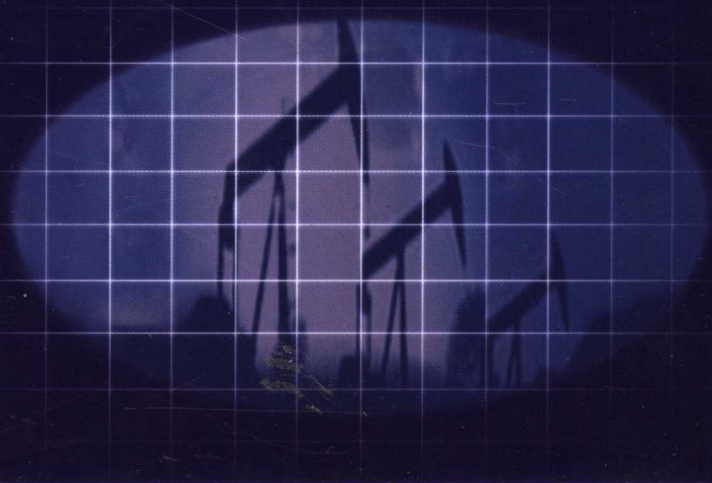


数据库新技术

在石油勘探中的应用

常冠华 熊华平 马玉书 等 编著



科学出版社

www.sciencep.com

数据库新技术在石油 勘探中的应用

常冠华 熊华平 马玉书 等 编著
郭万奎 金成志 主审

科学出版社

北 京

内 容 简 介

本书以大型企业信息化系统建设为例,较全面、系统地介绍了数据库新技术的应用,内容包括面向对象数据库、Web 数据库、主数据库(数据银行)、联机分析处理、数据仓库及数据挖掘技术等。书中介绍了这些新技术的基本概念、体系结构、实现方法与技术,深入分析和研究了这些新技术的实施过程与开发经验,列举了应用实例并指出了未来的发展趋势。

参加本书编写的均为长期从事数据库系统建设和大型 IT 项目开发的高级研究人员,具有开发数据库应用技术的丰富经验。在本书的编写过程中,作者参考了国内外的最新资料,并将从事大型 IT 项目的研究成果进行了分析总结,本书内容具有先进性、实用性和可读性。

本书可作为高等院校高年级本科生和研究生的教学参考书,也可供企事业单位从事数据库系统建设和信息系统开发的科技、工程、管理人员使用和参考,还可作为数据库培训班和数据库高级研讨班的教材。

图书在版编目(CIP)数据

数据库新技术在石油勘探中的应用 / 常冠华,熊华平,马玉书等编著.
—北京:科学出版社,2005
ISBN 7-03-015983-7

I. 数… II. ①常…②熊…③马… III. 数据库系统-应用-油气勘探
IV. P618.130.8-39

中国版本图书馆 CIP 数据核字(2005)第 083022 号

责任编辑:鞠丽娜 刘亚军 / 责任校对:柏连海
责任印制:吕春珉 / 封面设计:郝希平

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

铭浩彩色印装有限公司印刷

科学出版社发行 各地新华书店经销

*

2005 年 11 月第 一 版 开本:787×1092 1/16

2005 年 11 月第一次印刷 印张:16 1/2

印数:1—2 000 字数:376 000

定价:28.00 元

(如有印装质量问题,我社负责调换(环伟))

销售部电话 010-62136131 编辑部电话 010-62138978-8002 (B106)

前 言

数据库技术作为计算机科学的一个重要分支,一直是信息技术发展最活跃的领域,在计算机应用和企业信息化建设中发挥着重要的作用。

石油工业是一个高技术、高资金密集的产业。在石油勘探、生产的过程中会产生大量的数据,这些数据同地下的油气藏资源一样是国家的宝贵财富。大庆油田在勘探开发的过程中,一直重视数据库系统的建设。经过 20 多年的努力,已建立了庞大的数据库和超大型企业信息系统。随着计算机网络和数据库技术的进步与发展,近年来相继出现了 Web 数据库、联机分析处理、数据仓库、数据挖掘等数据库新技术,以及以主数据库(数据银行)为核心的信息化系统建设的整体开发方案,使我们有可能利用上述最新技术对石油勘探、生产中产生的大量数据进行有效的存储、组织、管理和传输,从而为油田勘探开发的正确决策、科学制定油气勘探开发方案和油田建设规划提供有效的解决方案和实现途径,以获取最大的经济利益。

本书较全面、系统地介绍了数据库新技术在油气勘探领域的应用及最新进展,内容包括面向对象数据库、Web 数据库、主数据库(数据银行)、联机分析处理、数据仓库、数据挖掘等技术,并深入分析和研究了勘探数据的流程、专业数据库、勘探主数据库、项目数据库等实施数据集成和专业应用软件集成的技术,列举了应用实例,指出了数据库技术的新应用及其未来发展。

本书共分 10 章。第 1~3 章讲述数据库技术的基本概念,包括关系数据库、面向对象数据库和 Web 数据库等;第 4 章描述油气勘探的业务、数据和管理流程;第 5 章讲述专业数据库的分类及其设计方法;第 6 章讲述主数据库的实现技术;第 7 章讲述项目数据库的设计及软件的集成方法;第 8 章讲述数据仓库与联机分析处理技术;第 9 章讲述数据挖掘及其实现方法;第 10 章讲述数据库技术的最新应用领域及其未来发展。

本书由多名长期从事数据库工作的专业技术人员编写。第 1、3 章由常冠华、马玉书编写;第 2 章由刘万伟、马玉书编写;第 4 章 4.1、4.2、4.3 节分别由丁淑霞、郭志勇、仲维鹏编写;第 5 章由闫海燕编写;第 6 章由王洪礼编写;第 7 章由燕汉业编写;第 8、9 章由熊华平、马玉书编写;第 10 章 10.1、10.2、10.5、10.6 节分别由丁淑霞、孙秀凤、王洪礼、崔敬伟编写,10.3、10.4 节由燕汉业编写。全书由郭万奎和金成志主审。

本书的出版得到了原大庆油田有限责任公司副总经理侯启军的指导;张铁刚、厉玉乐、韩殿杰、许显志、贺立强、岳兴举、李明等领导和专家为本书的编写提供了宝贵素材和修改意见,在此向他们表示诚挚的感谢。

由于作者水平有限,书中错漏和不妥之处在所难免,敬请读者批评指正。

目 录

第 1 章 数据库管理技术	1
1.1 数据库管理技术及其发展	1
1.1.1 数据库技术的发展	1
1.1.2 数据库系统的有关概念	2
1.2 数据模型	6
1.2.1 有关数据模型的一些基本概念	6
1.2.2 数据模型及其研究	9
1.2.3 关系数据模型	11
1.3 数据库技术的新进展和新一代数据库	17
1.3.1 传统数据库技术的局限性	17
1.3.2 当前数据库技术的发展途径	18
1.4 石油勘探数据库管理技术及其发展	20
1.4.1 石油勘探领域常用的数据库管理系统	20
1.4.2 油气勘探数据的标准化	22
1.4.3 石油勘探数据银行及勘探信息化服务体系	23
第 2 章 面向对象程序设计方法与面向对象数据库	27
2.1 面向对象程序设计方法学与面向对象数据库	27
2.1.1 面向对象程序设计的有关概念和特点	27
2.1.2 面向对象程序设计方法的研究与应用	31
2.1.3 面向对象数据库系统与面向对象数据库管理系统	31
2.1.4 面向对象数据库系统的标准化问题	33
2.2 面向对象建模技术	34
2.2.1 面向对象建模技术	34
2.2.2 面向对象的知识源模型	35
2.3 统一建模语言及其面向对象方法的描述	40
2.3.1 UML 在建模过程中采用的建模元素及图的表示方法	40
2.3.2 一个 UML 建模实例	42
第 3 章 Web 数据库技术	45
3.1 Web 数据库技术	45
3.1.1 Web 数据库技术概述	45
3.1.2 基于 Web 的数据库系统体系结构	45
3.1.3 基于中间件的软件体系结构	49
3.1.4 Web 与数据库的集成方式	50
3.1.5 Web 服务器	51

3.2 Web 数据库访问技术	52
3.2.1 Web 数据库访问技术概述	52
3.2.2 Web 数据库访问技术分类	53
3.2.3 几种常用的 Web 数据库接口技术与组件工具	54
3.3 组件对象模型技术及组件工具产品	66
3.3.1 面向对象的组件技术	67
3.3.2 组件对象模型技术	67
3.3.3 组件混合应用模型与组件技术的联合运用	75
3.4 XML 与数据库	77
3.4.1 XML 及其产生背景	77
3.4.2 XML 的定义	78
3.4.3 XML 应用的分类	79
3.4.4 XML 的特性	79
3.4.5 XML 的应用前景	80
第4章 石油勘探的业务、数据与管理流程	82
4.1 石油勘探业务及数据流程	82
4.1.1 物探工程	83
4.1.2 井筒工程	86
4.1.3 实验室分析化验	90
4.1.4 地质综合研究	92
4.2 石油勘探数据管理	94
4.2.1 数据整理	95
4.2.2 数据汇交归档	98
4.3 勘探动态数据管理应用实例	99
第5章 专业数据库	101
5.1 专业数据库的分类	101
5.1.1 物探专业数据库	101
5.1.2 钻井专业数据库	102
5.1.3 录井专业数据库	103
5.1.4 测井专业数据库	103
5.1.5 试油测试专业数据库	103
5.1.6 油气层改造专业数据库	103
5.1.7 分析化验专业数据库	104
5.2 专业数据库的设计	104
5.2.1 专业数据库的设计特点	104
5.2.2 专业数据库的设计规范	104
5.2.3 专业数据库的建设目标和功能特点	105
5.3 录井专业数据库设计实例	107
5.3.1 需求分析	107
5.3.2 录井专业数据库的总体方案	109
5.3.3 技术方案	110

第 6 章 主数据库	117
6.1 有关主数据库的概念	117
6.1.1 主数据库技术产生的背景	117
6.1.2 主数据库及其特点	117
6.2 主数据库系统的功能及配置	118
6.2.1 主数据库系统的基本功能	118
6.2.2 主数据库系统的基本配置	119
6.2.3 主数据库平台软件集成技术	121
6.3 主数据库的实施	124
6.3.1 主数据库的实施策略	124
6.3.2 主数据库实施的基本过程	125
6.4 主数据库的管理与应用支持	130
6.4.1 用户管理	131
6.4.2 数据管理	131
6.4.3 主数据库的应用支持	132
第 7 章 项目数据库	134
7.1 项目数据库的概念	134
7.1.1 基于区域数据的项目数据库	135
7.1.2 基于应用软件的项目数据库	135
7.1.3 项目数据库系统的功能特点	135
7.2 项目数据库的规划设计与建立	136
7.2.1 需求分析与概念设计	136
7.2.2 物理规划设计与实现	137
7.3 项目数据库与应用软件的集成	138
7.3.1 一体化的解决方案和途径	139
7.3.2 一体化专业软件的集成平台	140
7.3.3 应用软件的连接	142
7.4 一个项目数据库的实例	144
7.4.1 总体方案	145
7.4.2 数据库系统的物理模式规划	146
7.4.3 项目数据库的管理	149
7.4.4 查询统计分析软件	151
第 8 章 数据仓库与联机分析处理	153
8.1 数据仓库	153
8.1.1 数据仓库的产生背景	153
8.1.2 数据仓库的定义及其特点	154
8.1.3 数据仓库的组成与体系结构	158
8.1.4 数据仓库的数据库模式(数据仓库的多维分析)	168
8.1.5 数据仓库的设计与实现技术	170
8.1.6 数据仓库在油田勘探开发中的应用实例	176

8.2 联机分析处理	179
8.2.1 有关联机分析处理的概念	179
8.2.2 多维数据分析实例	183
8.2.3 多维数据模型上的 OLAP 操作	185
8.2.4 数据仓库查询语言	187
8.2.5 联机分析处理的实现方法与 Oracle 联机分析处理工具	192
第 9 章 数据挖掘技术	196
9.1 数据挖掘技术的概念和特点	196
9.1.1 什么是数据挖掘	196
9.1.2 数据挖掘产生的背景	197
9.1.3 数据挖掘的定义	197
9.1.4 数据挖掘的特点	198
9.2 数据挖掘算法的研究	200
9.2.1 数据挖掘算法研究的特点与分类	200
9.2.2 数据挖掘算法研究的内容	201
9.2.3 关联分析	205
9.2.4 聚类分析	210
9.2.5 基于神经网络的数据挖掘算法	215
9.3 数据挖掘的实现方法与应用	220
9.3.1 数据挖掘/知识发现的基本过程	220
9.3.2 数据挖掘的应用领域与数据挖掘工具的研究	222
9.3.3 数据挖掘技术在剩余油分布研究与潜力预测中的应用	223
第 10 章 数据库技术的应用及其未来发展	227
10.1 分销电子商务系统	227
10.1.1 电子商务	227
10.1.2 分销电子商务系统	227
10.2 电子政务系统	231
10.2.1 电子政务网站的建设与功能	231
10.2.2 电子政务应用系统的开发工具与技术	232
10.2.3 一个基于 Web 数据库的电子政务系统	232
10.3 数据库技术在企业资源计划中的应用	234
10.3.1 企业资源计划及其发展	234
10.3.2 企业资源计划	235
10.4 客户关系管理	238
10.4.1 客户关系管理	238
10.4.2 一个客户关系管理的体系结构	239
10.4.3 客户关系管理在销售商和银行金融行业中的应用	240
10.5 Web 数据挖掘技术	242
10.5.1 Web 数据挖掘技术概述	242
10.5.2 Web 数据挖掘技术的分类	242
10.5.3 Web 数据挖掘的实现步骤	244

10.5.4 Web 数据挖掘的应用	244
10.6 开放分布式环境下多智能体协同工作系统	245
10.6.1 有关多 Agent 系统的概念	245
10.6.2 Agent 与多 Agent	246
10.6.3 多 Agent 系统在石油勘探领域的应用前景及应用实例	249
主要参考文献	254

第 1 章 数据库管理技术

1.1 数据库管理技术及其发展

数据库技术是数据管理的最新技术,是研究用电子计算机对大量数据进行管理,实现数据共享的一门高技术学科,它是计算机科学的一个重要分支。数据库技术是计算机科学技术中发展最快、应用最广的技术。数据库技术的发展,已经成为先进信息技术的重要组成部分,是现代计算机信息系统和计算机应用系统的基础和核心。

1.1.1 数据库技术的发展

自从世界上第一台电子计算机问世以来,人们就开始利用计算机来进行各种数据处理。所谓数据处理(data processing)指的是对数据进行收集、组织、加工、储存、抽取和传播的过程。数据处理的目的是从大量的、凌乱的、难以理解的数据中,获得对某个特定的应用领域来说是有价值、有意义的信息,作为管理决策的依据。例如,气象预报中的数据处理和石油地震勘探资料的数据处理,都是先采集原始数据,然后按预先设计的数学模型对这些数据进行计算、处理,再把得到的结果提供给专业人员作为预报或预测的依据。从利用算盘、手摇计算机的手工数据处理的初级阶段,到目前基于数据库技术的数据处理和数据管理的新阶段,数据处理已经历了漫长的发展历程。

根据数据模型的演变,可以将数据库技术的发展大致划分为三个阶段。

(1) 第一代数据库系统

第一代数据库的代表是 1969 年 IBM 公司研制的基于层次模型的数据库管理系统 IMS 和 20 世纪 70 年代美国数据库系统语言协会 CODASYL 下属数据库任务组 DBTG 提议的网状模型。层次数据库的数据模型是有根的定向有序树;网状模型对应的是有向图。这两种数据模型统称为格式化的数据模型,它们奠定了现代数据库发展的基础。这两种数据模型的特点是:支持三级模式(外模式、模式、内模式),保证数据库系统具有数据与程序的物理独立性和一定的逻辑独立性;用存取路径来表示数据之间的联系;有独立的数据定义语言和导航式的数据操纵语言。但这类数据模型的缺点是使用者必须详细地了解 and 描述数据的存储结构和存取路径,进行数据库设计和开发数据库系统要凭经验。此外,数据库应用程序的可移植性和数据的独立性都较差。

(2) 第二代数据库系统

第二代数据库系统的主要特征是支持关系数据模型(数据结构、关系操作、数据完整性)。第二代数据库系统以 E. F. Codd 的关系数据模型和关系数据库理论为基础,开创了数据库关系方法和关系数据库理论研究,并给出了人们一致接受的关系模型的规范说明。关系模型具有以下特点:一是关系模型的概念单一,实体和实体之间的联系用关系来表示;二是以关系数学为基础,开展了关系方法和关系数据库理论研究(关系代数、关系演算、关系数据语言、SQL、QBE 等);三是数据的物理存储和存取路径对用户不透明;四是关系数据库语言是非过程化的,这些描述性语言一改以往的程序设计语言和层次网状数据库系

统中数据库语言的风格,而得到最终用户的喜爱,并为以后的数据库语言标准化奠定了基础。经过多年的努力已研制了大量的关系数据库管理系统 RDBMS,解决了查询优化、并发控制、故障恢复等多项技术问题。这不仅丰富了 DBMS 实现技术和数据库理论研究,更重要的是促进了 RDBMS 产品的蓬勃发展和广泛应用。目前,绝大多数数据库管理系统都是 RDBMS 产品,其主要的产品包括:IBM 公司的 DB2,CAI 公司的 Ingres II,Informix Software 公司的 Informix Dynamic Server,Microsoft 公司的 Microsoft SQL Server,Oracle 公司的 Oracle 9i,Sybase 公司的 Sybase Adaptive Server。

关系数据库系统最大的优点是使用非过程的数据操纵语言并有高度的数据独立性,但关系数据库系统还存在效率低,不适于许多应用领域等缺点。在处理复杂对象数据方面,关系数据模型还显得力不从心。

(3) 第三代数据库系统

自 20 世纪 80~90 年代以来,数据库技术在商业领域的巨大成功刺激了其他领域对数据库技术需求的迅速增长。这些新的领域(如 CAD/CAM、CIM、CASE、图像处理、办公自动化系统、地理信息系统、实时控制系统等)为数据库技术的发展和應用开辟了新的天地。随着数据库技术及计算机网络技术的飞速发展(如 Web 数据库等),原有的或新的行业领域对数据库技术又提出了更多、更高和更新的需求(如电子商务、电子政务、ERP、CRM、数据仓库、数据挖掘等)。此外,数据库处理的对象也越来越复杂(如图形图像、视频数据、空间数据、时态数据、多媒体、超文本、知识、语义信息检索等)。面对这些复杂的应用需求,关系型数据库已经不能完全满足,于是产生了以面向对象数据库系统为代表的新一代数据库系统或称之为第三代数据库系统。第三代数据库系统主要有以下特征:一是支持数据管理、对象管理和知识管理;二是保持和继承了第二代数据库系统的技术;三是对其其他系统开放,支持数据库语言标准,支持标准网络协议,有良好的可移植性、可连接性、可扩展性和互操作性等。第三代数据库支持多种数据模型(比如关系模型和面向对象模型等),并能与诸多新技术相结合,应用于多个领域,由此也衍生出多种数据库新技术。

1.1.2 数据库系统的有关概念

本节简略介绍数据、数据库、数据库系统和数据库管理系统的有关基本概念。

1. 数据与信息

(1) 数据

数据(data)泛指计算机能处理的各种事实,是数字、字符等各种符号的集合。数据是信息系统的基本概念,是计算机系统处理的基本对象。通常,任何一种符号、数字、文字、图形、图像都可以用来反映客观事物的性质、属性及其相互关系,这就是所谓的数据。例如,形式符号 5、101、五、five、cinq 等都可以用来表示数字 5。数据本身并没有什么意义,不代表任何一类具体的东西,它仅仅是一种抽象的量的概念,用不同的符号来物理地表示它,丝毫不会改变其抽象的含义。

严格地讲,数据可定义为:对客观事物记录下来的、可以鉴别的符号。数据是客观事物的属性、数量、位置及相互关系等的抽象表示。

数据作为一种符号串,是计算机软件领域的专门术语,它是软件中程序加工的原料和结果。在计算机软件中,只有有了程序和数据才构成完整的运行实体,亦即软件是由程序和数据这两部分所组成的。

(2) 数据的特性

在计算机软件中,数据具有许多重要的特性。

1) 数据表现形式的多样性:数据的表现形式多种多样,除常用的符号、数字、文字外,还包括图形、图像、语言、音乐、视频等多媒体数据,以及知识、规则、数学符号和推理策略等抽象数据,它们都可以用来反映客观事物的性质、属性及其相互关系,这就是所谓的数据。数据表现的多样性为数据的应用提供了保障。

2) 数据的可构造性:数据在计算机中有结构化、半结构化及非结构化三种存储形式。而数据库中的数据是具有一定结构的。首先,表示数据的属性有型(type)和值(value)之分,数据的型给出了数据表示的类型,如数字型(又分为整数型、实数型等)、字符型、日期型、逻辑型等;而数据的值给出了符合某给定型的数值。其次是数据的长度,如总长几位,小数点后几位等。对复杂对象(如空间数据、视频图像等)一类的数据,还有另外一些属性去描述它。随着应用需求的不断扩大,数据的属性有了进一步的扩大,它包括了将多种相关数据以一定的结构方式组合构成特定的数据框架,这种数据框架就称为数据结构(data structure)或(在特定条件下的)数据模式(data schema)。

3) 数据的挥发性和持久性:计算机中的数据分为两部分,一部分数据与程序紧密相关,只能与程序短时间交互,随着程序运行结束而消亡,这种数据一般只存放在内存中,称为临时性数据或挥发性数据(transient data);而另一部分数据则对系统起着长期持久的作用,它们被称为持久性数据(persistent data),这些数据作为数据库的内容一般存放在计算机的次级存储器(如磁盘)内。

4) 数据的私有性与共享性:数据按使用对象不同,分为私有数据和共享数据两种。为特定应用(程序)服务的数据称为私有数据(private data);而为多个应用(程序)服务的数据称为共享数据(share data)。

5) 数据“量”的表示:数据有量的区别。数据的量是衡量与区别数据的重要标志。随着数据处理技术的发展和数据库应用领域的扩大,数据的量由小量、大量,发展到海量。数据量的变化将引起数据“质”的变化。这些变化主要表现在:①数据的处理量已达到海量级别;②数据结构由简单变为复杂;③数据的服务范围由(单用户)私有发展到(多用户)共享;④数据在软件中由附属地位成为主导地位,数据这种根本性的变化完全是在数据库系统出现后产生的。这是因为:过去在软件系统中是以程序为主体,而数据则以私有及挥发性形式从属于程序。此时数据在程序中是分散、凌乱的。这造成数据管理的混乱,易产生数据冗余、数据不一致性和数据安全性差等弊病。

近年来,随着数据库技术的应用,已使数据在应用程序中占据主体地位。与此相反,程序已退居附属地位。在此种结构中需要对数据做集中、统一的管理,并使其能为多个应用程序所共享,其关系结构如图 1.1 所示。

本书所介绍和讨论的数据库应用系统的内容是以对海量的、复杂结构的、持久性和共享的数据的统一管理为目标,如不特别说明,所提及的数据即具有以上这 4 种特性。

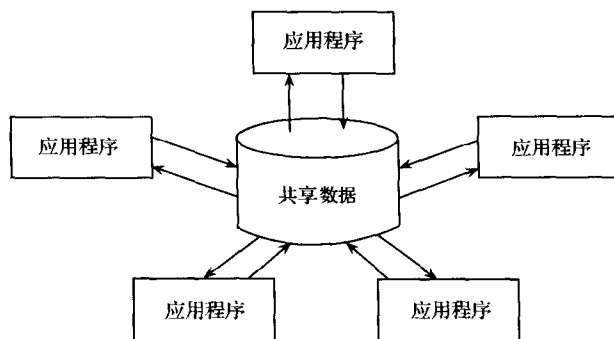


图 1.1 以数据为主体的应用程序系统

(3) 信息

前面我们介绍了数据,那么数据与信息又有什么不同呢?其实,信息与数据是两个既有联系,又有区别的概念。信息(information)是客观世界的一切真知,它向人们(或计算机)提供有关现实世界新的事实的知识。而数据是信息的具体物理表示,它是载荷信息的各种物理符号。数据经过处理、组织并赋予一定的意义后即可成为信息。信息和数据是不可分离而又有一定区别的概念。一方面并非任何数据都能表示为信息,信息只是消化了数据;另一方面信息是更基本的直接反映现实世界的概念,而数据则是信息的具体表现。信息不随载荷它的物理设备的改变而改变,而数据则不然。在实际应用中,有时对信息和数据这两个术语,并不严格加以区分,如信息处理又常称为数据处理。通常,我们在对数据库中的数据进行处理时,最关心的是要从大量的数据中获取有用的信息(这些信息可以是一些规律或知识),以此作为决策的依据。

2. 数据库

广义地讲,数据库(data base)是任何数据的集合。书库里的书,档案馆中的资料,油田的地质勘探资料,磁盘、磁带上的数据和程序、歌曲、录像,甚至山洞中的史前壁画等都可以被看成是数据库。不过我们这里所讨论的数据库含义并没有这样广泛。我们讲的数据库通常是指在计算机中用于数据处理的一种数据管理技术。简单地说,数据库是指为满足某一或多个部门中多个用户的多种应用需要,在计算机存储设备上,按一定组织方式存储在一起的、相互有关的全部数据的集合。这些数据集合具有如下的特点。

- 1) 数据尽可能不重复(最小冗余)。
- 2) 以最优的方式服务于一个或多个用户或应用程序(应用程序对数据资源的共享)。
- 3) 数据的存放尽可能地独立于使用它的应用程序(数据的独立性)。

4) 用一个系统软件——数据库管理系统(database management system, DBMS)统一管理和维护这些数据,在增加新的数据或更新、检索数据库中现有数据时,使用了通用的控制方法,并确保数据库中数据的安全性和完整性。

3. 数据库系统

数据库系统(data base system)是一个计算机系统,它不是指数据库本身,也不是指数据库管理系统,而是指计算机系统中引进数据库后系统的构成。通常,数据库系统可视为由人(各种用户、数据库管理员)、计算机系统、数据库管理系统、数据库及其描述机构等组成的一个高效能的信息处理系统。

数据库系统具有如下特点。

1) 数据结构化:数据库系统具有表达复杂数据结构的能力,数据不再针对某个应用,而是面向整个系统及整个机构来组织数据,并且数据是结构化的,存取方式也十分灵活。

2) 数据共享:数据库系统从整体描述数据,数据不再面向某一应用,并且数据可被多个用户及应用程序共享。数据共享可减少数据冗余、保持数据的一致性。

3) 数据的独立性:指数据的物理独立性和逻辑独立性。前者指用户的应用程序与存储在数据库中的数据是相互独立的;后者指用户的应用程序与数据库的逻辑结构是相互独立的。这样,当数据的物理结构或逻辑结构改变了,应用程序也不需要改变;反之亦然。

4) 数据统一由 DBMS 管理和控制:数据库的共享是并发的共享,即多个用户的应用程序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个数据。因此,数据库系统中的 DBMS 应提供数据控制功能来实现数据共享。

数据库系统的出现使信息系统从以加工处理数据为中心发展到围绕共享数据库为中心的新阶段。这既便于数据的集中管理,又有利于应用程序的维护,提高了数据的利用率和兼容性,也为基于数据库技术的管理信息系统的建立和发展奠定了基础。

4. 数据库管理系统

DBMS 是一种操纵和管理数据库的大型系统软件。它提供多种功能,确保数据库管理员能正常运转和维护数据库系统,其功能包括如下几个方面。

1) 数据库定义:定义全局逻辑数据结构(模式)、局部逻辑数据结构(子模式)、存储结构(存储模式)、保密模式以及信息格式等。

2) 数据库管理:包括系统控制、数据存取及更新管理、数据的完整性及安全性控制、并发控制等。①数据安全性:每个用户或应用程序只能按规定对某些数据以指定的方式进行存取或处理,以防止数据被非法使用造成数据泄密或被破坏;②数据完整性:指数据的正确性、有效性和兼容性;对数据完整性的检查是将数据控制在有效范围内,或保持数据之间满足一定的关系;③并发控制:对多用户的并发操作进行控制和协调,以避免多用户的并发进程破坏数据的完整性;④数据恢复:将数据库从错误状态恢复到某一已知正常状态(完整状态或一致状态),以避免数据库系统因硬软件故障、操作员失误和系统遭到病毒、黑客攻击造成数据库中全部或部分数据的丢失,破坏数据库中数据的完整性及一致性。

3) 数据库的建立和维护:数据库的建立、更新、再组织、维护、恢复和性能监视等。

4) 通信:具备与操作系统、分时系统的联机处理,网络及远程作业的输入/输出接口。通常,DBMS 由数据字典、数据描述语言及其编译程序、数据操纵(查询)语言及其编

译(或解释)程序和数据库管理例行程序等部分组成。

5. 数据库应用系统

利用数据库系统提供的各种功能,开展应用开发可构成一个数据库应用系统(database application system),数据库应用系统是由数据库系统、应用软件和用户界面组成。其中,应用软件是由数据库系统所提供的 DBMS(软件)及数据库系统开发工具所编写的程序构成,而用户界面大都由可视化开发工具和数据库接口软件工具开发而成。

6. 数据库系统的用户

数据库系统的用户主要包括三类。一类是应用程序员用户。该类用户通过使用如 COBOL、P L/I、C++、Java 等程序设计语言来编写数据库应用程序。这些程序通过向 DBMS 发出 SQL 语句请求来访问数据库。这些程序通常可以是批处理应用程序,也可以是交互式应用程序,目的是允许最终用户通过联机工作站或终端访问数据库。第二类用户是最终用户。最终用户可以通过应用程序员编写的应用程序,或者直接使用数据库系统软件提供的接口来访问数据库。Web 数据库的实现,更加简化了最终用户对数据库的操作,它们只需在客户端的浏览器可视化页面上进行一些简单的操作即可完成对数据库的各种复杂处理任务。第三类用户是数据库管理员(简称为 DBA)。数据库管理员的工作是创建实际的数据库以及执行需要实施各种决策的技术控制。数据库管理员也负责确保系统正确执行操作,并且提供各种其他技术服务。

1.2 数据模型

数据模型是数据库系统的重要概念,它决定了数据库系统中的数据结构与数据的组织方式。要理解什么是数据模型,首先要弄清与数据模型相关的一些基本概念。

1.2.1 有关数据模型的一些基本概念

1. 实体及其相互联系

(1) 实体与实体集

实体(entity)是现实世界的语义范畴,是客观存在并可相互区分的事物,是信息世界最主要的研究对象。实体具有存在和可区分两大特点。例如,一个人、一口油井是实体。而树上的一片树叶一般不被看成是一个实体。虽然树叶是实际存在的,但很难有人有能力将一片树叶与树上另一片树叶区分开来,否则树叶也是实体。

在数据库设计中,关心的不是某一个实体,而是具有相同性质的实体的集合,这些具有相同性质的实体的集合被称为实体集。例如,所有的学生、所有的油井等。

(2) 属性与实体集的标识符

实体具有的某一特性称为属性(attribute),一个实体可以有多个属性。属性的具体化称为属性值。例如,学生实体的属性有姓名、年龄、班级等,而某个学生的属性可能是姓名

=张红,年龄=20岁,班级=04班。在现实世界中,没有两个实体在所有属性值上是完全一样的,尽管在某些属性值上相同,但一定有若干属性能使两个实体彼此区分开。通常,一个实体集的一个或一组属性,如果其值能唯一区分实体集的每一个实体,就称这个属性或属性组为实体集的标识符(identify)。

(3) 实体间的联系

现实世界中的事物是彼此关联的,因此描述实体的数据也是相互关联的。联系(relationship)是实体相互之间关系的抽象表示,即现实世界中事物与事物之间的语义关系。例如,“职工张三在研究院工作”,即是职工实体集与部门实体集之间的一个联系。如果参与联系的实体集的数目为 n ,我们就称这种联系为 n 元联系。根据联系的元数,可以将实体间的联系归纳为三类。

1) 二元联系:二元联系是现实世界中最常见的联系,这种联系只有两个实体集参与。例如,职工与部门之间的联系即为二元联系。二元联系又可分为以下三种情况。① $1:m$ 联系:指两个实体集 E_1 、 E_2 ,如果 E_2 中的每个实体与 E_1 中的任意个实体有关,而 E_1 中每个实体至多与 E_2 中的一个实体有关,则称该联系为“从 E_2 到 E_1 的1对多关系”,记为 $1:m$ (或 $1:n$)。如图1.2(a)所示,一个部门可以聘任多个职工,但一个职工不能被多个部门聘任,这类联系就是 $1:m$ 联系;② $1:1$ 联系:如果两个实体集 E_1 、 E_2 中的每一个实体至多只能与其中另一个实体集中的一个实体有联系,则 E_1 、 E_2 两个实体集之间存在“一对一”联系,一般记为 $1:1$ 。如图1.2(a)所示,一个部门实体集与职工实体集除了有聘任联系外,还存在负责人联系,这种联系就是“ $1:1$ ”联系;③ $m:n$ 联系:如果两个实体集 E_1 、 E_2 中的每一个实体都能与另一个实体集中任意个实体有联系,则称这两个实体集是“多对多”联系,一般记为 $m:n$ 。如图1.2(b)所示,职工实体集与设备实体集之间存在着使用联系:一个职工可以使用多台设备,一台设备可以被多个职工使用。

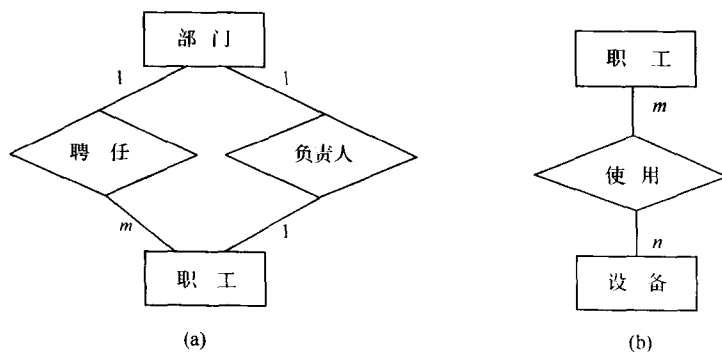


图 1.2 二元联系

2) 多元联系:虽然现实世界中以二元联系为多,但也存在高于二元的联系。这些参与的实体集至少为三个。例如,一居民李四在商店甲购买商品洗衣粉;另一居民王五在商店乙购买商品洗涤净。这种购物的语义,用联系表达出来,即为居民、商店和商品三个实体集之间的三元联系(如图1.3所示)。

3) 自反联系:自反联系是一种特殊的二元联系。参与联系的也是两个实体集,但它们

是相同的实体集,即表达的是同一实体集内两个实体之间的关系。例如,职工之间存在着上下级的领导与被领导的联系,就是一种自反联系(如图 1.4 所示)。与二元联系类似,自反联系也分为 $1:m$ 、 $1:1$ 和 $m:n$ 联系。

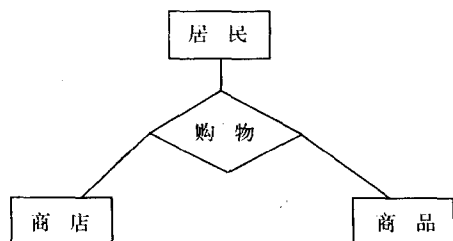


图 1.3 多元联系

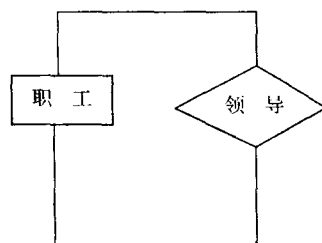


图 1.4 自反联系

在上述各种联系中, $1:1$ 是 $1:m$ 联系的特例,而 $1:m$ 联系又是 $m:m$ 联系的特例。通常,在数据模型描述中,用来表示实体(集)之间的联系的图形称为实体关联图(E-R图)。如部门与职工间的聘任联系可用图 1.5 表示。在数据库中,不仅实体(集)有标识符,而且联系(集)也有标识符。联系的标识符通常由其相关联的所有实体集的标识符组合而成。

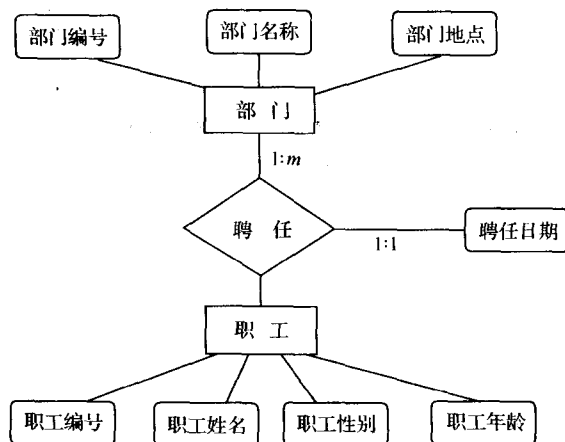


图 1.5 部门与职工间的聘任联系的 E-R 图

在数据库具体的实现中有两种类型的联系:一种是实体内部的联系,反映在数据上是记录内各属性(字段)之间的联系;另一种是实体之间的联系,反映在数据上是不同记录之间的联系。在数据库中,既要考虑记录内部的联系,还要考虑记录之间的联系。

2. E-R 图

数据库系统的一个核心问题就是研究如何建立一种数据模型来表示和处理实体间的联系。实体—联系模型(entity-relationship model),简称 E-R 模型,作为一种 E-R 模型的图示方法,它已成为一种直观表示现实世界的有力工具,在数据库设计中发挥了重要作用。这种模型直接列出所有的实体、实体的属性以及实体间的联系,并采用一些规定的符