



“十二五”职业教育国家规划教材
经全国职业教育教材审定委员会审定



数据库技术与应用 项目化教程 ——SQL Server 2012

岳国英 主 编
李 鹏 胡文红 王 樱 副主编



中国电力出版社
CHINA ELECTRIC POWER PRESS



“十
经全

家规划教材
委员会审定

数据库技术与应用 项目化教程 ——SQL Server 2012

主 编 岳国英
副主编 李 鹏 胡文红 王 樱
编 写 唐 亮 陈忠文 王 军
关晓惠 酆丽华
主 审 杨文柱

常州大学图书馆
藏书章



中国电力出版社
CHINA ELECTRIC POWER PRESS

内 容 提 要

本书为“十二五”职业教育国家规划教材。

本书共分14章,主要包括“校园一卡通”项目导入、数据库技术基础知识、关系数据库设计、SQL Server 2012安装、利用Management Studio创建和管理数据库、关系数据库标准语言SQL、SQL Server 2012数据查询、Transact-SQL编程基础、存储过程和触发器、数据库完整性及SQL Server的完整性控制、数据库的安全性及SQL Server安全管理、数据库并发控制及SQL Server的并发控制机制、数据库恢复技术与SQL Server的数据恢复机制、课程设计——《网上书店图书销售管理系统》的数据库设计与实现等。各章之后附有实训、习题,附录中给出了常用的内置函数、系统表和系统存储过程,还给出了各章习题的参考答案。本书运用CDIO工程教育理念,选取应用广泛的“一卡通”类数据库系统的代表“校园一卡通”为案例,以项目为载体,将数据库的知识、技术与方法融为一体。本书的内容从教、学、做3个层面展开,技术线路清晰,以具体项目贯穿始末,易于学习,可操作性强,循序渐进且层次分明,便于读者掌握数据库实用技术。

本书可作为高等职业技术学院、应用型本科院校计算机及其相关专业数据库课程的教材,也可作为计算机应用系统开发与设计者学习数据库的参考用书。

图书在版编目(CIP)数据

数据库技术与应用项目化教程:SQL Server 2012/岳国英主编. —北京:中国电力出版社,2014.9

“十二五”职业教育国家规划教材

ISBN 978-7-5123-6041-9

I. ①数… II. ①岳… III. ①关系数据库系统—高等教育—教材 IV. ①TP311.138

中国版本图书馆CIP数据核字(2014)第173936号

中国电力出版社出版、发行

(北京市东城区北京站西街19号 100005 <http://www.cepp.sgcc.com.cn>)

北京丰源印刷厂印刷

各地新华书店经售

*

2014年9月第一版 2014年9月北京第一次印刷
787毫米×1092毫米 16开本 19.75印张 492千字
定价 39.00元

敬告读者

本书封底贴有防伪标签,刮开涂层可查询真伪
本书如有印装质量问题,我社发行部负责退换

版权专有 翻印必究

前言

数据库技术的发展使计算机应用领域拓展出一片崭新的、广阔的空间,移动互联网、云计算、物联网等新一代 IT 技术的出现更是将数据库技术的应用推向一个巅峰。无论是“智慧城市”建设,还是各行各业的信息建设或是风靡世界的电子商务,数据的组织和管理均是核心和基础。数据库技术在信息技术的应用中无所不在,已成为信息技术的核心与基础。因此,数据库技术是计算机及其相关专业的重要专业课程之一。

得益于编者主编的《SQL Server 2000 数据库技术实用教程》及其升级版“十一五”国家级规划教材《数据库技术与 SQL Server 2005》,读者给予的积极评价和支持,给了编者继续做好这本书编写的推动力。编者将本书的编写作为对老读者的一种回报、对新读者的一种关爱。相信本书会给新老读者带来裨益。

本书的编写原则:以用为本,学以致用,以使读者掌握实用技术为核心,以积累项目经验为目标。以项目为载体,按 CDIO (Conceive——构思, Design——设计, Implement——实现, Operate——运行操作)的工程线路,将数据库知识、技术与方法融为一体,体现 CDIO 工程教育理念。全书的学习过程,即是项目实施的过程。

本书的内容组织从教、学、做三个层面展开。各章首先提出学习目的、重点、难点和任务描述,然后按“项目教学”、“项目案例”、“实训”、“习题”的顺序展开。①“教”主要体现在“项目教学”部分。这个层面主要是介绍知识,教授技术,选用极具代表性的“校园一卡通”消费管理子项目为载体,以数据库设计、实现、管理、查询、编程为主线贯穿始终。②“学”主要体现在“项目案例”部分。这个层面以“校园一卡通”图书借阅管理子项目为载体,从实际业务逻辑的角度提出问题,通过分析给出解决方法和参考代码,与“教”的内容相呼应,形成“学”的主线,不仅使读者通过模仿并有所拓展地“学”一些解决实际问题的方法和技术,而且也让读者体验一个大的项目“整体规划,分块实施”的过程。③“做”是指各章后的“实训”部分。该部分与“教”和“学”的内容相随,对两个子项目进行完善,使读者对项目的实施得到更深一层的认识,并自己亲自操作完成。当本书学习结束时,读者已经具有了一定的项目经验。

本书选择 SQL Server 2012 版本不仅因为它是目前最流行、应用最广泛的关系数据库管理系统,更重要的是将 SQL Server 2012 作为实现数据库技术的工具,起到触类旁通的作用。版本并不重要,本书的大部分内容在 SQL Server 2005 或以后推出的 SQL Server 更高版本都适用。

本书由浙江水利水电学院的岳国英教授担任主编,河北工程技术高等专科学校的李鹏副教授、浙江水利水电学院的胡文红副教授、上海电子信息职业技术学院的王樱副教授担任副主编,参加编写的还有加拿大达内 IT 培训集团教学督导总监唐亮。第 1~3 章、第 10 章、第 11 章由岳国英编写,第 4 章、第 5 章和第 13 章由李鹏编写,第 8 章、第 9 章和第 14 章由胡文红编写,第 6 章和第 7 章由王樱编写,第 12 章和附录由唐亮编写,浙江水利水电学院陈忠文、王军、关晓惠老师以及浙江育英职业技术学院酆丽华老师也参加了本书编写。全书由岳国英修改并统稿。

本书由河北大学教学与计算机学院杨文柱教授担任主审。同时，本书在编写过程中还参考了许多专家学者的文献资料，在此一并致谢！

为方便读者学习，本书将为大家提供教学 PPT、用例数据库、习题参考答案等电子资源，读者还可访问编写团队主要成员胡文红老师主持的浙江水利水电学院精品课程《数据库原理及应用》的网站：<http://shujuku.zjweu.edu.cn>，有比较丰富的课程资源库。

限于编者水平，加之时间仓促，书中疏漏之处在所难免，敬请广大读者批评指正。也可通过电子邮件与编者联系，E-mail:yuegy@zjwchc.com。

编 者

2014 年 6 月

目 录

前言	
第1章 “校园一卡通”项目导入	1
1.1 本教程的学习目标	1
1.2 “校园一卡通”项目简介	1
1.3 各章内容的组织结构	3
第2章 数据库技术基础知识	6
2.1 数据库技术概述	6
2.2 数据模型	13
2.3 概念模型	15
2.4 数据库设计	19
2.5 项目案例“校园一卡通”图书借阅管理子项目概念模型设计	21
实训 “校园一卡通”活动考勤管理子项目概念模型设计	23
习题	23
第3章 关系数据库设计	26
3.1 关系数据模型	26
3.2 E-R 模型到关系模型的转换	32
3.3 关系模型的三类完整性约束	34
3.4 关系数据库的规范化设计	38
3.5 项目案例“校园一卡通”图书借阅管理子项目关系模型设计	41
实训 “校园一卡通”活动考勤管理子项目关系模型设计	43
习题	43
第4章 SQL Server 2012 安装	45
4.1 SQL Server 2012 简介	45
4.2 SQL Server 2012 的安装准备	46
4.3 SQL Server 2012 安装与检验	51
4.4 SQL Server 2012 的配置	62
4.5 SQL Server 2012 的组件	65
4.6 系统数据库和示例数据库	69
习题	71
第5章 利用 Management Studio 创建和管理数据库	72
5.1 创建和管理数据库	72

5.2 创建和管理表	78
5.3 建立表间关系	84
5.4 创建和管理视图	86
5.5 创建和管理索引	90
5.6 项目案例“校园一卡通”图书借阅管理子项目数据库及表的管理	93
实训 “校园一卡通”活动考勤管理子项目物理结构设计	98
习题	99
第6章 关系数据库标准语言 SQL	101
6.1 SQL 概述	101
6.2 数据定义	103
6.3 数据更新	109
6.4 数据查询	112
6.5 视图	118
6.6 项目案例“校园一卡通”图书借阅管理数据库的数据管理	120
实训 完善“校园一卡通”数据库的消费数据管理	123
习题	124
第7章 SQL Server 2012 数据查询	125
7.1 简单查询	125
7.2 连接查询	130
7.3 子查询	136
7.4 联合查询	139
7.5 项目案例“校园一卡通”图书借阅管理数据库的高级数据查询	142
实训 “校园一卡通”消费管理数据库的高级数据查询	146
习题	147
第8章 Transact-SQL 编程基础	149
8.1 基本知识	150
8.2 常用函数和自定义函数	160
8.3 批处理和流程控制语句	168
8.4 项目案例“校园一卡通”图书借阅管理子项目的借、还书流程处理	174
实训 “校园一卡通”消费管理子项目的业务处理	176
习题	177
第9章 存储过程和触发器	179
9.1 存储过程	180
9.2 触发器	192
9.3 游标及其应用	204
9.4 项目案例 利用存储过程和触发器实现“校园一卡通”图书 借阅管理子项目的借、还书流程处理	207

实训	利用触发器和存储过程实现“校园一卡通”消费管理子项目的业务流程处理	210
习题		210
第 10 章	数据库完整性及 SQL Server 的完整性控制	212
10.1	完整性约束条件及完整性控制	212
10.2	SQL Server 的数据完整性及其实现	213
10.3	项目案例 “校园一卡通” 图书借阅管理子项目数据库的完整性控制	220
实训	完善“校园一卡通” 图书借阅管理子项目数据库完整性控制	225
习题		226
第 11 章	数据库的安全性及 SQL Server 安全管理	227
11.1	数据库安全性控制的一般方法	227
11.2	SQL Server 的安全体系结构	231
11.3	SQL Server 数据库安全性管理	233
11.4	项目案例 设计“校园一卡通” 图书借阅管理子项目的安全性管理	250
实训	完善“校园一卡通” 数据库的安全性	252
习题		253
第 12 章	数据库并发控制及 SQL Server 的并发控制机制	254
12.1	事务及并发控制的基本概念	254
12.2	封锁机制	257
12.3	SQL Server 的并发控制机制	259
12.4	项目案例 事务在“校园一卡通” 消费管理中的应用	261
实训	完善“校园一卡通” 消费管理子项目的并发控制	266
习题		266
第 13 章	数据库恢复技术与 SQL Server 的数据恢复机制	268
13.1	故障的种类	268
13.2	数据恢复的实现技术	269
13.3	数据库恢复策略	271
13.4	SQL Server 的数据备份和还原策略	272
13.5	SQL Server 2012 的数据备份	274
13.6	SQL Server 2012 的数据恢复	278
13.7	数据的导入与导出	282
13.8	项目案例 “校园一卡通” 数据库的数据维护计划	286
实训	“校园一卡通” 数据库的数据备份与恢复	288
习题		289
第 14 章	课程设计——《网上书店图书销售管理系统》的数据库设计与实现	291
14.1	任务描述	291

14.2 设计要求	291
14.3 提交资料	294
附录	295
附录 A SQL Server 2012 的常用内置函数	295
附录 B SQL Server 2012 的常用系统表	298
附录 C SQL Server 2012 的常用系统存储过程	300
附录 D 习题和实训参考答案	303
参考文献	308

第1章

“校园一卡通”项目导入

随着泛在网络、云计算、物联网等技术的迅猛发展,新一轮信息化革命已经到来。目前,世界各信息技术发达的国家和地区,都在探索和实践利用新一代互联网、云计算、智能传感、通信、遥感、卫星定位、地理信息系统等技术,通过对一切物品的智能化识别、定位、跟踪、监控与管理,使国家和地区达到“智慧”的状态,以实现“智慧城市”、“智慧国家”的目标。无论是“智慧城市”建设,还是各行各业的信息化建设或电子商务等,数据的组织和管理均是核心和基础。数据库技术作为数据组织与管理的实现技术和方法,在信息化中的重要性越来越凸显。

1.1 本教程的学习目标

本教程的学习目标在于使读者掌握数据库的设计、实现、管理和维护的技术与方法。将SQL Server 2012作为实现这些技术与方法的一种工具。因此,在内容组织上强化数据库标准技术,包括代数运算、数据建模及其表现、关系数据库标准语言SQL。数据库技术的工程应用性较强,使读者通过学习具有一定的项目经验也是本教程的学习目标之一。因此,在教程内容组织过程中,以“校园一卡通”项目为载体,介绍数据库的知识、理论、技术和方法。该项目贯穿本教程的全程。读者完成本教程学习的同时,也具有了一定的项目经验。

1.2 “校园一卡通”项目简介

1.2.1 “一卡通”

所谓“一卡通”,就是在同一张卡上实现多种不同功能的智能管理。其核心内容是利用IC卡这种特定的物理媒介,实现从业务数据的生成、采集、传输到汇总分析等信息资源管理的规范化和自动化。

“一卡通”是基于“信息共享,集中控制”的需求形成的信息技术产物,它给人们的生活、工作带来极大便利,因此被广泛应用于各行各业。例如,众多企业采用“企业一卡通”,用于身份识别、考勤、就餐等。又如,很多城市采用“城市一卡通”,实现市民身份识别、公共设施使用(如公共自行车的借用、公园门禁、图书借阅等)许可、市内交通消费、市内就医等。再如,我国的铁路客运中,“居民身份证”在购票、取票、乘车过程中,实际上起到了“一卡通”的作用。越来越多的高校也在使用“校园一卡通”。

“一卡通”是信息管理中的一类应用。在这类应用中,虽然不同的应用管理目标和管理对象有所不同,但对数据的组织和管理是“异曲同工”。不失一般性,本教程的用例选用“校园一卡通”项目。通过该项目的学习和实践,读者对其他“一卡通”的项目也会从数据库层面有整体的领会和理解。

1.2.2 “校园一卡通”

所谓“校园一卡通”，简言之，就是使全校所有师生员工每人持一张校园卡，此卡可取代以前的各种证件（包括学生证、工作证、借书证、医疗证、出入证等）的功能。师生员工凭此卡出入学校、借阅图书、参加活动和消费等，并与银行卡实现自助圈存。“校园一卡通”从用户视角来看，方便了学生、教工的生活，减轻携带各种证件的烦恼。从管理者视角来看，可以动态掌握每一持卡人情况，分析和挖掘一些决策辅助信息，提高学校的管理水平。

“校园一卡通”的拓扑结构如图 1-1 所示。

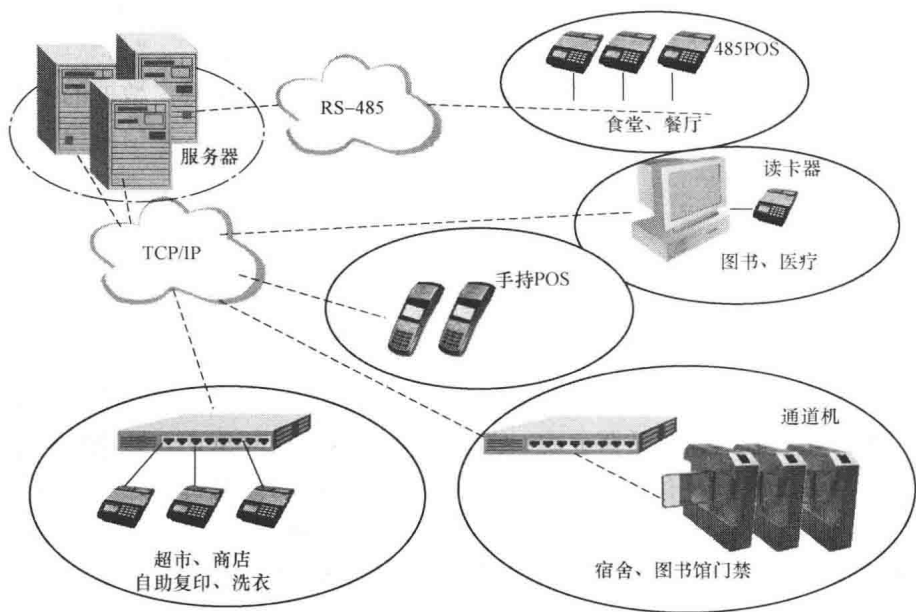


图 1-1 “校园一卡通” 拓扑结构图

1.2.3 案例简述

东南工程学院（虚构的学校名称）为了方便师生员工，提高管理现代化水平，决定实行“校园一卡通”，对校内一些工作和活动进行信息化管理。主要包括以下子项目。

1. 消费管理

该子项目的主要功能包括：持卡人（消费者）可以在校园的食堂、医院、商店通过 POS 机进行消费，也可以在充值点向校园卡充值。持卡人可以对消费、卡内余额等情况进行查询。管理人员可以对消费点的账务进行查询、汇总和划账。商家（收款单位）可以对隶属本家的 POS 机的交易情况进行查询和分析。

2. 图书借阅管理

该子项目的主要功能包括：持卡人（读者）可以通过读卡机进行身份识别，在规定数量内进行图书借阅、在规定日期内归还图书、查阅某本图书馆存情况、查询自己的借阅情况等。图书馆管理人员可以查询读者的借阅情况、超期未还情况，对图书借阅量、书目流通、馆藏等进行查询与统计。

3. 活动考勤管理

该子项目的主要功能包括：持卡人（活动参与者）对学校组织的需要考勤的活动采用刷卡的形式进行考勤，也可以在查询机上查询持卡人参加活动情况和积分情况等。管理人员可以维护各种活动，查询活动的参加情况，对个人、班级、院系参加活动情况进行查询分析。

4. 学生寝室管理

该子项目的主要功能包括：持卡人通过刷卡出入寝室，管理人员可以按个人、寝室、班级、院系查询学生住宿等情况。

5. 系统管理

该子项目的使用人员为“校园一卡通”管理员。主要功能包括卡务管理（新办卡、挂失、挂起、注销等）、商户管理、账务管理等。

“校园一卡通”功能结构简图如图 1-2 所示。

通过上述分析可知，“校园一卡通”是一个综合性比较强的项目，通过一张卡实现对校园内资源的使用（如图书借阅、寝室门禁、各类活动考勤）、消费（就餐、购物等）等的管理。该项目涉及大量的数据和多项管理功能，可以遵循整体规划，分块实施的原则，以简化和降低项目难度。

本教程以“校园一卡通消费管理”子项目为载体，进行数据库的知识、技术和方法的介绍，完成该部分数据库的设计、实现和应用。以“校园一卡通图书借阅管理”子项目为案例，对各章的知识、技术和方法进行综合应用。读者可以根据实际情况，自己完成其他子项目。特别提出的是，在各个子项目的设计与实施过程中，数据库设计与管理要从“校园一卡通”项目视角整体考虑。

为了丰富读者的项目经验，拓展读者解决问题的能力，本教程还安排了应用广泛的电子商务类型的项目“网上书店图书销售管理系统”作为课程设计，对数据库进行设计、实现和管理的体验。

考虑到本教程的读者定位是数据库初学者或有一定开发经验的工程人员，因此在数据库设计时，力求实体局部完整、属性精简，使初学者既能感知和领悟实体联系和属性设置的重要性，又不至于因为复杂的实体联系或因记忆繁多的属性投入太多精力，构造一种轻松的学习环境，工程人员也不会因此而影响阅读本教程的兴趣。

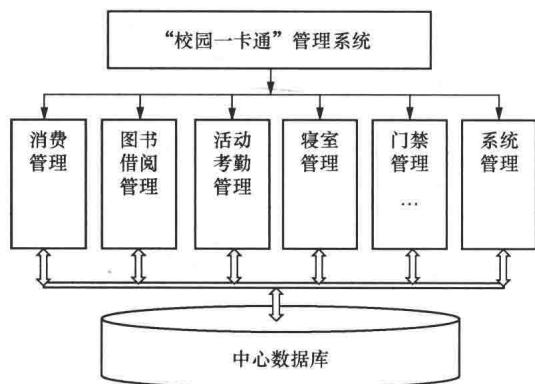


图 1-2 “校园一卡通”功能结构简图

1.3 各章内容的组织结构

这里介绍一下各章的内容组织方式，以使读者更轻松地学习本教程。各章将首先提出学习目的、学习重点、学习难点和任务描述。然后，按“项目教学”、“项目案例”、“实训”、“习题”的顺序展开。在本教程的基本内容之后，是一个留给读者的课程设计。

本教程的内容组织结构如图 1-3 所示。

1. 导入语

“导入语”的设置是使读者在学习本章的内容之前，明白该章要学什么、重点和难点是什

么、学习本章的知识和技术能够做什么。明确目标，有的放矢地去学去做，以达到更好的学习效果。

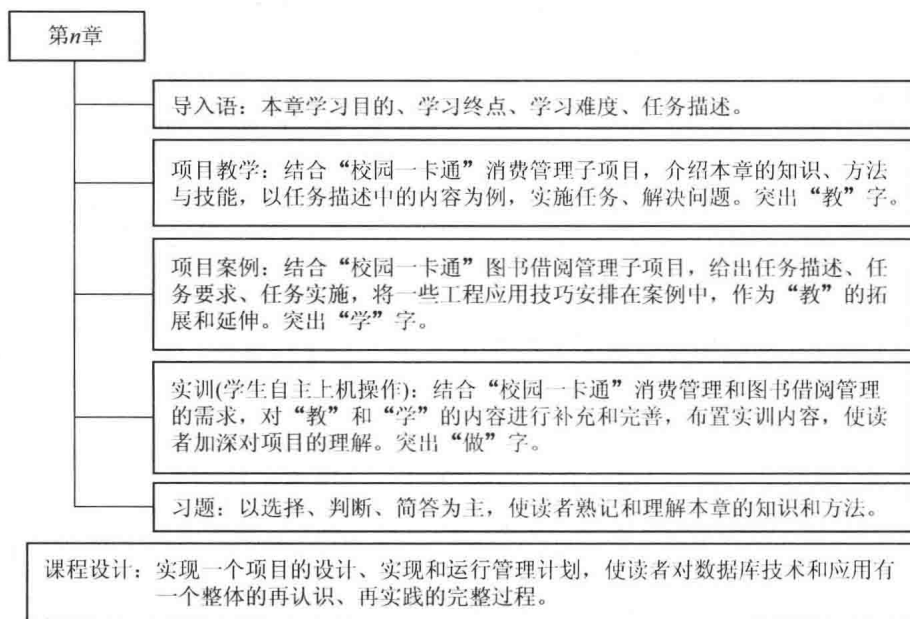


图 1-3 教程内容组织结构示意图

2. 项目教学

“项目教学”部分的内容在“教”的层面上，以“校园一卡通”消费管理子项目为载体，介绍数据库的知识、方法和技能，贯穿教材全程。该部分的内容组织以“教”的形式展开，“教”读者领会和理解利用数据库技术解决实际问题的方法，培养读者数据组织与管理的能力。

3. 项目案例

“项目案例”部分的内容在“学”的层面上，以“校园一卡通”图书借阅管理子项目为载体，锻炼和提高读者解决实际问题的能力，案例贯穿教材全程。该部分的内容组织以“学”的形式展开，让读者仿照“教”的内容，学着自己完成相应的任务。

4. 实训

“实训”是留给读者的作业。对“项目教学”和“项目案例”中“校园一卡通”消费管理和图书借阅管理两个子项目的需求进行补充和完善，将各章的知识、方法与技能以任务的形式布置实训内容，进一步理解和消化两个子项目，使读者在前面“教”和“学”的基础上，进行“做”的演练。

5. 习题

“习题”是根据本章学习的内容，将概念、知识与方法，以选择题、判断题和简答题的形式呈现给读者，通过作答加深读者的记忆、理解和领悟。

6. 课程设计

“课程设计”的安排，是使读者综合利用所学知识、方法和技能，自主完成电子商务类项目“网上书店图书销售管理系统”的需求分析、数据库设计、数据库实现、数据管

理、运行维护计划等，使读者对数据库技术及其应用有一个整体的体验，同时也拓展了项目经验。

本章的内容组织和设置，目的在于使读者对教程的读者定位、学习目标、内容组织结构、项目用例有一个了解，以利于后续章节的学习。

第2章

数据库技术基础知识

【学习目标】

- (1) 理解数据库基本概念——数据、数据库、数据库管理系统、数据库系统、大数据、云数据库等基本概念。
- (2) 了解数据库技术的发展过程。
- (3) 了解数据库的三级模式和二级映射。
- (4) 了解数据模型的组成和分类。
- (5) 理解和领会实体、实体型、实体集、属性、主码、域等术语。
- (6) 掌握实体的描述方法、正确确定实体之间的联系类型。
- (7) 熟练掌握用 E-R 模型描述概念模型的方法。

【重点】

- (1) 数据库技术的基本概念和术语。
- (2) 数据库系统的三级模式和二级映射。
- (3) 概念模型及其描述。

【难点】

- (1) 数据库系统的三级模式和二级映射。
- (2) 用 E-R 模型描述数据库的概念模型。

【任务】

针对东南工程学院的“校园一卡通”消费管理、图书借阅管理子项目，分析其实体集及其之间的联系，并用 E-R 图描述其概念模型。

2.1 数据库技术概述

2.1.1 数据库基本概念

1. 信息

信息 (Information) 是现实世界中事物的存在形式、运动形态、具体特征及其之间相互联系等要素在人脑中的反映，通过人脑抽象后形成的概念及描述。信息是人们在适应外部世界，控制外部世界的过程中，同外部世界交换的内容的统称。

2. 数据

数据 (Data) 是用物理符号对信息的表达和记录。或者说数据是载荷或记录信息的按一定规则排列组合的物理符号。物理符号有多种表现形式，包括数字、文字、图形、图像、声音及其他特殊符号。数据的各种表现形式，都可以经过数字化后存入计算机。

数据是信息的抽象和记录,信息是数据的解释和含义。

3. 数据库

数据库(DataBase, DB)是长期存储在计算机内、有组织的、可共享的数据集合。这种集合具有如下特点:

- (1) 最小的冗余度。以一定的数据模型来组织数据,避免出现不必要的重复数据。
- (2) 应用程序对数据资源共享。以最优方式为某个特定组织或企业的多种应用服务。
- (3) 数据独立性高。数据结构较强地独立于使用它的应用程序。
- (4) 统一管理和控制。对数据的定义、操纵和控制,由数据库管理系统进行统一管理和控制。

4. 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种数据管理软件,它的基本功能包括以下几个方面。

(1) 数据定义功能:DBMS 提供数据定义语言(Data Definition Language, DDL),使用 DDL 用户可以方便地创建数据库及其数据库中的对象。

(2) 数据操纵功能:DBMS 还提供数据操纵语言(Data Manipulation Language, DML),使用 DML 用户可以操纵数据,实现对 DB 中数据的基本操作,如查询、插入、删除和修改。

(3) 数据库的运行管理功能:数据库的建立、运行和维护由 DBMS 统一管理和控制,以保证数据的安全性、完整性、并发控制以及数据恢复等。

(4) 数据库的维护功能:它包括数据库初始数据的输入、转换功能、数据库的转储和恢复、数据库的重组、性能监视、性能分析等功能。

数据库管理系统软件有多种,比较著名的有 Oracle、Informix、Sybase、SQL Server、DB2 等。

5. 数据库系统

数据库系统(DataBase System, DBS)是指在计算机系统中引入数据库后构成的系统,一般由数据库、操作系统、数据库管理系统(及其开发工具)、应用程序、数据库管理员和用户构成。应当指出的是,数据库的建立、使用和维护等工作只有 DBMS 远远不够,还要有专门的人员来完成,这些人被称为数据库管理员(DataBase Administrator, DBA)。

6. 大数据与云计算

大数据是从英文 Big Data 翻译而来。大数据是一个比较抽象的概念,目前尚没有一个公认的定义^[1]。不同的定义基本上都是对大数据的特征进行阐述和归纳。在这些定义中,比较有代表性的是 3V 定义^[2]和维基(WiKi)百科的定义^[3]。

3V 定义认为,大数据要满足 3 个特点——规模性(volume)、多样性(variety)和高速性(velocity)。除此之外,还有尝试在 3V 的基础上增加一个特性的 4V 定义。但对第 4 个“V”的说法并不统一。例如,国际数据公司(International Data Corporation, IDC)认为大数据还应当具有价值(value),大数据的价值往往呈现稀疏性的特点。而 IBM 认为大数据必然具备真实性(veracity)。

WiKi 定义:大数据是指利用常用软件工具捕获、管理和处理数据所耗时间超过可容忍时间的数据集。该定义隐含了对大数据要采用非常用的技术与方法进行快速处理。

大数据的特点有四个层面,一是数据量巨大。这里的巨大是相对而言的,与相关技术的发展有关。目前大数据的量已从 TB 级别,跃升到 PB 级别。二是数据类型繁多。除了常见的

数据类型外,还有网络日志、视频、图片、地理位置信息等。三是价值密度低。以视频为例,连续不间断监控过程中,可能有用的数据仅仅有一两秒。四是快速处理。

云计算是从英文 Cloud Computing 翻译而来。目前,对云计算的定义比较有代表性的有维基百科的定义和美国国家标准与技术研究院(NIST)的定义。

WiKi 定义:云计算是一种通过 Internet 以服务的方式提供动态可伸缩的虚拟化资源的计算模式。

NIST 定义:云计算是一种按使用量付费的模式。这种模式提供可用的、便捷的、按需的网络访问,进入可配置的计算资源共享池(资源包括网络、服务器、存储、应用软件、服务),快速获得和使用这些资源,只需投入很少的管理工作,或服务供应商进行很少的交互。

云计算的概念已被大量运用到生产环境中。目前,IBM 的 Power8 芯片、Intel 公司的云计算整体解决方案及相应产品,已用于大数据与云计算领域。VMware 和 Apache Hadoop 云计算和大数据软件平台,以及国内的“阿里云”的应服务范围正日渐扩大。云计算与大数据已从概念阶段,逐步走向实际应用阶段。

7. 云数据库

目前,对于云数据库(CloudDB)的概念定义不尽相同,尚没有公认的定义。这里暂且定义云数据库是部署在云计算环境中的数据库。云数据库是在 SaaS (software as a service, 软件即服务)成为应用趋势的大背景下发展起来的云计算技术。

云数据库具有以下特性。

(1) 动态可扩展。理论上,云数据库具有无限可扩展性,可以满足不断增加的数据存储需求。在面对不断变化的条件时,云数据库可以表现出很好的弹性。例如,对于一个从事产品零售的电子商务公司,会存在季节性或突发性的产品需求变化。像天猫商城、Animoto (国外一个视频、图像、音乐剪辑网站)这样的网站,可能会经历一个指数级的增长阶段。这时,云数据库可以分配额外的数据库存储资源来处理增加的需求,这个过程只需要几分钟。一旦需求过去之后,就可以立即释放这些资源。

(2) 高可用性。云数据库不存在单点失效问题。如果一个结点失效了,剩余的结点就会接管未完成的事务。在云数据库中,数据通常是复制的,在地理上也是分布的。如 Google、Amazon 和 IBM 等大型云计算供应商具有分布在世界范围内的数据中心,通过在不同地理区间内进行数据复制,可以提供高水平的容错能力。例如,Amazon SimpleDB 会在不同的区间内进行数据复制。因此,即使整个区域内的云设施发生失效,也能保证数据继续可用。

(3) 较低的使用代价。云数据库通常采用多租户(multi-tenancy)的形式,这种共享资源的形式对于用户而言可以节省开销。而且用户采用按需付费的方式使用云计算环境中的各种软、硬件资源,不会产生不必要的资源浪费。另外,云数据库底层存储通常采用大量廉价的商业服务器,这也大幅度降低了用户开销。

(4) 易用性。使用云数据库的用户不必控制运行原始数据库的机器,也不必了解它身在何处。用户只需要一个有效的链接字符串就可以使用云数据库。

(5) 大规模并行处理。云数据库支持几乎实时(不支持严格实时)的面向用户的应用、科学应用和新类型的商务解决方案。

就目前阶段而言,虽然一些云数据库产品,如 Google 的 BigTable、Amazon 的 SimpleDB 开源的 Hypertable 和 Hadoop 等,在一定程度上实现了对于海量数据的管理,但是这些系统