



普通高等教育“十三五”规划教材

数据库系统与应用

吴汝明 辛小霞 赵慧青 ◆ 编著



科学出版社

普通高等教育“十三五”规划教材

数据库系统与应用

吴汝明 辛小霞 赵慧青 编著

科学出版社

北 京

内 容 简 介

本书从理论与应用相结合的角度出发,内容涵盖数据库基本原理、系统结构,以及使用 Access 2010 进行数据库建立、管理、开发等相关的概念、方法和技巧。全书共分 8 章,主要包括数据库系统概述、数据库与表、查询、窗体、报表、宏、模块与 VBA 程序设计、VBA 数据库访问技术等内容。

本书内容充实,层次清晰,原理和实践紧密结合,注重实用性和可操作性,语言叙述力求深入浅出、简明易懂,在适度的基础知识与鲜明的结构体系覆盖下,努力拓宽知识面,尽量反映数据库技术发展新动态。各章后面均配有精心设计的习题和上机实验,以便于教师教学和学生自我测试。

本书既可作为高等院校各专业数据库基础课程教材,也可以作为计算机爱好者的自学用书及全国计算机等级考试二级 Access 数据库程序设计的参考用书。

图书在版编目(CIP)数据

数据库系统与应用/吴汝明,辛小霞,赵慧青编著. —北京:科学出版社, 2017

(普通高等教育“十三五”规划教材)

ISBN 978-7-03-050704-4

I. ①数… II. ①吴… ②辛… ③赵… 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(2016)第 278354 号

责任编辑:吕燕新 王 惠 / 责任校对:刘玉靖

责任印制:吕春珉 / 封面设计:东方人华平面设计部

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>

三河市骏杰印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2017 年 1 月第 一 版 开本:787×1092 1/16

2017 年 1 月第一次印刷 印张:21

字数:485 000

定价:47.00 元

(如有印装质量问题,我社负责调换<骏杰>)

销售部电话 010-62136230 编辑部电话 010-62135397-2052

版权所有,侵权必究

举报电话:010-64030229; 010-64034315; 13501151303

前 言

随着计算机应用技术的迅猛发展及大数据时代的来临,数据处理技术已成为当代大学生知识结构的重要组成部分。数据库技术是目前存储和管理数据的主流技术,是信息管理的核心。随着社会对数据库人才的需求量越来越大,掌握基本的数据库应用技术,已成为高等学校计算机基础教学的新目标。

本书根据教育部高等学校计算机科学与技术教学指导委员会提出的《关于进一步加强高等学校计算机基础教学的意见》,并结合全国计算机等级考试二级 Access 数据库程序设计考试大纲的基本内容,由多年从事计算机应用及计算机基础教学、具有丰富教学经验和实践经验的教师编写而成。作者结合数据库技术的最新发展,并充分吸收了国内外教材的优点,努力把系统性、整体性与实用性结合起来,力求内容新颖、重点突出、文字精练、侧重应用;从实际出发,用读者容易理解的体系和叙述方法,深入浅出、循序渐进地帮助读者掌握课程的基本内容。

Access 2010 是美国 Microsoft 公司推出的基于关系数据模型的数据库管理系统,它作为 Office 2010 的组成部分,具有与 Word 2010、Excel 2010 和 PowerPoint 2010 等组件相同的操作界面和使用环境,其功能强大、使用方便,是学习和掌握数据库应用知识的有效工具。本书以 Access 2010 为基础,系统地介绍了数据库系统的基本概念和原理,使学生能较全面、系统地掌握基本的数据处理技术,同时兼顾了数据库技术与应用系统的结合,有利于提高学生的计算思维能力,为后续专业课程的学习打下良好的基础。

本书结构合理,图文并茂,在讲授理论知识的同时,穿插丰富、实用的操作实例,展示了使用 Access 2010 创建数据库及各种对象的方法,以及构成数据库应用系统的基本知识和技巧,激发学生的学习兴趣,由浅入深,引导学生自主学习。每章后面都配有相应的习题和上机实验,以实现培养学生的应用能力和实际动手操作能力的目标。

本书由吴汝明、辛小霞、赵慧青编著,其中第 1~3 章由辛小霞编写,第 4、5 章由赵慧青编写,第 6~8 章由吴汝明编写,全书由吴汝明审定。本书的编写得到中山大学新华学院计算机基础教研部全体老师的支持和配合,William Xin 提供了第 8 章编程部分的实例,在此表示衷心的感谢。同时,谨向帮助和指导我们的林卓然教授,致以深深的敬意和真挚的感谢!

由于编者水平所限,书中如有不足之处,敬请读者批评指正,以便修订时改进。如读者在使用本书的过程中有其他意见或建议,恳请向编者提出宝贵意见。

编 者

2016 年 12 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 数据管理的诞生	1
1.1.2 关系数据库的由来	2
1.1.3 面向对象数据库	2
1.1.4 数据仓库和数据挖掘	3
1.1.5 大数据时代	3
1.2 数据库系统的结构	6
1.2.1 数据及其相关概念	6
1.2.2 数据库系统	7
1.2.3 数据库系统的体系结构	10
1.2.4 DBMS 的工作过程	11
1.3 数据模型	13
1.3.1 数据模型概述	13
1.3.2 概念数据模型 (E-R 模型)	14
1.3.3 逻辑数据模型	16
1.3.4 物理数据模型	18
1.4 关系数据库的基本理论	18
1.4.1 关系模型	18
1.4.2 关系的范式及规范化	21
1.4.3 关系代数	23
1.5 数据库设计概述	27
1.5.1 数据库设计方法	27
1.5.2 数据库设计步骤	28
1.5.3 教学管理系统数据库设计实例	30
习题 1	39
第 2 章 数据库与表	43
2.1 Access 2010 数据库的创建与操作	43
2.1.1 创建 Access 数据库	43
2.1.2 数据库的打开和关闭	45

2.2 数据库对象的组织和管理	46
2.2.1 认识 Access 数据库对象	47
2.2.2 在导航窗格和工作区中操作对象	48
2.2.3 自定义操作环境	49
2.3 表的创建与编辑	50
2.3.1 设计表结构	50
2.3.2 创建表	53
2.3.3 修改表结构	58
2.3.4 字段属性的设置	59
2.3.5 设置和取消表的主键	62
2.4 表之间的关系	63
2.4.1 表间关系的类型	64
2.4.2 创建表之间的关系	65
2.4.3 编辑表之间的关系	67
2.4.4 子数据表	68
2.5 表的基本操作	68
2.5.1 打开和关闭表	69
2.5.2 调整表的外观	69
2.5.3 表的删除和重命名	71
2.5.4 在表中插入和编辑数据	72
2.5.5 记录的排序	74
2.5.6 记录的筛选	75
2.5.7 数据的汇总	78
2.5.8 数据的导入导出	79
2.5.9 管理链接表	80
2.6 数据库的管理与安全策略	81
2.6.1 数据库的备份与还原	81
2.6.2 数据库的压缩与修复	82
2.6.3 数据库的安全保护	84
2.6.4 数据库的性能分析与优化	85
习题 2	87
第 3 章 查询	91
3.1 查询概述	91
3.1.1 查询的特性与分类	91
3.1.2 查询视图	92
3.1.3 创建查询的方法	96

3.2 设置查询条件	96
3.2.1 Access 的常量	96
3.2.2 运算符	96
3.2.3 系统内置常用函数	98
3.2.4 查询条件举例	99
3.3 选择查询	100
3.3.1 使用查询向导创建选择查询	100
3.3.2 使用查询设计视图创建选择查询	101
3.3.3 在查询中进行计算	104
3.4 交叉表查询	108
3.4.1 使用查询向导创建交叉表查询	108
3.4.2 使用查询设计视图创建交叉表查询	110
3.5 参数查询	112
3.5.1 在设计视图中创建单参数查询	112
3.5.2 在设计视图中创建多参数查询	114
3.6 操作查询	115
3.6.1 生成表查询	115
3.6.2 追加查询	117
3.6.3 更新查询	119
3.6.4 删除查询	120
3.7 SQL 查询	121
3.7.1 SQL 在 Access 中的应用	121
3.7.2 SQL 数据查询	123
3.7.3 SQL 数据操纵	130
3.7.4 SQL 数据定义	132
习题 3	135
第 4 章 窗体	140
4.1 窗体概述	140
4.1.1 窗体结构	140
4.1.2 视图	141
4.2 窗体的分类	142
4.3 创建窗体	144
4.3.1 一键生成窗体	144
4.3.2 使用向导创建窗体	145
4.3.3 创建分割窗体	146
4.3.4 创建数据表窗体	147

4.3.5	创建多项目窗体	148
4.3.6	创建数据透视表窗体	149
4.3.7	创建数据透视图窗体	150
4.4	窗体控件	152
4.4.1	控件类型和属性	153
4.4.2	控件事件和事件过程	153
4.5	窗体的高级应用	154
4.5.1	标签和文本框控件的应用	154
4.5.2	复选框、选项按钮和切换按钮控件的应用	155
4.5.3	选项组控件的应用	157
4.5.4	列表框和组合框控件的应用	158
4.5.5	子窗体/子报表控件的应用	159
4.5.6	数据查询窗体的设计与创建	161
4.5.7	数据操作窗体的设计与创建	166
4.6	控制窗体的设计与创建	167
4.6.1	创建切换窗体	168
4.6.2	创建导航窗体	172
4.6.3	设置启动窗体	173
习题 4		174
第 5 章	报表	177
5.1	报表概述	177
5.1.1	报表结构	177
5.1.2	报表视图	178
5.2	报表的分类	179
5.3	创建报表	180
5.3.1	一键生成报表	180
5.3.2	使用向导创建报表	180
5.3.3	使用空报表设计报表	183
5.3.4	创建标签报表	184
5.4	报表的高级应用	187
5.4.1	在设计视图中自定义报表	187
5.4.2	报表数据的排序	189
5.4.3	报表数据的分组	190
5.4.4	两级分组统计及计算控件的实现	192
5.4.5	创建子报表	194
5.4.6	报表的外观设计和美化	196

习题 5	197
第 6 章 宏	200
6.1 宏的概述	200
6.1.1 宏的概念及类型	200
6.1.2 宏操作界面	201
6.1.3 常用宏操作命令	202
6.2 宏的创建	203
6.2.1 创建顺序操作独立宏	203
6.2.2 创建宏组	204
6.2.3 创建条件宏	205
6.2.4 创建嵌入宏	206
6.2.5 创建数据宏	208
6.2.6 创建自动宏	209
6.3 宏的调试与运行	209
6.3.1 调试宏	209
6.3.2 运行宏	210
6.4 宏与 Visual Basic	212
6.4.1 宏与 VBA 编程	212
6.4.2 将独立宏转换为 VB 代码	213
习题 6	215
第 7 章 模块与 VBA 程序设计	217
7.1 模块与 VBA 概述	217
7.1.1 模块的概念	217
7.1.2 模块的组成	218
7.1.3 VBA 的编程环境	219
7.1.4 模块的创建	226
7.2 VBA 编程基础	227
7.2.1 数据类型	227
7.2.2 常量	230
7.2.3 变量	231
7.2.4 常用函数	237
7.2.5 运算符与表达式	247
7.2.6 语句	250
7.3 程序流程控制	252
7.3.1 顺序结构程序设计	252

7.3.2 选择结构程序设计	253
7.3.3 循环结构程序设计	261
7.4 过程与作用域	269
7.4.1 过程的定义与调用	270
7.4.2 参数传递	275
7.4.3 过程的作用域	278
7.5 面向对象的程序设计概述	278
7.5.1 对象的概念	278
7.5.2 对象的属性	280
7.5.3 对象的方法	282
7.5.4 对象的事件	282
7.5.5 DoCmd 对象及其常用方法	284
7.5.6 事件过程	287
7.6 VBA 程序调试与错误处理	289
7.6.1 VBA 程序的调试方法	289
7.6.2 VBA 程序的错误处理	295
习题 7	297
第 8 章 VBA 数据库访问技术	300
8.1 常用的数据库访问接口技术	300
8.2 数据访问对象 (DAO)	301
8.2.1 DAO 数据访问对象模型	301
8.2.2 使用 DAO 访问数据库	302
8.3 ActiveX 数据对象 (ADO)	304
8.3.1 ADO 数据模型	305
8.3.2 利用 ADO 访问数据库的方法	307
8.4 域聚合函数	312
习题 8	313
附录 1 常用函数	315
附录 2 习题参考答案	318
附录 3 全国计算机等级考试二级 Access 数据库程序设计考试大纲 (2013 年版)	320
参考文献	324

第 1 章 数据库系统概述

数据库技术是现代信息科学与技术的重要组成部分，是计算机数据处理与信息管理系统核心。数据库技术的核心是数据管理。数据管理技术是以数据建模和数据库管理系统（Database Management System, DBMS）核心技术为主的一门学科，其研究的是如何科学地组织和存储数据，减少数据存储的冗余，实现数据共享和保障数据安全，以及高效地检索和处理数据，从各种数据中快速获得有价值的信息。

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。尤其是经过几十年的发展，数据库技术已具有坚实的理论基础、成熟的商业产品和广泛的应用领域。从企业资源规划、医院信息管理、数字校园、电子政务，到数据挖掘、商务智能，再到今天的大数据时代，在数据库技术的承载下，数据将成为一种核心竞争力。

1.1 数据库技术的发展史

数据作为信息的载体，其电子化过程经历了半个多世纪的发展，脉络清晰可见。从数据管理的角度来看，数据库技术经历了人工管理、文件系统、数据库系统三代演变。其中最具代表性的是关系数据库阶段。直到今天，关系数据库都是高价值数据的主要存储管理方式。相信在未来相当长的时间里，关系数据库都将继续发挥重要作用。今天的大数据技术，则充分吸收了关系数据库的简洁、易操作和文件系统的高效快速的优点，把数据的易用性和大容量结合起来，使海量数据的使用变为可能。下面介绍一下数据库的发展历程。

1.1.1 数据管理的诞生

数据库的历史可以追溯到 20 世纪 50 年代中期以前，即计算机发展的早期，当时的计算机主要用于科学计算，还未出现磁盘存储设备和操作系统软件，数据多以穿孔纸带这种裸文件方式进行物理的储存和处理，数据管理主要靠人工系统。20 世纪 50 年代末期，随着磁盘、磁鼓等直接存储设备的出现，以及软件领域有了操作系统，计算机不仅可用于科学计算，而且可用于数据管理方面，因而进入了文件系统阶段。这个阶段，数据是由操作系统中的文件系统模块来管理的。其主要特点是，数据以“文件”形式长期保存，实现了一定程度的数据共享和数据管理能力。但由于文件之间相对独立，文件系统缺乏完整和统一的管理与共享数据的能力，其缺陷是具有较多的数据冗余和数据的不一致，数据之间的联系弱，数据的逻辑独立性差，并且由于文件是为某一特定应用服务的，难以在已有数据上扩充新的应用。

数据库系统出现于 20 世纪 60 年代。当时计算机开始广泛地应用于数据管理,对数据共享提出了越来越高的要求,传统的文件系统已然无法满足人们的需要,数据库系统应运而生。应该说,数据库技术是在文件系统的基础上发展而来的,两者都是以数据文件的形式组织数据。由于在文件系统之上加入了 DBMS 对数据进行管理,数据库系统克服了文件系统的缺陷。其特点是采用数据模型表示复杂的数据结构,具有数据的集成性、数据的高共享性与低冗余性、数据的独立性、数据的统一管理与控制功能。

数据模型是数据库系统的核心基础,各种 DBMS 软件均基于某种数据模型。通常按照数据模型的特点将传统数据库系统分为网状数据库(Network Database)、层次数据库(Hierarchical Database)和关系数据库。最早出现的是网状数据库,由美国通用电气公司 Bachman 等人在 1961 年开发成功的世界上第一个数据库管理系统——集成数据存储(Integrated Data Store, IDS),奠定了网状数据库的基础,并在当时得到了快速的发展和广泛的应用。网状模型对于层次结构和非层次结构的事务都能比较自然地模拟,在数据库发展史上,网状数据库占有重要地位。层次数据库是紧随网状数据库而出现的。最著名的层次数据库管理系统是 IBM 公司在 1968 年推出的一种适合其主机的信息管理系统(Information Management System, IMS),这是 IBM 公司研制的最早的大型数据库管理系统产品。

1.1.2 关系数据库的由来

网状数据库和层次数据库已经很好地解决了数据的集中和共享问题,但它们衍生于文件系统,受文件的物理影响较大,在数据独立性和抽象级别上仍有很大欠缺。1970 年,IBM 公司的研究员 E. F. Codd 博士在 *Communication of the ACM* 上发表了一篇名为 *A Relational Model of Data for Large Shared Data Banks* 的论文,提出了关系模型的概念,奠定了关系模型的理论基础。这篇论文被普遍认为是数据库系统历史上具有划时代意义的里程碑。E. F. Codd 的心愿是为数据库建立一个优美的数据模型。后来 E. F. Codd 又陆续发表多篇文章,论述了范式理论和衡量关系系统的 12 条标准,用数学理论奠定了关系数据库的基础。

关系模型有严格的数学基础,抽象级别比较高,而且简单、清晰,便于理解和使用。但在当时有人认为,关系模型过于理想化,用来实现 DBMS 不现实。为了促进对问题的理解,1974 年 ACM 牵头组织了一次研讨会,在会上开展了一场关于支持和反对关系数据库的辩论,这次著名的辩论会推动了关系数据库的发展。经过几十年的发展和应用,关系数据库技术越来越成熟和完善,最终成为现代数据库产品的主流。其代表产品有甲骨文公司的 Oracle、IBM 公司的 DB2、Microsoft 公司的 Microsoft SQL Server 等。

1.1.3 面向对象数据库

随着信息技术和市场的发展,人们发现关系数据库系统虽然技术很成熟,但其局限性也显而易见,即它能够很好地处理“表格型数据”,却对技术界出现的越来越多复杂

类型的数据无能为力。20 世纪 90 年代后,技术界一直在研究和寻求新型数据库系统。受当时技术风潮的影响,在相当长一段时间内,人们把精力花在研究面向对象的数据库(Object Oriented Database, OOD)上,其基本思路是在关系数据库基础上融合面向对象技术,在关系数据库上封装面向对象的存储模型,以承载应用和数据的复杂性,拥有比关系技术更强的扩展性。

然而,数年的发展表明,面向对象的关系数据库不太可能取代传统的关系数据库。理论上的完美性并没有带来市场的热烈反应。其原因在于,对于许多已经运用关系数据库系统多年并积累了大量工作数据的客户,难以承受新旧数据库的转换而带来的巨大开销。此外,面向对象的关系数据库系统使查询语言变得极其复杂,使得部分数据库系统的应用集成商和应用客户因其复杂的应用技术而退缩。

1.1.4 数据仓库和数据挖掘

1988 年,为解决企业集成问题,IBM 公司研究人员创造性地提出了一个新的术语——数据仓库。1991 年,数据仓库真正开始应用。随着数据仓库、联机分析技术的发展和成熟,形成了基本的商务智能框架,但真正给商务智能赋予“智能”生命的则是数据挖掘。数据挖掘通过分析大量的数据来揭示数据之间隐藏的关系、模式和趋势,从而为决策者提供新的知识,是一种深层次的数据分析方法。它与传统数据分析的本质区别是,数据挖掘是在没有明确假设的前提下去挖掘信息、发现知识。数据仓库是数据挖掘的数据基础,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。数据仓库和数据挖掘的研究和应用,需要把数据库技术、统计分析技术、人工智能、模式识别、高性能计算、神经网络等技术相结合,目的是充分利用已有数据资源,把数据转换为信息,从中挖掘出知识,提炼成智慧,最终创造出效益。

知识链接

数据仓库在商业应用中最为脍炙人口的当属“啤酒和尿布”的故事。话说沃尔玛公司拥有世界上最大的数据仓库,在一次购物篮分析之后,研究人员发现跟尿布一起搭配购买最多的商品竟是风马牛不相及的啤酒!这是对历史数据进行挖掘和深层次分析的结果,反映的是数据层面的规律。但这是一个有用的知识吗?沃尔玛公司的分析人员不敢妄下结论。经过大量的跟踪调查,终于发现事出有因:在美国,一些年轻的父亲经常要被妻子派到超市去购买婴儿尿布,30%~40%的新生儿爸爸会顺便买点啤酒犒劳自己。沃尔玛公司随后对啤酒和尿布在卖场的摆放位置做了调整,并进行了捆绑销售,不出意料,销售量双双增加。这种点“数”成金的能力,是数据挖掘真正的“灵魂”和魅力所在。

1.1.5 大数据时代

阿基米德说过:“给我一个支点,我能撬起地球。”仿照类似的语调,Microsoft 公司的史密斯这样说:“给我提供一些数据,我就能做一些改变。如果给我提供所有数据,

我就能拯救世界。”我们现在正处于这样的“大数据”时代，带给人们极大憧憬和想象的“大数据”，正在掀起一场数据技术的革命，它将对人类的世界观、知识发现方式、思维方式及伦理道德等产生全方位的影响。

1. 大数据的定义

一般意义上，“大数据”是指无法使用传统流程或工具在合理的时间和成本内处理或分析的信息。此定义或许还不足以完全描述大数据，因此附加了一个普遍认可的 4V 特征。

1) Volume: 数据体量大，增长速度快。这个特征隐含的应用提示是，在具体的部署实施过程中，数据存取必须支撑海量的数据并发访问，并且数据处理的性能必须支持海量吞吐率。企业为提高整个企业决策需要利用庞大的数据，并以前所未有的速度持续增加，数据量从原有的 TB 级发展到了 PB 级甚至 ZB 级。相关数据指出，互联网上的数据每年增长 50%，每两年翻一番。仅 Facebook 每天就会产生超过 10TB 的数据，某些企业每小时就会产生数 TB 的数据。

2) Velocity: 数据的流动性大，变化迅速。这个特征隐含的应用提示是，在具体的部署实施过程中，数据采集速度要快，数据存取速度要快，数据分析速度要快，而所有的这些要求都对相关的技术选型、策略定位有很大的影响。在许多应用场景中，数据的价值在于其时效性。例如，企业希望基于大数据解决生产和销售的不平衡，准确洞悉应该生产多少、配送多少，并将所有的配送中心纳入体系中，形成一个动态网状结构，让退货、残次等问题与生产基地实时连接起来。试想，如果今天数据的分析结果要等到明天才能得到，显然这些数据将失去分析的意义。

3) Variety: 数据的来源多样和类型多样。这个特征隐含的应用提示是，在具体的部署实施过程中，数据存取必须支持多格式、多模式的高效管理，以及数据分析手段必须支持多格式、多模式的有效分析。随着无线感知设备、监控设备、智能设备及社交协作技术的激增，企业中的数据也变得更加复杂，不仅包含传统结构化数据，而且包含 Web 日志、网页、搜索引擎、电子邮件、文档、传感器数据、音频、视频等非结构化和半结构化的数据。

4) Value: 数据密度低，商业价值高。这个特征隐含的应用提示是价值密度稀疏。以视频为例，一部一小时的视频，在连续不间断监控过程中，可能有用的数据仅仅一两秒。大浪淘沙而又弥足珍贵，正是大数据的价值特征。

上述定义及 4V 特征的描述只是一个普适的通用描述，在具体实施过程中，不同的公司也会根据不同的价值观和方法论提出其他具体特征。IBM 公司提出的真实性 (Veracity) 特征是从大数据部署实施过程中数据质量的维度考虑的。IBM 公司认为，真实性是当前企业亟待考虑的重要维度，将促使他们利用数据融合和先进的数学方法进一步提升数据的质量，从而创造更高的价值。

2. 大数据支撑技术

大数据的 4V 特征给大数据分析带来了极大的挑战, 算法分析的高复杂度和海量分析吞吐率的矛盾, 必须依赖更快、更高效的计算架构来解决。一个大数据项目的开展, 需要回答的基本问题如下: ①数据从哪里获得, 如何获得; ②数据将存储在哪里, 如何透明存取; ③如何分析使用这些海量数据; ④面对海量数据, 如何高效计算; ⑤如何运维。图 1-1 在概念上给出了大数据支撑技术框架应包含的要素。

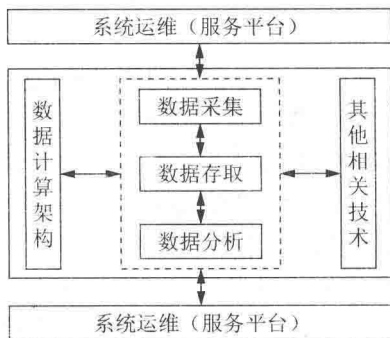


图 1-1 大数据技术框架

需要强调的是, 大数据要求能够处理几乎所有类型的海量数据, 如文档、图片、视频、音频、电子邮件等, 要求的处理速度几乎实时, 而且这种大量数据的计算要求必须是面向普通用户的, 必须是廉价的, 这种情况就促使了大数据走向云端。云计算的处理能力和它的分布式结构, 为大数据的商业模式的实现提供了可能。应该说, 云计算是大数据的 IT 基础, 强调的是计算; 而大数据则是云计算的应用, 是计算的对象。两者类似于一个硬币的两面。

3. NoSQL 的定义

大数据时代的超大数据体量和占相当比例的半结构化数据和非结构化数据的存在, 已经超越了传统数据库的管理能力。随着大数据的兴起, NoSQL 数据库成为极其热门的新领域。NoSQL 是 Not Only SQL 的缩写, 它的意义是: 适用关系数据库时就使用关系数据库, 不适用时也没有必要非使用关系数据库不可, 可以考虑使用更加合适的数据存储方式。关系数据库的最大优点是可以保证数据的一致性和处理的完整性, 它以连接为前提, 也就是说, 各个数据之间存在关联是关系数据库得名的主要原因, 但为了进行连接处理, 关系数据库不得不把数据存储在同一个数据库内, 这不利于数据的分散。NoSQL 数据库由于不进行复杂处理, 不支持连接, 每个数据都是独立设计的, 在这一点上恰恰弥补了关系数据库的不足, 很容易把数据分散到多个服务器上, 适用于大量数据的写入操作和简单查询的快速处理。

1.2 数据库系统的结构

数据库系统是一个基于计算机的、统一集中的数据管理体系。学习数据库技术,首先需要对数据库系统的基本概念、原理和方法有一个基本的认识。

1.2.1 数据及其相关概念

1. 数据

我们常说,水的温度是 100°C , 旅游景点门票是 90 元。通过水、温度、 100°C 、旅游景点、门票、90 元这些关键词,我们的大脑里就形成了对客观世界的印象。这些约定俗成的关键词就构成了我们探讨的数据基础。提到的关键词必须是人们约定俗成的,这就表示不同宗教、不同国家的人,对于关键词的约定必然会有差异。由此可以推导出数据其实也具有使用范围。不同领域的人在描述同一事物会出现不同的数据。数据的有范围性导致由此建立的信息世界、知识世界,在不同的国家、不同的宗教中会产生差异。认识到数据的有范围性可以帮助我们在一个领域进行知识管理时,首先要统一关键词或数据的约定。

因此,对数据可以进行这样的描述:数据是使用约定俗成的关键词,对客观事物的数量、属性、位置及其相互关系进行抽象表示,以适合在这个领域中用人工或自然的方式进行保存、传递和处理。

2. 信息

作为知识层次的中间层,有一点可以确认,那就是信息必然来源于数据并高于数据。我们知道,像 100°C 、90 元这些数据是没有联系的、孤立的。只有当这些数据用来描述一个客观事物和客观事物的关系,形成有逻辑的数据流时,它们才能被称为信息。除此之外,信息还具有一个非常重要的特性,即时效性。例如,新闻说广州气温 37°C , 这个信息是无意义的,必须加上今天或明天等表示时效性的信息。

信息的时效性对于使用和传递信息有重要的意义。它提醒我们,失去时效性的信息就不是完整的信息,甚至会变成毫无意义的信息流。所以我们认为,信息是具有时效性的、有一定含义的、有逻辑的、经过加工处理的、对决策有价值的信息流。

3. 知识

信息虽给出了数据中一些有意义的东西,但它的价值往往会在时间效用失效后开始衰减,只有通过归纳、演绎、比较等手段对信息进行挖掘,使其有价值的部分沉淀下来,并与已存在的人类知识体系相结合,这部分有价值的信息才会转变成知识。例如,广州 7 月 1 日的气温为 37°C , 12 月 1 日的气温为 7°C 。这些信息一般会在时效性消失后变得

没有价值，但当人们对这些信息进行归纳和对比时就会发现，广州每年7月的气温会比较高，12月气温比较低，于是总结出一年有春、夏、秋、冬4个季节。因此，我们认为知识就是沉淀并与已有人类知识库进行结构化的有价值信息。

1.2.2 数据库系统

1. 数据库系统的组成

数据库系统（DataBase System，DBS）宏观上由三大部分组成，即计算机硬件平台（包括计算机设备、网络及通信设施）、软件平台（包括操作系统、数据库管理系统、数据库、语言工具与开发环境、数据库应用软件等）、各类人员（包括数据库管理员、应用集成开发人员、终端用户等）。这三大部分构成了一个以 DBMS 为核心的完整运行实体，称为数据库系统。数据库系统的组成如图 1-2 所示。

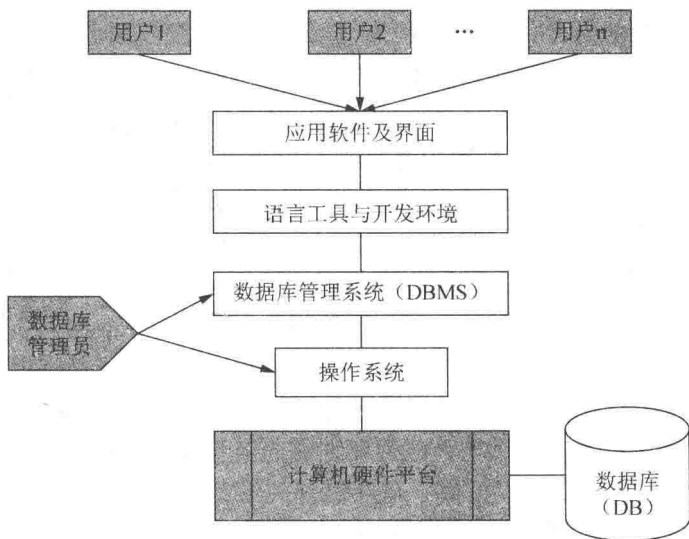


图 1-2 数据库系统的组成

（1）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、统一管理的相关数据集合，即在计算机系统中按一定的数据模型组织、存储和使用的相关联数据的集合。它不仅包括描述事物的数据本身，还包括相关事物之间的联系。数据库中的数据以文件形式存储在存储介质上，它是数据库系统操作的对象和结果。

（2）DBMS

DBMS 是数据库系统的核心，是为数据库的建立、使用和维护而配置的软件系统。它建立在操作系统的基础上，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，包括数据库的创建、查询、更新及各种数据控制等。