



“十三五” 高等教育规划教材
高等院校电气信息类专业“互联网+”创新规划教材



数据库原理与应用(SQL Server版)

(第2版)

主 编 毛一梅 郭 红



教材预览、申请样书



微信公众号: pup6book



北京大学出版社
PEKING UNIVERSITY PRESS

“十三五”普通高等教育本科规划教材
高等院校电气信息类专业规划教材

数据库原理与应用（SQL Server 版） （第2版）

毛一梅 郭 红 主编



北京大学出版社
PEKING UNIVERSITY PRESS

内 容 简 介

本书第2版从数据库原理入手,结合 Microsoft SQL Server 2012 的具体应用,详细介绍了数据库技术的相关知识。本书第2版分3篇共10章:第1篇为导入篇,包括第1章和第2章,主要介绍数据库的相关概念和发展进程;第2篇为基础篇,包括第3章至第8章,以 Microsoft SQL Server 2012 为实施工具介绍数据库的基本操作技巧;第3篇为应用篇,包括第9章和第10章,介绍数据库编程和数据库设计的方法,本书的特点在于理论与实际的紧密结合,各章中有大量的应用实例供读者学习和提高。

本书适合作为本科院校数据库相关课程的教材,也可供数据库技术初、中级水平的读者自学使用。

图书在版编目(CIP)数据

数据库原理与应用:SQL Server 版/毛一梅,郭红主编. —2版. —北京:北京大学出版社,2017.6
(高等院校电气信息类专业“互联网+”创新规划教材)

ISBN 978-7-301-28262-5

I. ①数… II. ①毛… ②郭… III. ①关系数据库系统—高等学校—教材 IV. ①TP311.138

中国版本图书馆 CIP 数据核字(2017)第 095408 号

书 名 数据库原理与应用(SQL Server 版)(第2版)

SHUJUKU YUANLI YU YINGYONG

著作责任者 毛一梅 郭 红 主编

策划编辑 郑 双

责任编辑 李娉婷

数字编辑 陈颖颖

标准书号 ISBN 978-7-301-28262-5

出版发行 北京大学出版社

地 址 北京市海淀区成府路 205 号 100871

网 址 <http://www.pup.cn> 新浪微博: @北京大学出版社

电子信箱 pup_6@163.com

电 话 邮购部 62752015 发行部 62750672 编辑部 62750667

印刷者 北京鑫海金澳胶印有限公司

经 销 者 新华书店

787 毫米×1092 毫米 16 开本 22.75 印张 536 千字

2012 年 2 月第 1 版

2017 年 6 月第 2 版 2017 年 6 月第 1 次印刷

定 价 52.00 元

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有,侵权必究

举报电话: 010-62752024 电子信箱: fd@pup.pku.edu.cn

图书如有印装质量问题,请与出版部联系,电话: 010-62756370

第2版前言

数据库技术是当今世界高新技术潮流中的主流技术之一，在信息量高速迸发的今天，数据库技术越来越被重视，各行各业均离不开数据库，如何更好地管理和使用数据库也是企业和政府格外关注的话题。

作为应用型本科类的学校，人才培养目标不是造就研究型的知识精英，而是要打造有一定文化素养的、有实用价值的人才。因此，编者于2010年编写了《数据库原理与应用(SQL Server版)》。在本书的编写过程中，编者充分考虑应用型本科学生自身的特点和发展方向，把数据库技术中的原理与具体的应用紧密结合，把数据库技术的运用方法和技巧融入具体的数据库应用实例中去，深入浅出、循序渐进地讲解了数据库系统的基本概念和基本理论。本书第1版出版后受到了广大读者的好评。鉴于近年来数据库技术的快速发展，应读者的要求，编者对本书第1版进行修订改编，对数据库新技术方面的知识进行了普及性介绍，如数据仓库、分布式数据库、大数据等概念；本书第2版以数据库原理为知识背景，以主流数据库 Microsoft SQL Server 2012 为实施工具，并提供了实验指导，详细列出了操作方法和步骤，方便教师教学和学生自学；另外，在本书第2版的最后一章给出了针对具体的信息系统进行数据库设计的完整分析思路和实施步骤。

本书第2版分3篇共10章：第1篇为导入篇，包括第1章和第2章，主要介绍数据库的相关概念和发展进程，如数据库相关定义、数据库的起源与近年来所应用的新技术、数据模型的概念、关系的数学定义、关系的完整性及关系的规范化等知识；第2篇为基础篇，包括第3章至第8章，以 Microsoft SQL Server 2012 为实施工具介绍数据库的基本操作技巧，如数据库的管理、数据表的管理、数据的查询、视图的管理与应用、数据的安全管理等；第3篇为应用篇，包括第9章和第10章，介绍数据库编程和数据库设计的方法，包括函数、批命令、事务、存储过程、触发器等创建与管理，数据库的设计步骤等，并给出了相应的应用案例。

本书第2版的特点是涵盖知识比较全面，既包括了数据库的基础理论知识，又包括了数据库的应用技术，并提供了大量实例。编者根据6年来的教学实践活动，总结了经验，对第1版内容中的第10章和第12章的重点内容归并到了与之相关的章节，并对整本教材的章节顺序进行了部分调节，增加了相应的实验指导，为读者理解相关知识点、提高实际应用能力提供了方便。本书在第1、4、6、10章相关内容旁嵌入了微课视频的二维码，学生通过扫描二维码学习可以更好地预习和复习所学操作。

本书由毛一梅、郭红担任主编，本书第2版具体编写分工如下：第1、2、6、7、10章由上海商学院毛一梅、张晶老师改编，第3、4、5、9章由华北科技学院郭红老师改编，

第8章由毛一梅和华北科技学院吴晓丹老师共同改编，最后由毛一梅、郭红统稿！

在本书的编写过程中，编者得到了上海商学院计算机学院老师们的大力支持，他们对本书的修订提出了宝贵的建议，在此深表谢意。鉴于编者学识水平有限，书中不当之处望广大读者不吝赐教，联系方式：yimei_mao@163.com。

编者
2017年2月



【资源索引】

目 录

第 1 篇 导 入 篇

第 1 章 绪论	2	第 2 章 关系数据理论	16
1.1 数据库系统概述	3	2.1 关系的数学定义	17
1.2 数据管理技术的产生和发展	5	2.2 关系代数	20
1.3 数据库的模式结构	8	2.3 关系的完整性	26
1.4 数据库模型	10	2.4 关系的规范化	29
本章小结	14	2.5 案例分析	37
习题 1	15	本章小结	40
		习题 2	40

第 2 篇 基 础 篇

第 3 章 SQL Server 2012 基础	44	第 5 章 数据表的管理	112
3.1 SQL Server 2012 的新特点	45	5.1 SQL Server 2012 的数据类型	113
3.2 SQL Server 2012 的安装与配置	48	5.2 表结构的创建与修改	118
3.3 SQL Server 2012 系统数据库	57	5.3 表数据的约束	125
3.4 SQL Server 2012 的启动方式	58	5.4 表数据的操作	139
3.5 SQL Server 2012 与旧版文件的兼容性 操作	61	5.5 索引操作	143
3.6 SQL Server 2012 管理控制器	64	5.6 实验指导	152
3.7 疑难分析	66	本章小结	159
本章小结	69	习题 5	160
习题 3	69	第 6 章 数据查询	162
第 4 章 数据库的管理	71	6.1 基本的 SELECT 查询	163
4.1 用户数据库的创建与管理	72	6.2 条件查询	164
4.2 数据的导入与导出	88	6.3 排序查询	170
4.3 用户数据库的分离与附加	96	6.4 分组查询	171
4.4 实验指导	100	6.5 筛选查询	173
本章小结	110	6.6 联结查询	174
习题 4	111	6.7 嵌套查询	180
		6.8 实验指导	182
		本章小结	184



习题 6	185	第 8 章 数据的安全性	204
第 7 章 视图	186	8.1 数据库的安全性机制	205
7.1 基本概念	187	8.2 SQL Server 的安全策略	206
7.2 视图的创建	188	8.3 权限管理	214
7.3 视图的查询	193	8.4 角色管理	219
7.4 视图的修改	195	8.5 数据的备份与还原	224
7.5 疑难分析	199	8.6 疑难分析	253
7.6 实验指导	200	8.7 实验指导	254
本章小结	202	本章小结	257
习题 7	203	习题 8	258

第 3 篇 应 用 篇

第 9 章 数据库编程	260	10.1 数据库设计概述	315
9.1 T-SQL 概述	261	10.2 需求分析	318
9.2 T-SQL 基础	262	10.3 概念结构设计	327
9.3 T-SQL 编程	277	10.4 逻辑结构设计	333
9.4 事务	286	10.5 物理结构设计	337
9.5 存储过程	290	10.6 数据库的实施	341
9.6 触发器	299	10.7 数据库的运行和维护	342
9.7 实验指导	310	10.8 综合案例	343
本章小结	312	本章小结	354
习题 9	312	习题 10	354

第 10 章 数据库设计	314	参考文献	355
--------------------	-----	------------	-----

第 1 篇

导 入 篇

绪 论

教学目标

1. 掌握数据库的基本概念。
2. 了解数据管理的发展历史。
3. 熟悉基本的数据模型。

在信息时代的今天，各行各业都有各自不同的信息管理系统，几乎所有的信息管理系统都要用到数据库，相信大家也曾听过许多关于数据库的专业术语，那么，也许大家会问：

- 什么是数据？数据和信息之间有什么关系？
- 什么是数据库？数据库、数据库系统、数据库管理系统之间有什么联系？
- 数据库系统是如何发展起来的？近年来产生了哪些新的数据库技术？
- 数据库系统是如何将复杂的内部结构屏蔽起来的？
- 什么是数据仓库？数据仓库和传统的数据库系统有什么不同之处？
- 什么是分布式数据库系统？它有什么特点？
- 什么是大数据？它的 4V 特征的具体内容是什么？
- 什么是数据库结构模型？不同的数据库结构模型各有什么优缺点？

本章将详细介绍数据库及数据库技术的相关概念和知识。在完成本章的学习任务之后，相信读者可以轻松地回答以上问题。



【微课视频】

1.1 数据库系统概述

在科技飞速发展的今天,信息无处不在,为了及时获取有效的信息,人们通常要把数据收集起来,然后进行加工处理,从中发现有用的信息。因此,数据处理是当前计算机的主要应用之一,数据库技术是作为一门数据处理技术而发展起来的,所研究的问题就是如何科学地组织和存储数据,如何高效地获取和处理数据。

在当今这个信息爆炸的年代,随着数据量的日益膨胀,数据库技术作为信息分析的核心和基础得到了越来越广泛的应用。

1.1.1 数据

众所周知,做任何一件事情决策很重要,而正确的决策必须有正确的信息作为依据,这些信息来源于事实。

什么是数据?数据(Data)就是对客观事实的记录,它是可以鉴别的符号,这种符号可以是数字、文字、图形、图像、声音等多种表现方式。

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解决烦琐的数字计算,处理的数据主要是整数、实数、浮点数等传统数学中的数据。现代计算机能够存储和处理的对象十分广泛,不仅可以是数字、文本、图形,还可以是音频、视频等多媒体数据。因此,数据的形式越来越复杂。

数据和其语义是不可分的。所谓数据的语义就是指对数据的解释,例如,402 是一个数据,它可能是一个门牌号,也可能是一个货品的编号或价格,如果只有一个数据而没有对它的解释,那么这个数据是毫无意义的。因此,数据不是一个孤立的符号,伴随着数据的出现必须要对该数据的含义进行说明,也就是说,数据是要有语义的。

数据与信息有什么关系?数据与信息不同,数据指的是用符号记录下来的可区别的一种事物的特征或事实,信息是反映现实世界的知识。信息以数据的形式表示,即数据是信息的载体;信息是抽象的,而数据是具体的,信息不随数据设备所决定的数据形式而改变;信息是经过加工,并对客观世界产生影响的数据;信息是对数据的解释。而数据的表现方式可以是不同的,数据既是对客观事实的记录,也是对信息的一种描述。

同一个数据经过不同人的处理可以产生不同的信息,同一个数据在不同背景下也会产生不同的信息。例如,同样一个产品的销售数据对于一个大型企业来说,可能会觉得销售量太小,需要减少产量,调整产品销售策略;而这个销售数据对于一个小型企业来说,可能就觉得销售量很大,需要增加该产品的产量。

数据处理是指将数据向信息转换的过程,包括对数据的收集、存储、传播、检索、分类、加工和输出等活动。

1.1.2 数据库

什么是数据库?数据库(Database, DB)是长期存储在计算机内的、有组织的、可共享的大量数据的集合。数据库中的数据不是杂乱无章地堆积在一起的,而是按照一定的数据

模型组织、描述和存储的。数据库中的数据相互关联,可以为多个用户、多个程序所共享,具有较小的冗余度,数据间联系密切,但又有较高的数据独立性。

数据库技术要解决的主要问题就是如何科学地组织和存储数据,如何高效地获取、更新和加工处理数据,并保证数据的安全性、可靠性和共享性。

数据库技术从诞生到现在,在不到半个世纪的时间里,形成了坚实的理论基础,成熟的商业产品和广泛的应用领域,吸引了越来越多的研究者加入。数据库的诞生和发展给计算机信息管理带来了一场巨大的革命。30多年来,国内外已经开发建设了成千上万个数据库,它已成为企业、部门乃至个人日常工作、生产和生活的基础设施。同时,随着应用的扩展与深入,数据库的数量和规模越来越大,数据库的研究领域也已经大大地拓宽和深化了。

1.1.3 数据库管理系统

数据库管理系统(Database Management System, DBMS)是位于用户与操作系统之间的一层数据管理软件,为用户或应用程序提供访问数据库的方法,是用来管理数据库的计算机应用软件,可以让用户很方便地对数据库进行维护、排序、检索和统计等操作。数据库管理系统的主要目标就是使数据成为方便用户使用的资源,易于为各类用户所共享,它建立在操作系统的基础之上,对数据库进行统一的管理和控制。

数据库管理系统是用户与数据库的接口,应用程序只有通过数据库管理系统才能和数据库“打交道”。数据库管理系统的基本功能有定义数据、组织和管理数据、数据库运行管理、数据库创建和维护等。

数据库管理系统是一个大型的、复杂的软件系统,是信息管理系统中的基础软件。目前专门研制数据库管理系统的厂商及其研制的DBMS产品有很多,比较著名的有IBM公司的DB2关系数据库管理系统和IMS层次数据库管理系统、Oracle公司的Oracle关系数据库管理系统、Sybase公司的Sybase关系数据库管理系统和Microsoft公司的Access、SQL Server关系数据库管理系统等。

1.1.4 数据库系统

数据库系统(Database System, DBS)是实现有组织地、动态地存储大量关联数据,方便多用户访问的计算机软件、硬件和数据资源组成的系统,即采用了数据库技术的计算机系统。从狭义上来讲,数据库系统主要是指数据库、数据库管理系统、应用程序和用户或应用软件。从广义上来讲,它不仅包括数据库、数据库管理系统、应用程序和用户或应用软件,还包括计算机硬件、操作系统和维护人员。其中,数据库管理系统是数据库系统的核心和主体,它保证了数据库的独立性和共享性。

数据库、数据库系统、数据库管理系统之间又有什么联系呢?可以用一个图书馆系统来比拟一个数据库系统,把数据库看作图书馆里的书库,数据库中的数据看作图书馆中的图书,把数据库管理系统看作图书馆管理的操作规程,图书馆中的一切操作如书的存储、查阅、借还等及所有的日常管理都必须按照图书馆指定的操作规程进行,而数据库中对数据的任何操作包括数据的定义、数据的查询、数据的维护、数据库的运行控制等也都必须在数据库管理系统的管理之下进行。



【微课视频】

1.2 数据管理技术的产生和发展

数据管理指的是对数据的分类、组织、编码、存储、检索和维护等。计算机的数据管理主要经历了人工管理、文件系统、数据库系统3个阶段。

1.2.1 人工管理阶段

在20世纪50年代中期以前,计算机主要用于科学计算。由于当时的外存只有纸带、卡片、磁带,没有磁盘等直接存取的存储设备,数据只是在需要时输入,用完后撤走,且没有专门用于管理硬件设备的操作系统,也没有管理数据的专门软件。数据开发人员在应用程序中自己设计、定义和管理数据,应用程序中不仅要规定数据的逻辑结构,还要设计物理结构,包括存储结构、存取方法、输入方式等,数据处理方式是批处理。所有的数据完全由人工进行管理,因此这个阶段被称为人工管理阶段。在这个阶段,数据本身不能独立存储和提供应用,只能是附属于计算机程序的一部分,不能在应用程序之间共享,随着应用程序一起运行与消失。

1.2.2 文件系统阶段

20世纪50年代后期到20世纪60年代中期,随着计算机硬件的发展,有了磁盘、磁鼓等直接存取的存储设备,计算机的应用范围不再局限于科学计算,操作系统中已经有了专门的数据管理软件,一般称为文件系统,处理方式上不仅有了文件批处理,而且能够联机进行实时处理。在文件系统阶段,数据可以以文件的组织方式长期保存在外存上,使用程序反复进行查询、修改、插入、删除等操作;程序和数据之间有了一定的独立性,操作系统提供文件管理功能和访问文件的存取方法,程序和数据之间有了数据存取的接口,数据具有一定的共享性,但是它的共享性是有一定局限的。当不同的应用程序使用部分相同的数据时,仍必须建立各自的文件,而不能共享相同的数据,造成了数据的冗余度大,且不能确保数据的一致性。文件系统阶段数据与程序之间的相互依赖性依然较强。

1.2.3 数据库系统阶段

文件系统中存在的各种问题使人们把希望寄托在数据库系统中。数据库是由逻辑上关联的数据组成的,它存储在一个数据“储藏室”中,数据库在最终用户数据的存储、访问和管理方面采用了不同的方式。通过数据库系统中的数据库管理系统软件,人们可以在很大程度上消除在文件系统中存在的数据不一致、数据异常、数据依赖和结构依赖等问题。

数据库系统与文件系统的主要区别在于:文件系统是操作系统的重要组成部分,而数据库系统是独立于操作系统、在操作系统之上实现的软件,数据库中数据的组织和存储是通过操作系统中的文件系统来实现的;文件系统中的文件是为某一特定应用服务的,文件的逻辑结构对该应用程序来说是优化的,但要想为现有的数据增加一些新的应用会很困难,系统不容易扩充,而数据库系统面向现实世界,它共享性高、冗余度小、易扩充,具有较高的物理独立性和一定的逻辑独立性,整体数据结构化,用数据模型来描述,由数据库管理系统提供数据的安全性、数据的完整性、并发控制能力和数据恢复能力。

使用数据库系统可以大大提高应用程序开发的效率。因为在数据库系统中,应用程序

不必考虑数据的定义、存储和数据存取的具体路径,这些工作都由数据库管理系统来完成。使用数据库系统可以减少开发人员的工作量。当因应用逻辑发生改变而需要改变数据的逻辑结构时,由于数据库系统提供了数据与程序之间的独立性,当数据逻辑结构改变时开发人员不必修改应用程序,或者只需要修改很少的应用程序就能保证应用程序的正常运行,既简化了应用程序的编制又减少了应用程序的维护和修改。

使用数据库系统可以减轻数据库系统管理人员维护系统的负担。因为数据库系统在数据库建立、运行和维护时对数据库进行统一的管理和控制。

总之,使用数据库系统的优点有很多,既便于数据的集中管理,控制数据冗余,提高数据的利用率和一致性,又有利于应用程序的开发和维护。

1.2.4 数据库系统的发展

近年来,随着 Internet 技术及其应用的高速发展,新的数据库技术不断涌现,数据仓库、分布式数据库系统、大数据等名词大家也都耳熟能详,限于篇幅,这里不再做详细的探讨,但为了帮助大家了解近年来数据库系统的发展进程,这里简单地介绍一下相关概念。

1. 数据仓库

随着 20 世纪 90 年代后期 Internet 的兴起与飞速发展,大量的信息和数据要求我们用科学的方法整理,从不同视角对企业经营各方面的数据和信息进行精确分析、准确判断,数据仓库系统也就应运而生。

数据仓库技术是基于数学及统计学的严谨逻辑思维达成“科学的判断、有效的行为”的一个工具。数据仓库技术也是一种实现“数据整合、知识管理”的有效手段。

数据仓库(Data Warehouse, DW 或 DWH)是企业所有级别的决策制定提供所有类型数据支持的战略集合。数据仓库之父比尔·恩门(Bill Inmon)在 1991 年出版的 *Building the Data Warehouse* (《建立数据仓库》)一书给出的定义:数据仓库是一个面向主题(Subject Oriented)的、集成(Integrated)的、相对稳定(Non-Volatile)的、反映历史变化(Time Variant)的数据集合,用于支持管理决策(Decision Making Support)。

与传统数据库面向应用进行数据组织不同,数据仓库中的数据是面向主题进行组织的。主题是指用户使用数据仓库进行决策时所关心的重点问题,一个主题通常与多个操作型信息系统相关。面向主题的数据组织方式,就是在较高层次上对分析对象的数据的一个完整、一致的描述,能完整、统一地体现各个分析对象所涉及的各项数据及数据间的联系。

数据仓库是集成的,数据仓库中的数据是从原有分散的数据库中抽取出来的,由于数据仓库的每一主题所对应的源数据在原有分散的数据库中可能有重复或不一致的地方,加上综合数据不能从原有数据库中直接得到。因此,数据在进入数据仓库之前必须经过统一和综合处理形成集成化的数据。也就是说,将所需数据从原来的数据中抽取出来,进行加工与集成、统一与综合之后才能进入数据仓库。

数据仓库的数据主要供企业决策分析使用,所涉及的数据操作主要是数据查询,一旦某个数据进入数据仓库以后,一般情况下将被长期保留,即数据仓库中一般有大量的查询操作,但修改和删除操作很少,通常只需要定期地加载、刷新。

数据仓库中的数据通常包含历史信息,系统地记录了企业从过去某一时点(如开始应

用数据仓库的时点)到当前的各个阶段的信息,通过这些信息,可以对企业的发展历程和未来趋势做出定量分析和预测。

数据仓库和数据库系统在其功能、面向的用户和存储对象等诸多方面有所不同,其主要区别见表 1-1。

表 1-1 数据仓库和数据库系统的主要区别

数据仓库	数据库系统
管理层使用	一线人员使用
支持战略决策	支持日常运营
用于联机分析	用于事务处理
面向主题	面向应用程序
存储历史数据	存储当前数据
不可预测查询模型	可预测查询模型

随着数据仓库技术的日趋成熟,基于数据仓库的决策支持系统正在许多大型企业中发挥着重要的作用。广义地说,基于数据仓库的决策支持系统由 3 个部件组成:数据仓库技术、联机分析处理技术和数据挖掘技术,其中数据仓库技术是系统的核心。

2. 分布式数据库系统

随着网络的普及与应用,传统集中式数据库系统的不足日渐明显,大量的数据需要在分布的网络上采集和发布。采用集中式处理数据的方式不仅大大加重了网络通信负担,还导致信息系统的规模和配置受到局限,且一旦集中式数据库服务器出现故障,整个系统会完全瘫痪。于是,分布式数据库系统于 20 世纪 70 年代末诞生,80 年代进入成长阶段。20 世纪 90 年代起分布式数据库系统开始进入商品化应用阶段,一些数据库厂商也在不断推出和改进自己的分布式数据库产品。

分布式数据库系统(Distributed Database System, DDBS)是物理上分散而逻辑上集中的数据库系统。一个分布式数据库系统具有物理分布性、逻辑整体性、站点自治性等特点。

1) 物理分布性

分布式数据库系统中的数据不是存储在一个站点上的,而是分散存储在由计算机网络连接起来的多个站点上,而且这种分散存储对用户来说是感觉不到的。所以,分布式数据库系统的数据具有物理分布性,这是与集中式数据库系统的区别之一。

2) 逻辑整体性

分布式数据库系统中的数据物理上分散在各个站点中的,但这些分散的数据逻辑上构成一个整体,它们被分布式数据库系统的所有用户(全局用户)共享,并由一个分布式数据库管理系统统一管理。这使得“分布”对于用户来说是透明的,也就是说用户感觉不到数据是分布在不同地方的。这是分布式数据库的“逻辑整体性”特点,也是与分散式数据库的最大区别。区分一个数据库系统是分散式的还是分布式的,只要判断该数据库系统是否支持全局应用即可。所谓全局应用,是指一个应用所使用的数据涉及两个或两个以上站点上的数据。分布式数据库系统中包含全局数据库(Global Database, GDB)和局部数据库(Local Database, LDB)。全局数据库由全局数据库管理系统(Global Database Management System, GDBMS)进行管理。所谓全局,是指从整个系统角度出发研究问题。局部数据库

由局部数据库管理系统(Local Database Management System, LDBMS)进行管理。所谓局部,是指从各个站点自己的角度出发研究问题。

3) 站点自治性

站点自治性也称场地自治性,各站点上的数据由本地的 DBMS 管理,具有自治处理能力,完成本站点的应用(局部应用),这是分布式数据库系统与多处理机系统的区别。多处理机系统虽然把数据也分散存放于不同的数据库中,但从应用角度来看,这种数据分布与应用程序没有直接的联系,所有的应用程序都由前端机处理,只不过对应用程序的执行由多个处理机进行,这样的系统仍然属于集中式数据库系统的范畴。

3. 大数据

大数据(Big Data),或称巨量资料,指的是所涉及的资料量和规模量巨大到无法通过传统的软件工具,在合理时间内达到撮取、管理、处理并整理成为帮助企业经营决策实现更为积极目的的资讯。麦肯锡全球研究所给出的大数据定义:一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合,它具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征。这四大特征也被人们称为大数据的 4V 特征,即 Volume(大量化)、Variety(多样化)、Velocity(快速化)、Value(价值化)。

大量化:全球在 2010 年正式进入 ZB 时代,互联网数据中心(Internet Data Center, IDC)预计到 2020 年,全球将总共拥有 35ZB 的数据量,企业面临着数据量的大规模增长。目前,大数据的规模尚是一个不断变化的指标,单一数据集的规模范围从几十 TB 到数 PB 不等。

多样化:传统的数据处理对象是结构化数据,然而,实际应用中有许多是非结构化的数据。数据多样性的增加主要是由于新型多结构数据,包括网络日志、社交媒体、互联网搜索、手机通话记录及传感器网络等数据类型造成。其中,部分传感器安装在火车、汽车和飞机上,每个传感器都增加了数据的多样性。

快速化:大数据高速描述的是数据被创建和移动的速度。在高速网络时代,通过基于实现软件性能优化的高速计算机处理器和服务器的,创建实时数据流已成为流行趋势。企业不仅需要了解如何快速创建数据,还必须知道如何将数据快速处理、分析并返回给用户,以满足他们的实时需求。

价值化:大量的不相关信息,浪里淘沙却又弥足珍贵。大数据的价值深藏于浩瀚的数据中,需要使用数据挖掘、机器学习、人工智能等先进技术对其进行分析,将看似不关联的数据和信息关联起来,从而获取较高的价值回报。

大数据技术描述了新一代技术和架构,目的是通过高速捕获、发现和分析数据,经济高效地从种类繁多的大量数据中获益。

1.3 数据库的模式结构

数据库系统是相关数据集和管理这些数据的软件程序的集合。数据库系统的主要作用就是为用户提供数据的抽象视图,大多数数据库用户是非专业用户,因此,系统将数据的存储结构和维护操作细节的复杂性屏蔽起来,为用户提供更为高效的数据检索和维护界面。为此,数据库系统需建立特有的模式结构。



【微课视频】

1.3.1 模式与实例

模式(Schema)是数据库逻辑结构和特征的描述,是对数据库的型的描述,反映了数据的结构及其联系。模式是相对稳定的。

模式的一个实例(Instance)就是模式的一个具体值,反映数据库某一时刻的状态。同一个模式可以有很多实例,实例随数据库中数据的更新而变动,实例是不稳定的。

数据库的模式结构被抽象为3层:视图层、逻辑层和物理层,人们通常也把这3层的模式结构称为外模式、逻辑模式和内模式。数据库的模式结构如图1.1所示。

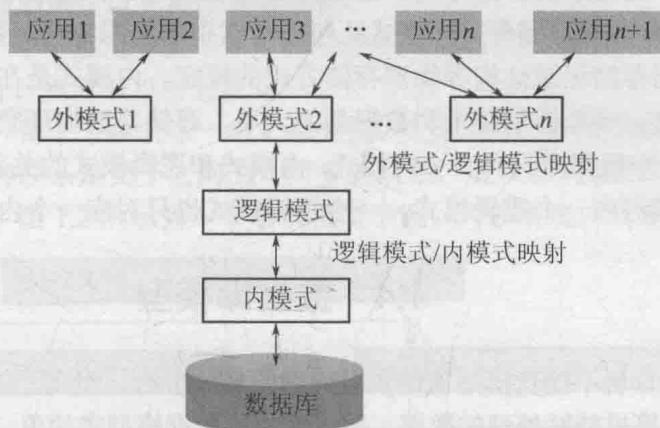


图 1.1 数据库的模式结构

从图1.1中可以看到,数据库的逻辑模式是数据库模式结构中的核心,数据库系统通过外模式到逻辑模式的映射和逻辑模式到物理模式的映射把数据库的物理结构与用户所看见的数据结构隔离开来,这样既方便用户检索和维护数据,也方便数据库设计者对存储的数据结构进行优化,确保数据最小冗余,最大限度满足数据的完整性、一致性和安全性等要求。

1.3.2 外模式

外模式也称子模式(Subschema)或用户模式,是数据库用户(包括应用程序员和最终用户)能够看见和使用的局部数据的逻辑结构和特征的描述,是数据库用户的数据视图,是与某一应用有关的数据的逻辑表示。

一个数据库可以有多个外模式,不同用户或应用程序中能看到不同的数据库外模式,通过外模式,将数据库的内部结构隐藏了起来,因此,外模式是保证数据安全性的一个有力措施。

外模式通常是模式的子集,一个数据库可以有多个外模式。反映了不同用户的应用需求、看待数据的方式、对数据保密的要求。对于模式中的同一数据来说,其在外模式中的结构、类型、长度、保密级别等都可以不同。

模式与外模式是一对多的关系,也就是说一个模式可以对应多个外模式,而一个外模式只对应一个模式;外模式与应用的关系也是一对多的关系,同一外模式可以为某一用户的多个应用程序所使用,但一个应用程序只能使用一个外模式。

1.3.3 逻辑模式

逻辑模式是数据库中全体数据的逻辑结构和特征的描述。逻辑模式是所有用户的公共

数据视图,综合了所有用户的需求,是数据库的核心层,一个数据库只有一个逻辑模式。逻辑模式是数据库系统模式结构的中间层,既与数据的物理存储细节和硬件环境无关,又与具体的应用程序、开发工具及高级程序设计语言无关。

数据的逻辑结构主要指数据项的名称、类型、取值范围等。

1.3.4 内模式

内模式也称存储模式或物理模式,是数据物理结构和存储方式的描述,是数据在数据库内部的表示方式,是记录的存储方式。记录的存储方式规定了数据是顺序存储、按B树结构存储还是按Hash方法存储等。内模式还包括对索引的组织方式、数据是否压缩存储、数据是否加密、数据存储记录结构等物理存储方式的规定。内模式是在逻辑模式的基础上根据用户对检索速度、安全性的要求和数据量的大小、存储器的物理空间、所在位置等实际情况来设计的。一个数据库只有一个内模式,内模式和逻辑模式的关系是一一对应的关系,也就是说一个内模式对应一个逻辑模式,一个逻辑模式也只对应一个内模式。

1.4 数据库模型



【微课视频】

由于计算机不能直接处理现实世界中的具体事物,因此,必须将那些具体的事物转换成计算机能够处理的数据。数据库中用数据模型来抽象、描述和处理现实世界中的数据。数据模型是数据库的核心概念,每个数据库中的数据都是按照某种特定的数据模型来组织的。

1.4.1 数据库模型的基本概念

数据库模型是数据库中用来表示数据结构和数据联系的逻辑概述的集合,数据库模型可以分为两种类型:概念模型和结构模型。

数据库的概念模型是独立于计算机系统的模型,强调的是数据库中描述的是什么,而不是如何描述它。概念模型通常用来描述某个特定组织所关心的信息结构。关于概念模型本书将在第10章中做详细的介绍。这里重点要讨论的是数据的结构模型,它直接面向数据库的逻辑结构,是现实世界的第二层抽象。

数据的结构模型好坏直接影响数据库的性能,因此,选择数据的结构模型是设计数据库的一项重要任务,现有的各种数据库管理系统软件都是基于某种结构模型的。数据库的结构模型包含数据结构、数据操作、数据完整性约束3个部分。

数据结构是所研究对象的数据模型,是指同一数据对象中各数据元素间存在的关系。常见的数据结构有层次结构、网状结构和关系结构等;数据操作是指对数据库各种对象的实例允许执行的操作的集合,包括操作及有关的操作规则;数据完整性约束条件是数据完整性规则的集合,完整性规则是给定的数据模型中数据及其联系所具有的依存规则,用以限定符合数据模型的数据库状态及其状态的变化。

1.4.2 数据库结构模型

数据库结构模型的3个方面内容即数据结构、数据操作和数据完整性约束完整地描述