.8 4NF	188	7.5 物理结构设计	234
6.2.9 规范化小结	189	7.5.1 数据库物理设计的内容和方法	235
6.3 数据依赖的公理系统	190	7.5.2 关系模式存取方法选择	235
*6.4 模式的分解	194	7.5.3 确定数据库的存储结构	237
6.4.1 模式分解的三个定义	194	7.5.4 评价物理结构	238
6.4.2 分解的无损连接性和保持 函数依赖性	196	7.6 数据库的实施和维护	238
6.4.3 模式分解的算法	198	7.6.1 数据的载入和应用程序的调试	238
6.5 小结	201	7.6.2 数据库的试运行	239
习题	202	7.6.3 数据库的运行和维护	240
本章参考文献	204	7.7 小结	241
第7章 数据库设计	205	习题	241
7.1 数据库设计概述	205	实验	242
7.1.1 数据库设计的特点	206	本章参考文献	242
7.1.2 数据库设计方法	207	第8章 数据库编程	245
7.1.3 数据库设计的基本步骤	207	8.1 嵌入式SQL	245
7.1.4 数据库设计过程中的各级模式	210	8.1.1 嵌入式SQL的处理过程	245
7.2 需求分析	211	8.1.2 嵌入式SQL语句与主语言 之间的通信	246
7.2.1 需求分析的任务	211	8.1.3 不用游标的SQL语句	249
7.2.2 需求分析的方法	212	8.1.4 使用游标的SQL语句	251
7.2.3 数据字典	213	8.1.5 动态SQL	252

8.2 过程化 SQL	253	8.4.2 ODBC 工作原理概述	260
8.2.1 过程化 SQL 的块结构	253	8.4.3 ODBC API 基础	261
8.2.2 变量和常量的定义	253	8.4.4 ODBC 的工作流程	263
8.2.3 流程控制	254	*8.5 OLE DB	267
8.3 存储过程和函数	255	*8.6 JDBC 编程	269
8.3.1 存储过程	256	8.7 小结	270
8.3.2 函数	258	习题	270
*8.3.3 过程化 SQL 中的游标	258	实验	271
8.4 ODBC 编程	259	本章参考文献	271
8.4.1 ODBC 概述	259		
第三篇 系 统 篇			
第 9 章 关系查询处理和查询优化	275	10.2 数据库恢复概述	294
9.1 关系数据库系统的查询处理	275	10.3 故障的种类	295
9.1.1 查询处理步骤	275	10.4 恢复的实现技术	297
9.1.2 实现查询操作的算法示例	277	10.4.1 数据转储	297
9.2 关系数据库系统的查询优化	279	10.4.2 登记日志文件	298
9.2.1 查询优化概述	280	10.5 恢复策略	300
9.2.2 一个实例	280	10.5.1 事务故障的恢复	300
9.3 代数优化	282	10.5.2 系统故障的恢复	300
9.3.1 关系代数表达式等价变换规则	283	10.5.3 介质故障的恢复	301
9.3.2 查询树的启发式优化	284	10.6 具有检查点的恢复技术	301
9.4 物理优化	286	10.7 数据库镜像	303
9.4.1 基于启发式规则的存取路径 选择优化	286	10.8 小结	304
9.4.2 基于代价估算的优化	287	习题	305
*9.5 查询计划的执行	289	实验	306
9.6 小结	289	本章参考文献	306
习题	290	第 11 章 并发控制	309
实验	291	11.1 并发控制概述	310
本章参考文献	291	11.2 封锁	312
第 10 章 数据库恢复技术	293	11.3 封锁协议	312
10.1 事务的基本概念	293	11.4 活锁和死锁	315
		11.4.1 活锁	315

11.4.2 死锁	315	12.2 数据库管理系统的系统结构	330
11.5 并发调度的可串行性	317	12.2.1 数据库管理系统的层次结构	330
11.5.1 可串行化调度	317	12.2.2 关系数据库管理系统的运行 过程示例	331
11.5.2 冲突可串行化调度	317	12.3 语言处理层	333
11.6 两段锁协议	319	12.3.1 语言处理层的任务和工作步骤	333
11.7 封锁的粒度	320	12.3.2 解释方法	335
11.7.1 多粒度封锁	321	12.3.3 预编译方法	336
11.7.2 意向锁	321	12.4 数据存取层	337
*11.8 其他并发控制机制	323	12.4.1 数据存取层的系统结构	338
11.8.1 多版本并发控制	323	12.4.2 数据存取层的功能子系统	339
11.8.2 改进的多版本并发控制	324	12.5 缓冲区管理	341
11.9 小结	325	12.6 数据库的物理组织	343
习题	326	12.7 小结	344
实验	327	习题	345
本章参考文献	327	本章参考文献	346
*第12章 数据库管理系统	329		
12.1 数据库管理系统的基本功能	329		

第四篇 新技术篇

第13章 数据库技术发展概述	349	习题	362
13.1 数据库技术发展历史回顾	349	本章参考文献	363
13.2 数据库发展的三个阶段	350	第14章 大数据管理	365
13.2.1 第一代数据库系统	350	14.1 大数据概述	365
13.2.2 第二代数据库系统	351	14.1.1 什么是大数据	365
13.2.3 新一代数据库系统	351	14.1.2 大数据的特征	366
13.3 数据库系统发展的特点	352	14.2 大数据的应用	369
13.3.1 数据模型的发展	352	14.2.1 感知现在 预测未来——互联网 文本大数据管理与挖掘	369
13.3.2 数据库技术与相关技术相结合	355	14.2.2 数据服务 实时推荐——基于大 数据分析的用户建模	375
13.3.3 面向应用领域的数据库新技术	357	14.3 大数据管理系统	377
13.4 数据管理技术的发展趋势	359	14.3.1 NoSQL 数据管理系统	377
13.4.1 数据管理技术面临的挑战	359	14.3.2 NewSQL 数据库系统	378
13.4.2 数据管理技术的发展与展望	360		
13.5 小结	362		

14.3.3 MapReduce 技术	379	15.5 小结	405
14.3.4 大数据管理系统的新格局	380	习题	405
14.4 小结	382	本章参考文献	406
习题	383	第 16 章 数据仓库与联机分析处理技术	409
本章参考文献	383	16.1 数据仓库技术	409
第 15 章 内存数据库系统	385	16.2 联机分析处理技术	413
15.1 概述	385	16.3 数据挖掘技术	414
15.2 内存数据库的发展历程	386	16.4 大数据时代的新型数据仓库	416
15.3 内存数据库的特性	389	16.5 小结	419
15.4 内存数据库的关键技术	389	习题	420
15.4.1 数据存储	390	本章参考文献	420
15.4.2 查询处理及优化	392	数字资源使用说明	421
15.4.3 并发与恢复	403		

第一篇

基础篇

本篇介绍数据库系统的基本概念和基础知识，是读者进一步学习后面各个章节以及数据库系统其他课程的基础。

基础篇包括5章。

第1章绪论，初步讲解数据库的基本概念，介绍数据模型的组成要素和常用的数据模型、数据库系统的三级模式结构和数据库系统的主要组成部分。

第2章关系数据库，系统讲解关系数据库的重要概念，包括关系模型和关系代数。

第3章关系数据库标准语言 SQL，系统而详尽地讲解 SQL 的数据定义、数据查询和数据更新三部分功能。

第4章数据库安全性，全面讲解实现数据库系统安全性的技术和方法，包括用户身份鉴别、自主存取控制、强制存取控制、视图机制、审计功能、数据加密存储和加密传输等。

第5章数据库完整性，系统而详尽地讲解实体完整性、参照完整性和用户定义的完整性，包括这些完整性约束的定义方法、完整性检查机制和违约处理，断言与触发器的作用和使用方法。

数据库是数据管理的有效技术，是计算机科学的重要分支。今天，信息资源已成为各个部门的重要财富和资源。建立一个满足各级部门信息处理要求的行之有效的信息系统也成为企业或组织生存和发展的重要条件。因此，作为信息系统核心和基础的数据库技术得到越来越广泛的应用，从小型单项事务处理系统到大型信息系统，从联机事务处理（On-Line Transaction Processing, OLTP）到联机分析处理（On-Line Analysis Processing, OLAP），从一般企业管理到计算机辅助设计与制造（CAD/CAM）、计算机集成制造系统（CIMS）、电子政务（e-Government）、电子商务（e-Commerce）、地理信息系统（GIS）等，越来越多的应用领域采用数据库技术来存储和处理信息资源。特别是随着互联网的发展，广大用户可以直接访问并使用数据库，例如通过网上订购图书、日用品、机票、火车票，通过网上银行转账存款取款、检索和管理账户，等等。数据库已经成为每个人生活中不可缺少的部分。

因此，数据库课程不仅是计算机类专业、信息管理专业的重要课程，也是许多非计算机专业的选修课程。

本章介绍数据库系统的基本概念，包括数据管理技术的发展过程、数据库系统的组成部分等。读者从中可以学习到为什么要使用数据库技术、数据库技术的重要性。本章是后面各章节的准备和基础。

1.1 数据库系统概述

在系统地介绍数据库的基本概念之前，这里首先介绍一些数据库最常用的术语和基本概念。

1.1.1 数据库的4个基本概念

数据、数据库、数据库管理系统和数据库系统是与数据库技术密切相关的4个基本概念。

1. 数据 (data)

数据是数据库中存储的基本对象。数据在大多数人头脑中的第一个反应就是数字，例如 93、1 000、99.5、-330.86、¥6 880、\$726 等。其实数字只是最简单的一种数据，是数据的一种传统和狭义的理解。广义的理解认为数据的种类很多，例如文本 (text)、图形 (graph)、图像 (image)、音频 (audio)、视频 (video)、学生的档案记录、货物的运输情况等，这些都是数据。

可以对数据做如下定义：**描述事物的符号记录称为数据**。描述事物的符号可以是数字，也可以是文字、图形、图像、音频、视频等，数据有多种表现形式，它们都可以经过数字化后存入计算机。

在现代计算机系统中数据的概念是广义的。早期的计算机系统主要用于科学计算，处理的数据是数值型数据，如整数、实数、浮点数等。现在计算机存储和处理的对象十分广泛，表示这些对象的数据也随之变得越来越复杂。

数据的表现形式还不能完全表达其内容，需要经过解释，数据和关于数据的解释是不可分的。例如，93 是一个数据，可以是一个同学某门课的成绩，也可以是某个人的体重，还可以是计算机系 2013 级的学生人数。数据的解释是指对数据含义的说明，数据的含义称为数据的语义，**数据与其语义是不可分的**。

在日常生活中，人们可以直接用自然语言（如汉语）来描述事物。例如，可以这样来描述某校计算机系一位同学的基本情况：李明同学，男，1995 年 5 月生，江苏省南京市人，2013 年入学。在计算机中常常这样来描述：

（李明，男，199505，江苏省南京市，计算机系，2013）

即把学生的姓名、性别、出生年月、出生地、所在院系、入学时间等组织在一起，构成一个记录。这里的学生记录就是描述学生的数据。这样的数据是有结构的。记录是计算机中表示和存储数据的一种格式或一种方法。

2. 数据库 (DataBase, DB)

数据库，顾名思义，是存放数据的仓库。只不过这个仓库是在计算机存储设备上，而且数据是按一定的格式存放的。

人们收集并抽取出一个应用所需要的大量数据之后，应将其保存起来，以供进一步加工处理，抽取有用信息。在科学技术飞速发展的今天，人们的视野越来越广，数据量急剧增加。过去人们把数据存放在文件柜里，现在人们借助计算机和数据库技术科学地保存和管理大量复杂的数据，以便能方便而充分地利用这些宝贵的信息资源。

严格地讲，数据库是长期储存在计算机内、有组织的、可共享的大量数据的集合。数据库中的数据按一定的数据模型组织、描述和储存，具有较小的冗余度 (redundancy)、较高的数据独立性 (data independency) 和易扩展性 (scalability)，并可为各种用户共享。