

21世纪高等学校计算机专业

核心课程规划教材

数据库原理与应用教程 —— SQL Server 2008

(第2版)

◎ 尹志宇 郭晴 主编

清华大学出版社



21世纪高等学校计算机专业
核心课程规划教材

数据库原理与应用教程 —— SQL Server 2008 (第2版)

◎ 尹志宇 郭 晴 主编

李青茹 解春燕 于富强 陈敬利 副主编

清华大学出版社
北京

内 容 简 介

本书全面讲述了数据库的基本原理和 SQL Server 2008 的应用,全书以理论够用、实用,实践第一为原则,使读者能够快速、轻松地掌握 SQL Server 数据库技术与应用。第 1~3 章讲述数据库的基本理论知识,其内容包括数据库系统概述、关系数据库和数据库设计;第 4~13 章讲述数据库管理系统 SQL Server 2008 的应用,其内容包括 SQL Server 2008 数据库基础、数据库的概念和操作、表的基本操作、数据库查询、视图和索引、T-SQL 编程、存储过程和触发器、事务与并发控制、数据库的安全管理、数据库的备份与还原;第 14 章利用一个实例介绍基于 C#.NET 的 SQL Server 数据库系统的开发过程。

本书理论和实践相结合,既阐述了数据库的基本理论,又结合了 SQL Server 2008 数据库管理系统的应用,内容翔实、实例丰富、图文并茂、体系完整、通俗易懂,有助于读者理解数据库的基本概念,掌握要点和攻克难点;为便于学习,每章节还配有丰富的习题。

本书可以作为大学本科、专科、高等职业院校“SQL Server 数据库”课程的教学用书,也可以作为培养数据库系统工程师的培训教材,还可以作为数据库管理人员及数据库应用系统开发人员的参考用书。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

数据库原理与应用教程:SQL Server 2008/尹志宇,郭晴主编.—2 版.—北京:清华大学出版社,2017
(21 世纪高等学校计算机专业核心课程规划教材)
ISBN 978-7-302-45410-6

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

中国版本图书馆 CIP 数据核字(2016)第 260155 号

责任编辑:魏江江 王冰飞

封面设计:刘 键

责任校对:白 蕾

责任印制:宋 林

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社总机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课 件 下 载: <http://www.tup.com.cn>, 010-62795954

印 刷 者:清华大学印刷厂

装 订 者:三河市新茂装订有限公司

经 销:全国新华书店

开 本:185mm×260mm 印 张:21 字 数:535 千字

版 次:2013 年 8 月第 1 版 2017 年 1 月第 2 版 印 次:2017 年 1 月第 1 次印刷

印 数:18001~23000

定 价:39.50 元

前言

数据库最初是在大公司或大机构中用作大规模事务处理的基础,后来随着个人计算机的普及,数据库技术被移植到 PC 上,供单用户个人数据库应用。接着,由于 PC 在工作组内连成网,数据库技术就移植到工作组级。如今,数据库正在 Internet 和内联网中被广泛使用。

数据库技术是目前计算机领域发展最快、应用最广泛的技术,它的应用遍及各行各业,大到操作系统程序,如全国联网的飞机票、火车票订票系统,银行业务系统;小到个人的管理信息系统,如家庭理财系统。在互联网日渐流行的动态网站中,数据库的应用显得尤为重要。

SQL Server 2008 是一个功能完备的数据库管理系统,提供了完整的关系数据库的创建、开发和管理功能。它功能强大、操作简便,日益被广大数据库用户所喜爱,而且越来越多的开发工具提供了与 SQL Server 的接口。

本书在第 1 版的基础上对部分内容做了调整和扩充,实例更丰富、实践性更强,对各章的习题也进行了合理的补充和调整。本书系统地介绍数据库技术的基本理论,全面介绍 SQL Server 2008 的各项功能、数据库系统设计方法、维护和管理以及数据库系统开发应用的相关技术。

全书共 14 章,分为 3 个部分,第一部分是第 1~3 章,系统讲述数据库的基本理论知识,内容包括数据库系统概述、关系数据库、数据库设计;第二部分是第 4~13 章,全面讲述 SQL Server 2008 数据库基础、数据库的概念和操作、表的基本操作、数据库查询、视图和索引、T-SQL 编程、存储过程和触发器、事务与并发控制、数据库的安全管理、数据库的备份与还原;第三部分是第 14 章,利用一个实例介绍基于 C#.NET 的 SQL Server 数据库系统的开发过程。

本书编者长期从事本科计算机类专业的教学工作,具有丰富的教学经验和多年的数据库开发经验,编者深知数据库的主要知识点和重点、难点,了解什么样的教材适合教学使用,学生及各类读者对数据库的学习方式和兴趣所在,以及如何组织书本的内容更有利于教学和自学,从而形成本书的结构体系。

本书第 1、5、9 和 10 章由尹志宇编写,第 2~4 章和第 8 章由郭晴编写,第 6 章由于富强编写,第 7 章由陈敬利编写,第 11 和 12 章由李青茹编写,第 13 和 14 章由解春燕编写,全书由尹



志宇统稿。

为了配合教学和方便读者参考,本书提供了配套的电子教案,读者可以到清华大学出版社网站(<http://www.tup.com.cn>)下载。

由于编者水平有限,书中难免有疏漏与错误之处,衷心希望广大读者批评、指正。

编 者

2016年6月

目 录

第 1 章 数据库系统概述	1
1.1 数据库技术的发展史	1
1.1.1 数据处理技术.....	1
1.1.2 数据库技术的 3 个发展阶段.....	2
1.1.3 数据库技术的新进展.....	3
1.2 数据库系统介绍	7
1.2.1 数据库系统的组成.....	7
1.2.2 数据库的体系结构.....	8
1.3 数据模型.....	10
1.3.1 信息的 3 种世界	10
1.3.2 概念模型	11
1.3.3 常见的 3 种数据模型	14
习题 1	17
第 2 章 关系数据库	19
2.1 关系数据结构.....	19
2.1.1 关系的定义和性质	19
2.1.2 关系模式和关系数据库	21
2.2 关系的完整性.....	22
2.3 关系运算.....	23
2.3.1 传统的集合运算	23
2.3.2 专门的关系运算	25
2.4 关系的规范化.....	29
2.4.1 函数依赖	29
2.4.2 关系规范化的目的	30
2.4.3 关系的规范化过程	32
习题 2	34

第3章 数据库设计	36
3.1 数据库设计概述	36
3.2 需求分析	36
3.2.1 需求分析的任务	36
3.2.2 需求分析的方法	37
3.2.3 数据流图和数据字典	38
3.3 概念结构设计	40
3.3.1 概念结构设计的方法	40
3.3.2 概念结构设计的步骤	40
3.4 逻辑结构设计	41
3.4.1 将 E-R 图转化为关系数据模型	41
3.4.2 关系模式的优化	44
3.4.3 设计用户外模式	44
3.5 物理结构设计	45
3.5.1 确定数据库的物理结构	45
3.5.2 评价物理结构	46
3.6 数据库的实施、运行和维护	46
3.6.1 数据库的实施	46
3.6.2 数据库的运行与维护	47
3.7 数据库设计实例	47
3.7.1 图书借阅管理系统设计	47
3.7.2 钢材仓库管理系统设计	49
习题 3	52
第4章 SQL Server 2008 基础	53
4.1 SQL Server 2008 简介	53
4.1.1 SQL Server 发展史	53
4.1.2 SQL Server 2008 的新增功能	54
4.1.3 SQL Server 2008 的新特性	57
4.2 SQL Server 2008 的安装与配置	58
4.2.1 SQL Server 2008 的版本	58
4.2.2 SQL Server 2008 的环境需求	59
4.2.3 SQL Server 2008 的安装过程	60
4.3 SQL Server 2008 的管理工具	71
4.3.1 服务器管理	71
4.3.2 SQL Server Management Studio	72
4.3.3 SQL Server 配置管理器	73
4.3.4 其他管理工具	74
4.4 T-SQL 语言基础	75

4.4.1	T-SQL 语言的特点	75
4.4.2	T-SQL 语言的分类	76
4.4.3	T-SQL 语言的基本语法	77
习题 4	78
第 5 章	数据库的概念和操作	79
5.1	数据库的基本概念	79
5.1.1	物理数据库	79
5.1.2	逻辑数据库	80
5.2	数据库的操作	81
5.2.1	创建数据库	81
5.2.2	修改数据库	87
5.2.3	删除数据库	89
习题 5	89
第 6 章	表的操作	91
6.1	创建表	91
6.1.1	数据类型	91
6.1.2	使用 SQL Server Management Studio 创建表	94
6.1.3	使用 T-SQL 语句创建表	95
6.2	修改表	98
6.2.1	在 SQL Server Management Studio 中修改表	98
6.2.2	使用 T-SQL 语句修改表	98
6.3	列约束和表约束	99
6.3.1	创建和删除 PRIMARY KEY 约束	99
6.3.2	创建和删除 UNIQUE 约束	100
6.3.3	创建和删除 FOREIGN KEY 约束	101
6.3.4	创建和删除 CHECK 约束	104
6.3.5	创建和删除 DEFAULT 约束	105
6.4	表数据的操作	106
6.4.1	插入数据	106
6.4.2	修改数据	109
6.4.3	删除数据	109
6.4.4	使用 MERGE 语句插入、修改和删除数据	110
6.5	删除表	112
6.6	数据的导入和导出	113
6.6.1	导出数据	113
6.6.2	导入数据	119
习题 6	125



第 7 章 数据库查询	127
7.1 SELECT 查询语法	127
7.2 简单查询	128
7.2.1 投影查询.....	128
7.2.2 选择查询.....	130
7.2.3 聚合函数查询.....	134
7.3 分组和汇总	135
7.3.1 分组查询.....	135
7.3.2 数据汇总.....	138
7.4 连接查询	139
7.4.1 内连接.....	139
7.4.2 自连接.....	140
7.4.3 外连接.....	141
7.4.4 交叉连接.....	143
7.5 子查询	144
7.5.1 无关子查询.....	144
7.5.2 相关子查询.....	146
7.6 其他查询	147
7.6.1 集合运算查询.....	147
7.6.2 对查询结果排序.....	149
7.6.3 存储查询结果.....	150
7.7 在数据操作中使用 SELECT 子句	151
7.7.1 在 INSERT 语句中使用 SELECT 子句	151
7.7.2 在 UPDATE 语句中使用 SELECT 子句	152
7.7.3 在 DELETE 语句中使用 SELECT 子句	153
习题 7	153
第 8 章 视图和索引	155
8.1 视图	155
8.1.1 视图概述.....	155
8.1.2 创建视图.....	156
8.1.3 修改视图.....	160
8.1.4 使用视图.....	162
8.1.5 删除视图.....	164
8.2 索引	165
8.2.1 索引简介.....	165
8.2.2 索引的类型.....	165
8.2.3 创建索引.....	166
8.2.4 查看索引信息.....	171

8.2.5 删除索引	174
习题 8	174
第 9 章 T-SQL 编程	176
9.1 T-SQL 编程基础	176
9.1.1 标识符	176
9.1.2 变量	176
9.1.3 运算符	178
9.1.4 批处理	181
9.1.5 注释	181
9.2 流程控制语句	182
9.2.1 SET 语句	182
9.2.2 BEGIN...END 语句	183
9.2.3 IF...ELSE 语句	183
9.2.4 CASE 语句	184
9.2.5 WHILE 语句	186
9.2.6 GOTO 语句	187
9.2.7 RETURN 语句	187
9.3 函数	188
9.3.1 系统内置函数	188
9.3.2 用户定义函数	190
9.4 游标	194
9.4.1 游标概述	194
9.4.2 游标的类型	195
9.4.3 游标的操作	196
习题 9	200
第 10 章 存储过程和触发器	202
10.1 存储过程	202
10.1.1 存储过程概述	202
10.1.2 存储过程的类型	203
10.1.3 创建存储过程	204
10.1.4 执行存储过程	208
10.1.5 查看存储过程	211
10.1.6 修改和删除存储过程	213
10.2 触发器	215
10.2.1 触发器概述	215
10.2.2 触发器的分类	215
10.2.3 创建 DML 触发器	217
10.2.4 查看触发器信息及修改触发器	227



10.2.5 禁止、启用和删除触发器	229
习题 10	230
第 11 章 事务与并发控制	232
11.1 事务概述	232
11.2 事务的类型	233
11.2.1 根据系统的设置分类	233
11.2.2 根据运行模式分类	234
11.3 事务处理语句	235
11.4 事务的并发控制	237
11.4.1 并发带来的问题	237
11.4.2 锁的基本概念	238
11.4.3 锁的类型	239
11.4.4 锁的信息	241
11.4.5 死锁的产生及解决办法	241
11.4.6 手工加锁	243
习题 11	245
第 12 章 数据库的安全管理	246
12.1 身份验证	246
12.1.1 SQL Server 的身份验证模式	246
12.1.2 设置身份验证模式	247
12.2 账号管理	249
12.2.1 服务器登录账号	249
12.2.2 数据库用户账号	252
12.3 角色管理	254
12.3.1 固定服务器角色	254
12.3.2 数据库角色	256
12.3.3 应用程序角色	261
12.4 权限管理	262
12.4.1 权限的类别	262
12.4.2 权限操作	263
习题 12	269
第 13 章 数据库的备份与还原	270
13.1 数据库备份概述	270
13.1.1 数据库备份计划	270
13.1.2 数据库备份的类型	271
13.2 数据库还原概述	273
13.2.1 数据库还原策略	273

13.2.2	数据库恢复模式	274
13.3	数据库备份操作	276
13.4	数据库还原操作	280
13.4.1	自动还原	280
13.4.2	手动还原	281
13.5	数据库的分离与附加	284
13.5.1	分离数据库	284
13.5.2	附加数据库	286
习题 13	289

第 14 章 基于 C#.NET 的数据库系统开发

290

14.1	C# 语言简介	290
14.2	使用 ADO.NET 访问 SQL Server 2008 数据库	291
14.2.1	ADO.NET 的对象模型	291
14.2.2	利用 ADO.NET 访问数据库的基本操作	293
14.3	数据库系统开发实例	295
14.3.1	数据库设计	295
14.3.2	系统设计与实现	297
习题 14	319

数据库技术是信息系统的一个核心技术,是一门信息管理自动化学科,是计算机学科的一个重要分支。数据库技术所研究的问题是如何科学地组织和存储数据,在数据库系统中减少数据存储冗余,实现数据共享,以及如何保障数据安全、有效地获取和处理。

在数据库系统中是如何抽象、表示、处理现实世界中的信息和数据的呢?客观事物是信息之源,是设计、建立数据库的出发点,也是使用数据库的最后归宿。计算机不能直接处理现实世界中的具体事物,所以人们必须事先将具体事物转换成计算机能够处理的数据,这就是数据库的数据模型。

本章主要介绍数据库技术的发展史,数据库系统的组成和功能,数据库的体系结构;信息的 3 种世界,概念模型和 E-R 图的画法,最常见的 3 种数据模型。

1.1 数据库技术的发展史

从 20 世纪 60 年代末开始到现在,数据库技术已经发展了 40 多年。在这 40 多年的历程中,人们在数据库技术的理论研究和系统开发上取得了辉煌的成就,数据库系统已经成为现代计算机系统的重要组成部分。数据库技术最初是在大公司或大机构中用作大规模事务处理,随着个人计算机的普及,数据库技术被移植到 PC 上供单用户应用;接着,由于 PC 在工作组内连成网,数据库技术就移植到工作组级;如今,数据库技术正在 Internet 中被广泛使用。

1.1.1 数据处理技术

1. 数据

数据(Data)是描述现实世界中各种具体事物或抽象概念的符号记录,除了常用的数字数据外,文字(如名称)、图形、图像、声音等信息也都是数据。在日常生活中,人们使用交流语言(如汉语)去描述事物;在计算机中,为了存储和处理这些事物,就要抽出人们对这些事物感兴趣的特征组成一个记录来描述。例如,在学生管理系统中可以将学生的学号、姓名、性别和年龄情况描述为“201601001,张强,男,18”。

2. 数据处理

数据处理(Data Process)是指对数据进行的收集、分类、组织、编码、存储、加工、计算、检索、维护、传播以及打印等一系列活动。数据处理的目的是从大量的数据中根据数据自身的规律和它们之间固有的联系,通过分析、归纳、推理等科学手段提取出有效的信息资源。

在数据处理中,通常数据的加工、计算等比较简单,而数据的管理比较复杂。数据管理是数据处理的核⼼,是指对数据的收集、分类、组织、编码、存储、检索、维护等操作,这部分操作是数据处理业务的基本环节,是任何数据处理业务中必不可少的共有部分,因此学习和掌握数据

管理技术能对数据处理提供有利的支持。

1.1.2 数据库技术的3个发展阶段

随着计算机硬件和软件的发展,数据库技术也在不断发展。从数据管理的角度而言,数据库技术经历了人工管理、文件系统和数据库系统3个阶段。

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,从硬件上看,外存只有磁带、卡片、纸带,没有磁盘等直接存取的存储设备;从软件上看,没有操作系统,没有管理数据的软件。

这个时期数据管理的特点是数据由计算或处理它的程序自行携带,数据和应用程序一一对应,应用程序依赖于数据的物理组织,数据的独立性差,数据不能被长期保存,数据的冗余度大,等等,给数据的维护带来许多问题。

人工管理阶段应用程序与数据之间的关系如图1-1所示。

2. 文件系统阶段

在20世纪50年代后期至60年代中后期,计算机的应用范围逐渐扩大,不仅用于科学计算,还大量用于管理。在硬件方面,磁盘成为计算机的主要外存储器;在软件方面,出现了高级语言和操作系统,出现了新的数据处理系统——文件系统,该系统把计算机中的数据组织成相互独立的数据文件,可以按照文件的名称对其进行访问,实现对文件中记录的查询、修改、插入和删除。文件系统实现了记录内的结构化,即给出了记录内各种数据间的关系。但是,文件从整体来看却是无结构的,其数据面向特定的应用程序,所以依然存在数据共享性、独立性差,冗余度大,管理和维护的代价大等缺点。

文件系统阶段应用程序与数据之间的关系如图1-2所示。

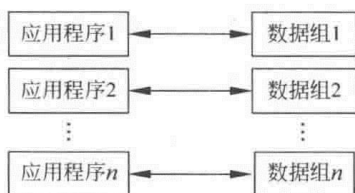


图 1-1 人工管理阶段应用程序与数据的对应关系

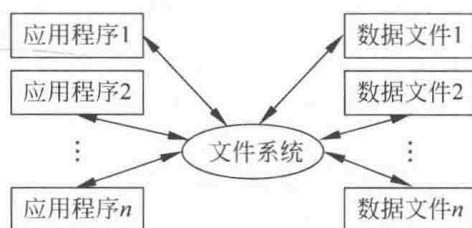


图 1-2 文件系统阶段应用程序与数据的对应关系

3. 数据库系统阶段

自20世纪60年代后期以来,计算机的应用越来越广泛,数据量急剧增加,而且数据的共享要求越来越高。计算机的硬件和软件都有了进一步的发展,在硬件方面,有了大容量的磁盘;在软件方面,传统的文件系统已经不能满足人们的需求,能够统一管理和共享数据的数据库管理系统(DataBase Management System, DBMS)应运而生。所以,此阶段将数据集中存储在一台计算机上(数据库中)进行统一组织和管理,从处理方式上讲,联机实时处理要求更多了,并开始提出和考虑分布处理。

数据库系统阶段的特点(也是优点)如下。

(1) 数据结构化:数据结构化是数据库系统与文件系统的根本区别。有了DBMS以后,数据库中的数据不再针对某一应用,而是面向整个应用系统,它是对整个组织的各种应用(包

括将来可能的应用)进行考虑后建立起来的总的数据结构。

(2) 较高的数据共享性: 数据共享是指允许多个用户同时存取数据而互不影响, 该特征正是数据库技术先进性的体现。数据库系统从整体角度描述数据, 数据不再面向某个应用而是面向整个系统, 因此数据可以被多个用户、多个应用共享使用。数据共享可以大大减少数据冗余, 节约存储空间。

(3) 较高的数据独立性: 所谓数据独立是指数据与应用程序的彼此独立, 它们之间不存在相互依赖的关系。应用程序不随数据存储结构的变化而变化, 简化了应用程序的编制和程序员的工作负担。

(4) 数据由 DBMS 统一管理和控制: 数据库的共享是并发的共享, 即多个用户可以同时存取数据库中的数据, 甚至可以同时存取数据库中的同一数据, 因此 DBMS 还必须提供数据的统一管理和控制功能。

DBMS 加入了安全保密机制, 可以防止数据被非法存取。DBMS 的数据完整性保护可以保障数据的正确性、有效性和相容性, 完整性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。当多个用户的并发进程同时存取、修改数据库时可能会发生相互干扰而得到错误的结果, 或使得数据库的完整性遭到破坏, 因此 DBMS 必须对多用户的并发操作加以控制和协调。另外, DBMS 还采取了一系列措施, 以实现数据库破坏后的恢复。

数据库系统阶段应用程序与数据之间的关系如图 1-3 所示。

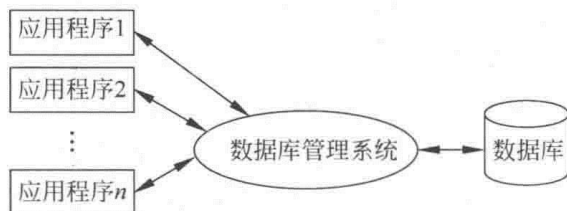


图 1-3 数据库系统阶段应用程序与数据的对应关系

1.1.3 数据库技术的新进展

数据库技术发展之快、应用之广是计算机科学其他领域的技术无可比拟的, 例如 CAD(计算机辅助设计)、CAM(计算机辅助制造)、CIMS(计算机集成制造系统)、CASE(计算机辅助软件工程)、OA(办公自动化)、GIS(地理信息系统)、MIS(管理信息系统)、KBS(知识库系统)等应用领域都需要数据库新技术的支持。

1. 分布式数据库

分布式数据库系统(Distributed DataBase System, DDBS)通常使用较小的计算机系统, 每台计算机可单独放在一个地方, 每台计算机中都有 DBMS 的一份完整副本, 并具有自己局部的数据库, 位于不同地点的许多计算机通过网络互相连接, 共同组成一个完整的、全局的大型数据库。

这种组织数据库的方法克服了物理中心数据库组织的弱点。首先, 降低了数据传送代价, 因为大多数对数据库的访问操作都是针对局部数据库的, 而不是对其他位置的数据库访问; 其次, 系统的可靠性提高了很多, 因为当网络出现故障时仍然允许对局部数据库的操作, 而且一个位置的故障不影响其他位置的处理工作, 只有当访问出现故障位置的数据时在某种程度上才受影响; 最后, 便于系统的扩充, 增加一个新的局部数据库或在某个位置扩充一台适当的

小型计算机都很容易实现,然而有些功能要付出较高的代价。例如,为了调配在几个位置上的活动,事务管理的性能比在中心数据库时花费更高,而且甚至抵消了许多其他的优点。

分布式数据库系统的主要特点如下。

(1) 数据是分布的:数据库中的数据分布在计算机网络的不同结点上,而不是集中在一个结点,区别于数据存放在服务器上由各用户共享的网络数据库系统。

(2) 数据是逻辑相关的:分布在不同结点的数据逻辑上属于同一个数据库系统,数据间存在相互关联,区别于由计算机网络连接的多个独立数据库系统。

(3) 结点的自治性:每个结点都有自己的计算机软/硬件资源、数据库、数据库管理系统(Local DataBase Management System, LDBMS, 局部数据库管理系统),因而能够独立地管理局部数据库。

2. 面向对象数据库

面向对象(Object-Oriented, OO)是一种认识方法学,也是一种新的程序设计方法学。把面向对象的方法和数据库技术结合起来可以使数据库系统的分析、设计最大程度地与人们对客观世界的认识相一致。面向对象数据库系统(Object-Oriented DataBase System, OODBS)是为了满足新的数据库应用需要而产生的新一代数据库系统。

面向对象数据库系统强调在数据库框架中发展类型、数据抽象、继承和持久性。它的基本设计思想是一方面把面向对象语言向数据库方向扩展,使应用程序能够存取并处理对象,另一方面扩展数据库系统,使其具有面向对象的特征,提供一种综合的语义数据建模概念集,以便对现实世界中复杂应用的实体和联系建模。因此,面向对象数据库系统首先是一个数据库系统,具备数据库系统的基本功能,其次是一个面向对象的系统,针对面向对象的程序设计语言的永久性对象的存储管理而设计,充分支持完整的面向对象概念和机制。

面向对象数据库将面向对象的能力赋予了数据库设计人员和数据库应用开发人员,从而扩展了数据库系统的应用领域,并能提高开发人员的工作效率和应用系统的质量。

3. 多媒体数据库

多媒体数据库系统(Multi-media DataBase System, MDBS)是数据库技术与多媒体技术相结合的产物。在许多数据库应用领域中都涉及大量的多媒体数据,它们与传统的数字、字符等格式化数据有很大的不同,都是一些结构复杂的对象。

因此,多媒体数据库不是对现有的数据进行界面上的包装,而是从多媒体数据与信息本身的特性出发,考虑将其引入到数据库中之后而带来的有关问题。多媒体数据库从本质上来说要解决3个难题,第一个是信息媒体的多样化,不仅仅是数值数据和字符数据,还要扩大到多媒体数据的存储、组织、使用和管理;第二要解决多媒体数据集成或表现集成,实现多媒体数据之间的交叉调用和融合,集成粒度越细,多媒体一体化表现越强,应用的价值才越大;第三是多媒体数据与人之间的交互性,没有交互性就没有多媒体,要改变传统数据库查询的被动性,能以多媒体方式主动表现。

从实际应用的角度考虑,多媒体数据库管理系统(MDBMS)应具有以下基本功能:

(1) 能够有效地表示多种媒体数据,对不同媒体的数据(如文本、图形、图像、声音等)能够按应用的不同采用不同的表示方法。

(2) 能够处理各种媒体数据,正确识别和表现各种媒体数据的特征以及各种媒体间的空间或时间关联。

(3) 能够像其他格式化数据那样对多媒体数据进行操作,包括对多媒体数据的浏览、查询

检索,对不同的媒体提供不同的操作,如声音的合成、图像的缩放等。

(4) 具有开放功能,提供多媒体数据库的应用程序接口等。

4. 数据仓库

数据仓库之父 Bill Inmon 在 1991 年出版的 *Building the Data Warehouse* 一书中所提出的定义被人们广泛接受——数据仓库(Data Warehouse)是一个面向主题的、集成的、相对稳定的、反映历史变化的数据集合,用于支持管理决策。

(1) 面向主题:操作型数据库的数据组织面向事务处理任务,各个业务系统之间各自分离,而数据仓库中的数据是按照一定的主题域进行组织的。

(2) 集成的:数据仓库中的数据是在对原有分散的数据源数据抽取、清理的基础上经过系统加工、汇总和整理得到的,必须消除数据源中的不一致性,以保证数据仓库内的信息是关于整个企业的一致全局信息。

(3) 相对稳定的:数据仓库的数据主要供企业决策分析之用,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,也就是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。

(4) 反映历史变化:数据仓库中的数据通常包含历史信息,系统记录了企业从过去某一时点(如开始应用数据仓库的时点)到目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。

企业数据仓库的建设是以现有企业业务系统和大量业务数据的积累为基础的。数据仓库不是静态的概念,只有把信息及时交给需要这些信息的使用者,供他们做出改善其业务经营的决策,信息才能发挥作用,信息才有意义。而把信息加以整理归纳和重组,并及时提供给相应的管理决策人员,是数据仓库的根本任务。

数据仓库的出现并不是要取代数据库。目前,大部分数据仓库还是用关系数据库管理系统来管理的。可以说,数据库、数据仓库相辅相成、各有千秋。

数据库是面向事务的设计,数据仓库是面向主题的设计;数据库一般存储在线交易数据,数据仓库一般存储历史数据;数据库在设计时尽量避免冗余,一般采用符合范式的规则来设计,数据仓库在设计时有意引入冗余,采用反范式的方式来设计;数据库是为捕获数据而设计的,数据仓库是为分析数据而设计的,它的两个基本元素是维表和事实表。

5. 数据挖掘

数据挖掘(Data Mining)又称为数据库中的知识发现(Knowledge Discovery in DataBase, KDD),就是从大量数据中获取有效的、新颖的、潜在有用的、最终可理解的模式的过程。简单地说,数据挖掘就是从大量数据中提取或“挖掘”知识。

并非所有的信息发现任务都被视为数据挖掘。例如使用数据库管理系统查找个别的记录,或通过因特网的搜索引擎查找特定的 Web 页面,这是信息检索(Information Retrieval)领域的任务。虽然这些任务是重要的,可能涉及使用复杂的算法和数据结构,但是它们主要依赖传统的计算机科学技术和数据的明显特征来创建索引结构,从而有效地组织和检索信息。

从数据本身来考虑,数据挖掘通常需要 8 个步骤。

(1) 信息收集:根据确定的数据分析对象抽象出在数据分析中所需要的特征信息,然后选择合适的信息收集方法,将收集到的信息存入数据库。对于海量数据而言,选择一个合适的数据存储和管理的数据仓库是至关重要的。

(2) 数据集成:把不同来源、格式、特点性质的数据在逻辑上或物理上有机地集中,从而