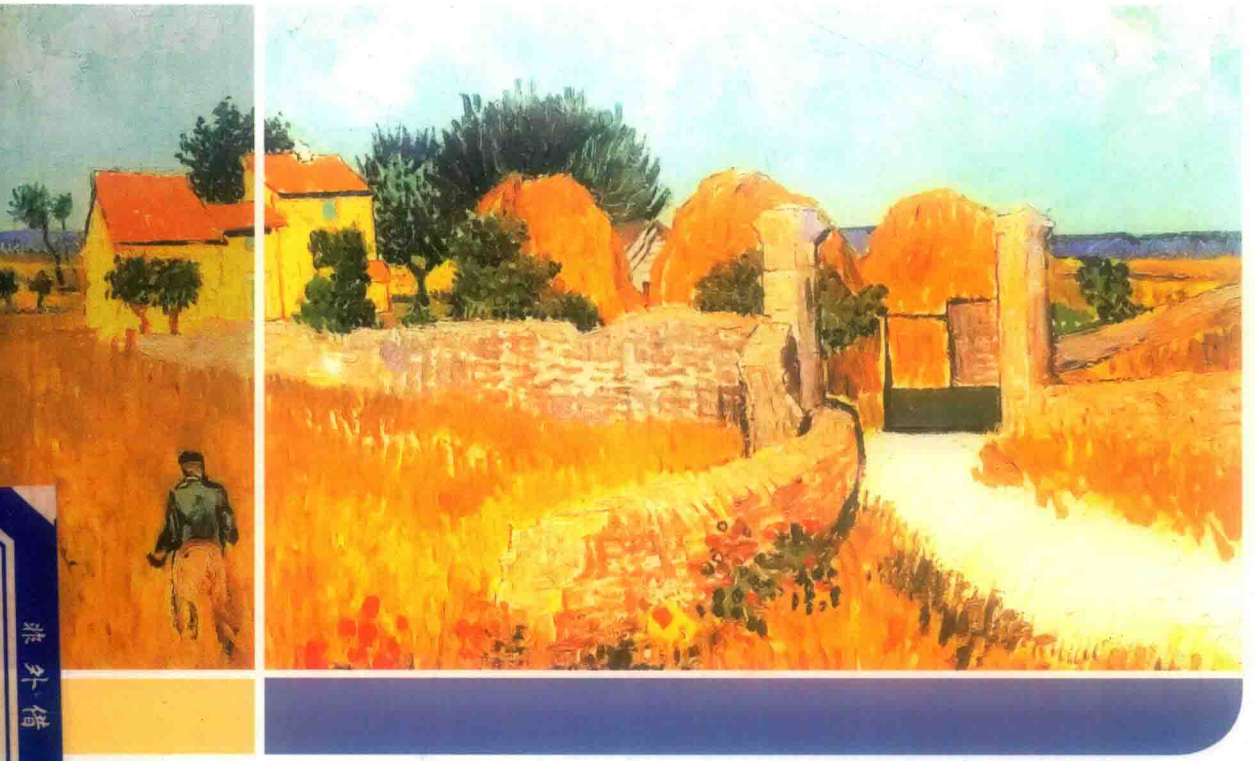
21世纪高等学校计算机
基础实用规划教材

数据库技术与应用
—— SQL Server 2012

◎ 张建国　主　编
黄庆凤　张晓芳　王　芬　副主编



清華大學出版社

21世纪高等学校计算机基础实用规划教材

数据库技术与应用
——SQL Server 2012

张建国　主　编
黄庆凤　张晓芳　王　芬　副主编
黄晓涛　阙向红　编　著

清华大学出版社
北京

内 容 简 介

本书以一个“学生成绩管理系统”演示案例为主线，分三部分介绍数据库的基础知识和数据库系统的开发方法。第一部分(第1～2章)为基础部分，介绍现代数据管理技术的发展，大数据时代的数据的特征和处理方法，数据库的基本概念，数据库设计的方法与步骤；第二部分(第3～6章)为技术部分，选用目前流行的关系型数据库管理系统SQL Server 2012，介绍其常用数据库对象的操作使用方法，包括数据库、表、约束、索引、视图、存储过程等，重点、详细地讲解了各种查询命令的设计方法；第三部分(第7章)为应用部分，介绍演示案例的设计实现过程以及所用到的相关知识，分别采用了VB.NET和VC++6.0作为前台开发工具来实现。

本书配有相应的实验内容，且每章后面均附有大量习题。

本书针对非计算机专业的学生学习数据库编写，可作为各高等院校非计算机专业相关课程的教材，也可作为其他人员学习数据库的参考教材。

图书在版编目(CIP)数据

数据库技术与应用：SQL Server 2012/张建国主编. —北京：清华大学出版社，2019
(21世纪高等学校计算机基础实用规划教材)
ISBN 978-7-302-51420-6

Ⅰ. ①数… Ⅱ. ①张… Ⅲ. ①关系数据库系统－高等学校－教材 Ⅳ. ①TP311.132.3

中国版本图书馆CIP数据核字(2018)第242453号

责任编辑：刘　星　薛　阳
封面设计：刘　键
责任校对：焦丽丽
责任印制：丛怀宇

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 装 者：北京鑫海金澳胶印有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：18.75　　**字　　数**：455千字
版　　次：2019年5月第1版　　**印　　次**：2019年5月第1次印刷
印　　数：1～1500
定　　价：49.50元

产品编号：069325-01

前 言

随着“互联网＋”和信息处理技术的不断发展，大数据时代的到来，以及人工智能、机器学习的发展进步，数据库在当今计算机应用中的应用越来越广泛，已成为不可或缺的数据管理基础工具。数据库的使用以及数据库系统的开发应用是很多人必须掌握的一种技能。作为当代大学生，无论何种专业，或多或少都需要处理各种各样的大量数据，没有数据库或不会使用数据库进行数据的管理和操作是不可想象的。

“数据库技术与应用”是高等学校非计算机专业一门非常重要的计算机公共课，华中科技大学几乎所有专业（理、工、医、文、管）都开设了这门课，为了适应普通高等院校各种专业的需求，以及学时数少的现实情况，我们编写了本书。为了和前期课程相呼应，本书用两种程序设计语言来讲解和开发示例应用程序。通过对本书的学习，读者可以掌握数据库的基本概念，数据库的设计实现步骤和方法，以及数据库应用系统的开发方法，也可为后续课程的学习和提高打下良好的基础。

当今社会，数据管理技术的掌握程度和数据处理能力水平的高低，是衡量大学生计算机使用水平的一个非常重要的指标，因而“数据库技术与应用”是当今各种专业的大学生必须学习和掌握的一门公共基础技能课程。为了方便、快捷地使读者适应社会，了解社会的使用情况和需求，本书分三部分组织。

第一部分数据库理论概述（第1～2章），首先介绍当今社会“互联网＋”、大数据、数据处理技术的发展情况，再讲解数据库的一些基本概念，以及数据库设计的基本步骤和方法，最后通过学生经常使用的HUB系统的模拟系统“学生成绩管理系统”来讲解系统的设计开发过程，这样做是为了简单、不局限于特定专业、易于理解实现。本书通篇都是以此模拟系统为主线讲解，力求做到通俗、易懂、不枯燥、趣味性强。

第二部分是数据库技术（第3～6章），以社会上使用较普及的微软公司的SQL Server 2012进行讲解，主要介绍常用的数据库、表、索引、视图和存储过程等各种常用的数据库对象的操作使用方法，包括通过管理平台的操作和通过命令的操作两种方式。

第三部分是数据库应用系统的开发（第7章），本书采用VB.NET和VC++6.0 Console两种环境平台进行系统的开发，主要考虑的是不同专业的学生前期所学的程序设计语言的不同。在这部分介绍了常用的应用系统的架构，不同开发环境所使用的API的使用方法以及“学生成绩管理系统”的功能划分和开发。本书“学生成绩管理系统”采用C/S架构实现，有兴趣的读者也可改用B/S架构实现。

本书每章均配有大量的习题，通过这些习题的练习，可以加深和巩固所学的知识。另外，针对本书的内容，在书的附表中还附有相应的实验。为了方便读者学习和上机实践，本书例题的数据库脚本和实验用的数据库脚本、教学课件PPT、教学大纲和部分习题答案等

资料可到清华大学出版社官网本书页面下载。

本书由张建国主编。第1章和第5章由黄晓涛编写，第2章由王芬编写，第3章由张晓芳编写，第4章由阙向红编写，第6章由黄庆凤编写，第7章由张建国编写。

在本书的酝酿和编写过程中得到了我校网络与计算中心于俊清主任（书记）、李战春副书记、康玲副主任和计算机基础教研室胡兵主任的大力支持和帮助，在此衷心地表示感谢！

限于编者的水平有限、经验不足，加之编者过多，书中难免存在错误或不妥之处，恳请广大读者给予批评指正，有意见或建议可发送邮件到 workemail6@163.com。

编　者

2018年12月于华中科技大学

目 录

第1章 数据管理技术及其发展

数据库技术是从20世纪60年代末开始逐步发展起来的计算机软件技术，它的产生推动了计算机在各行各业数据处理中的应用。目前，数据处理已成为计算机应用的主要方面。在数据库系统中，通过数据库管理系统来对数据进行统一管理。为了能开发出合适的数据库应用系统，就需要熟悉和掌握数据库管理系统。SQL Server是目前广泛使用的大型数据库管理系统，本书以SQL Server 2012为背景，介绍数据库的基本操作和数据库应用系统开发方法。作为学习的理论先导，本章介绍一些数据库系统的基础知识。

1.1 数据与数据爆炸

1.1.1 数据和信息

数据(**Data**)和**信息**(**Information**)是数据处理中的两个基本概念，有时可以混用，如平时所讲的数据处理就是信息处理，但有时必须分清。一般认为，数据是事实或观察的结果，是对客观事物的逻辑归纳，是用于表示客观事物的未经加工的原始素材。数据是信息的表现形式和载体，可以是符号、文字、数字、语音、图像、视频等。

例如，王雪峰的基本工资为8350元，职称为教授，这里的“王雪峰”“8350”“教授”就是数据。在实际应用中，有两种基本形式的数据，一种是可以参与数值运算的数值型数据，如表示成绩、工资的数据；另一种是由字符组成、不能参与数值运算的字符型数据，如表示姓名、职称的数据。此外，还有图形、图像、声音等多媒体数据，如照片、商品的商标等。

信息是数据中所蕴含的意义。通俗地讲，信息是经过加工处理并对人类社会实践和生产活动的决策产生影响的数据。不经过加工处理的数据只是一种原始材料，对人类活动产生不了决策作用，它的价值只是在于记录客观世界的事实。只有经过提炼和加工，原始数据才会发生质的变化，给人们以新的知识和智慧。

数据与信息既有区别，又有联系。数据和信息是不可分离的，数据是信息的载体，是信息的表达，数据本身没有意义；信息是数据的内涵，数据只有对实体行为产生影响时才成为信息。但并非任何数据都能成为信息，只有经过加工处理之后有用的数据才能成为信息。另外，信息不随其数据形式的表示不同而改变，它是反映客观现实世界的知识，而数据则具有任意性，用不同的数据形式可以表示同样的信息。例如，一个城市的天气预报情况是一条信息，而描述该信息的数据形式可以是文字、图像或声音等。

“数据”与“信息”、“知识”和“智慧”等概念之间存在一定的区别与联系，图1.1为DIKW金字塔(from Data to Information to Knowledge to Wisdom Pyramid)。从图1.1可

以看出，从“数据”到“智慧”的认识转变过程，同时也是“从认识部分到理解整体、从描述过去(或现在)到预测未来”的过程。

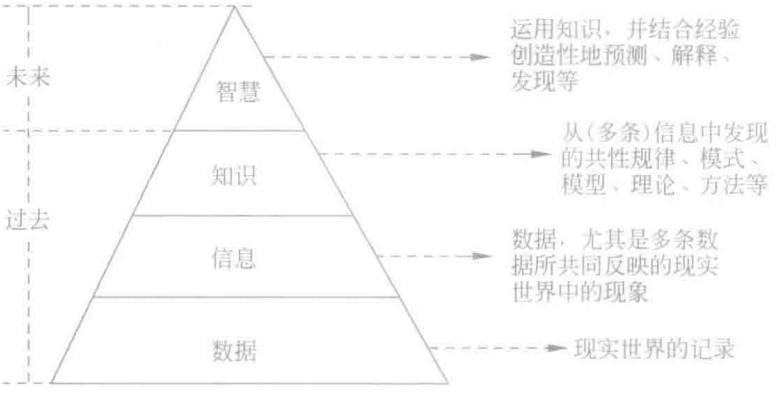


图 1.1 DIKW 金字塔

1.1.2 数据爆炸

由于网络快速发展，以及物联网时代的帷幕拉开，一场信息变革已经悄然开始。谁能在这场信息变革中领先一步，谁就能掌握新纪元的先发优势。什么是这场变革的制胜关键？既不是软件也不是硬件，而是数据。1998 年图灵奖获得者詹姆士·格雷(James Gray)曾经断言，现在每 18 个月新增的数据量等于有史以来的数据量之和。工业界每年产生的数据已达到 PB(拍字节)数量级；科学研究领域也面临相同的难题，例如欧洲核子研究中心每年产生的数据就高达 15PB。人们在信息活动中不断产生的数字化信息，如手机通信数据、出租车 GPS 数据、视频监控数据等，其总量不仅成几何级数增长，其结构也呈现连续的高维时空特性，较传统的二维关系表和< key，value >结构的万维网(Web)数据更复杂多变。“数据在，找不到”的问题日益严重，如何有效地存储和管理海量时空数据，成为这个时代的难题。

全球新产生的数据量年增长率达 40%，全球信息总量每两年就可以翻一番。2011 年全球新产生和复制的数据量达到 1.8ZB(1ZB=10^3EB=10^6PB=10^9TB=10^{12}GB)，如果用内存为 32GB 的 iPad 来存储的话，数量需要 562.5 亿个，足以砌起两座长城，由此可见大数据时代已经到来。全球的数据是由无数的数据集构成的，按照数据来源分类可分为社会数据、通过传感器收集的数据和网络数据。社会数据包括政府数据，例如国家税务总局每月收集全国数据约 4TB，已集中的结构化数据量约为 260TB。通过传感器收集的数据包括空客飞机等，空客飞机装有大量传感器，每个引擎每飞行 1 小时产生约 20TB 数据，一架飞机四个引擎，从伦敦到纽约每次飞行产生约 640TB 的数据。网络数据可细分为三类，即自媒体数据，包括在社交网络、博客、微博等应用中用户产生的数据；日志数据，包括搜索引擎、运营商、网购服务、金融服务等网络服务所产生的用户行为、交易等日志数据；富媒体数据，包括文本、音视频、图片、文字等。淘宝单日产生的日志数据量超过 50TB，存储量超过 40PB。服务行业也会累积大量的日志数据，例如国家电网公司年均产生数据 510TB(不含视频)，目前累计数据 5PB。近年来，数据规模与利用率之间的矛盾日益凸显，数据规模的“存量”和“增量”在快速增长。近年来，伴随着云计算、大数据、物联网、人工智能等信息技术的快速发展和传统产业数字化的转型，数据量呈现几何级增长。据 IDC 预测，全球数据总量预计

2020 年达到 44ZB，我国数据量将达到 8060EB，占全球数据总量的 18%。

互联网数据大爆炸，符合摩尔定律。近几年，世界基于大数据技术的人工智能学科的大发展，也适应了全球数据大爆炸的新形势。实际上，数据爆炸是无止境的。只有从大数据中提取有益人工智能(AI)、造福全人类的数据，才可以删除不必要存储的海量互联网数据，才能有效地控制数据增长、避免存储资源的浪费。

1.1.3 数据分类

数据分类是帮助人们理解数据的一个较重要的途径。一般从数据的结构化程度来分类，可以分为结构化数据、半结构化数据和非结构化数据三种，如表 1.1 所示。数据的结构化程度对于数据处理方法的选择具有重要影响，例如，结构化数据的管理可以采用传统关系型数据库技术，这是本书的主要内容，而非结构化数据的管理往往采用 NoSQL、NewSQL 或云技术等。

表 1.1 结构化数据、非结构化数据与半结构化数据对比

类型	含义	本质	举例
结构化数据	直接可以用传统关系型数据库存储和管理的数据	先有结构，后有数据	关系型数据库中的数据
非结构化数据	无法用关系型数据库存储和管理的数据	没有(或难以发现)统一结构的数据	语音、图像文件等
半结构化数据	经过一定转换处理后可以用传统关系型数据库存储和管理的数据	先有数据，后有结构(或较容易发现其结构)	HTML、XML 文件等

1. 结构化数据

结构化数据是以“先有结构，后有数据”的方式生成的数据。通常，人们所说的“结构化数据”主要指的是在传统关系数据库中捕获、存储、计算和管理的数据。在关系数据库中，需要先定义数据结构(如表结构、字段的定义、完整性约束条件等)，然后严格按照预定义的结构进行数据的捕获、存储、计算和管理。当数据与数据结构不一致时，需要按照数据结构对数据进行转换处理。

2. 非结构化数据

非结构化数据是没有(或难以发现)统一结构的数据，即在未定义结构的情况下或并不按照预定义的结构捕获、存储、计算和管理的数据。非结构化数据通常指无法在传统关系型数据库中直接存储、管理和处理的数据，包括所有格式的办公文档、文本、图片、图像和音频、视频等数据。

3. 半结构化数据

半结构化数据是介于完全结构化的数据(如关系型数据库、面向对象数据库中的数据)和完全无结构的数据(如语音、图像文件等)之间的数据，例如 HTML、XML 等，其数据的结构与内容耦合度高，需要进行转换处理后才可发现其结构。

目前，非结构化数据占比最大，绝大部分数据或数据中的绝大部分属于非结构化数据，因此，非结构化数据是数据科学中的重要研究对象之一，也是与传统数据管理的主要区别之一。

1.1.4 数据处理和数据管理

数据处理(Data Processing)是指将数据转换成信息的过程,其基本目的是从大量的、杂乱无章的、难以理解的数据中整理出对人们有价值、有意义的数据(即信息)作为决策的依据。例如,全体考生各门课程的考试成绩记录了考生的考试情况,属于原始数据,对考试成绩进行分析和处理,如按成绩从高到低的顺序排列、统计各分数段的人数等,进而可以根据招生人数确定录取分数线。

数据管理(Data Management)是指数据的收集、组织、存储、检索和维护等操作,这些操作是数据处理的基本环节,是任何数据处理业务中不可缺少的部分。数据管理的基本目的是为了提高数据的独立性、降低数据的冗余度、提高数据共享性、提高数据的安全性和完整性,从而能更加有效地管理和使用数据资源。

数据管理是利用计算机硬件和软件技术对数据进行有效的收集、存储、处理和应用的过程,其目的在于充分有效地发挥数据的作用。实现有效数据管理的关键是数据组织。随着计算机技术的发展,数据管理经历了人工管理、文件系统、数据库系统三个发展阶段,每一阶段的发展以数据存储冗余不断减小、数据独立性不断增强、数据操作更加方便和简单为标志,各有各的特点。

数据库系统的核心任务是数据管理。数据库技术是一门研究如何存储、使用和管理数据的技术,是计算机数据管理技术的较新发展阶段。走进数据库应用领域,就要涉及数据、信息、数据处理和数据管理等基本概念。

1.2 数据管理技术的发展过程

在计算机发展的初期,计算机主要应用于科学计算,虽然此时同样有数据管理的问题,但这时的数据管理是以人工的方式进行的,后来发展到文件系统,再后来才是数据库。也就是说,数据库技术的产生与发展是随着数据管理技术的不断发展而逐步形成的。

1.2.1 人工管理

20 世纪 50 年代中期以前,计算机主要应用于科学计算,数据量较少,一般不需要长期保存数据。硬件方面,没有磁盘等直接存取的外存储器;软件方面,没有对数据进行管理的系统软件。在此阶段,对数据的管理是由程序员个人考虑和安排的,他们既要设计算法,又要考虑数据的逻辑结构、物理结构以及输入输出方法等问题。数据依附于处理它的应用程序,数据和应用程序一一对应、互相依赖。程序与数据是一个整体,一个程序中的数据无法被其他程序使用,因此程序与程序之间存在大量的重复数据。数据存储结构一旦有所改变,则必须修改相应的程序,应用程序的设计与维护负担繁重。

以一所学校的信息管理为例,在人工管理阶段,应用程序与数据之间的关系如图 1.2 所示。

1.2.2 文件管理

20 世纪 50 年代后期至 20 世纪 60 年代后期,计算机开始大量用于数据管理。硬件方

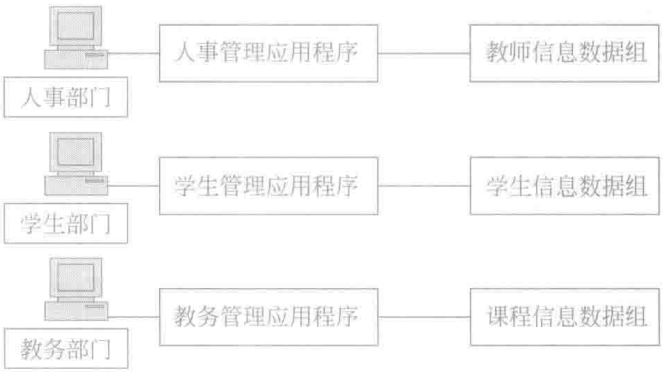


图 1.2　应用程序和数据的依赖关系

面，出现了直接存取的大容量外存储器，如磁盘、磁鼓等，这为计算机系统管理数据提供了物质基础，软件方面，出现了操作系统，其中包含文件系统，这为数据管理提供了技术支持。

数据处理应用程序利用操作系统的文件管理功能，将相关数据按一定的规则构成文件，通过文件系统对文件中的数据进行存取、管理，实现数据的文件管理方式。

文件系统为程序和数据之间提供了一个公共接口，使应用程序采用统一的存取方法来存取、操作数据，程序和数据之间不再直接对应，因而有了一定的独立性。文件的逻辑结构与存储结构有一定区别，数据的存储结构变化，不一定影响到程序，因此程序员可集中精力进行算法设计，并大大减少了维护程序的工作量。

文件管理使计算机在数据管理方面有了长足的进步。时至今日，文件系统仍是一般高级语言普遍采用的数据管理方式。然而，当数据量增加、使用数据的用户越来越多时，文件管理便不能适应更有效地使用数据的需要了，其症结表现在以下三个方面。

(1) 数据的冗余度大。

由于数据文件是根据应用程序的需要而建立的，当不同的应用程序所需要使用的数据有许多部分相同时也必须建立各自的文件，即数据不能共享，会造成大量重复。这样不仅浪费存储空间，而且使数据的修改变得非常困难，容易产生数据不一致，即同样的数据在不同的文件中所存储的数值不同，造成矛盾。

(2) 数据独立性差。

在文件系统中，数据和应用程序是互相依赖的，即程序的编写与数据组织方式有关，如果改变数据的组织方式，就必须修改有关应用程序，这无疑会增加用户的负担。此外，数据独立性差也不利于系统扩充、系统移植等开发推广工作。

(3) 缺乏对数据的统一控制管理。

在同一个应用项目中的各个数据文件没有统一的管理机构，数据完整性和安全性很难得到保证。数据的保护等均交给应用程序去解决，使得应用程序的编写相当繁琐。

在文件管理阶段，学校信息管理中应用程序与数据文件之间的关系如图 1.3 所示。

1.2.3　数据库管理

20 世纪 60 年代后期，计算机在管理中的应用规模更加庞大，数据量急剧增加，数据共享性更强。硬件价格下降，软件价格上升，编写和维护软件所需成本相对增加，其中维护成本更高，这些都成为数据管理在文件系统的基础上发展到数据库系统的原动力。

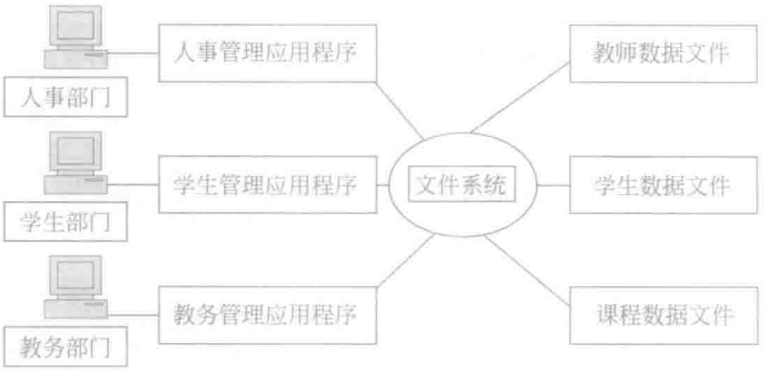


图 1.3　应用程序和数据文件的关系

数据库(DataBase,DB)是在数据库管理系统的集中控制之下,按一定的组织方式存储起来的、相互关联的数据集合。在数据库中集中了一个部门或单位完整的数据资源,这些数据能够为多个用户同时共享,且具有冗余度小、独立性和安全性高等特点。

在数据库管理阶段,由一种叫做**数据库管理系统(DataBase Management System,DBMS)**的系统软件来对数据进行统一的控制和管理,把所有应用程序中使用的相关数据汇集起来,按统一的数据模型,以记录为单位用文件方式存储在数据库中,为各个应用程序提供方便、快捷的查询和使用。在应用程序和数据库之间保持高度的独立性,数据具有完整性、一致性和安全性,并具有充分的共享性,有效地减少了数据冗余。

在数据库管理阶段,学校信息管理中应用程序与数据库之间的关系如图 1.4 所示。

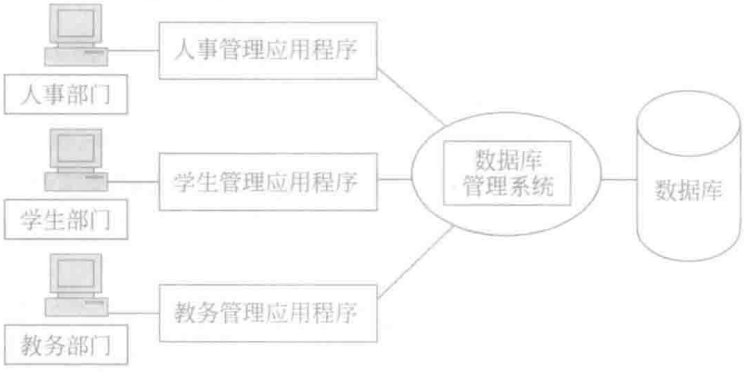


图 1.4　应用程序和数据库的关系

1.2.4　从数据库到大数据

从数据库到大数据,看似只是简单的技术演进,但细细考究不难发现两者有着本质上的差别。大数据的出现必将颠覆传统的数据管理方式。在数据来源、数据处理方式和数据思维等方面都会对其带来革命性的变化。

如果要用简单的方式来比较传统的数据库和大数据的区别,我们认为"池塘捕鱼"和"大海捕鱼"是个很好的类比。"池塘捕鱼"代表着传统数据库时代的数据管理方式,而"大海捕鱼"则对应着大数据时代的数据管理方式,"鱼"是待处理的数据。"捕鱼"环境条件的变化导致了"捕鱼"方式的根本性差异,这些差异主要体现在如下几个方面。

(1) 在数据规模上,“池塘”规模相对较小,即便是先前认为比较大的“池塘”,例如VLDB(Very Large Database)和“大海”相比仍旧偏小,“池塘”的处理对象通常以MB为基本单位,而“大海”则常常以GB,甚至是TB或PB为基本处理单位。

(2) 在数据类型上,“池塘”中的数据种类单一,往往仅有一种或少数几种,这些数据又以结构化数据为主;而在“大海”中数据种类繁多,数以千计,而这些数据又包含着结构化、半结构化以及非结构化的数据,并且半结构化和非结构化数据所占份额越来越大。

(3) 在模式(Schema)和数据的关系上,传统的数据库是先有模式,然后才会产生数据,这就好比是先选好合适的“池塘”,然后才会向其中投放适合在该“池塘”环境中生长的“鱼”;而大数据时代很多情况下难以预先确定模式,模式只有在数据出现之后才能确定,且模式随着数据量的增长处于不断的演变之中。这就好比先有少量的鱼类,随着时间推移,鱼的种类和数量都在不断地增长,鱼的变化会使大海的成分和环境处于不断的变化之中。

(4) 在处理对象上,在“池塘”中捕鱼,“鱼”仅仅是其捕捞对象;而在“大海”中,“鱼”除了是捕捞对象之外,还可以通过某些“鱼”的存在来判断其他种类的“鱼”是否存在。也就是说传统数据库中数据仅作为处理对象,而在大数据时代,要将数据作为一种资源来辅助解决其他诸多领域中的问题。

(5) 在处理工具上,捕捞“池塘”中的“鱼”,一种渔网或少数几种基本渔网就可以应对,也就是所谓的one-size-fits-all;但是在“大海”中,不可能存在一种渔网能够捕获所有的鱼类,也就是说no-size-fits-all。

从“池塘”到“大海”不仅仅是规模的变大,传统的数据库代表着数据工程(Data Engineering)的处理方式,大数据时代的数据已不仅仅只是工程处理的对象,需要采取新的数据思维来应对。图灵奖获得者、著名数据库专家James Gray博士观察并总结人类自古以来,在科学研究上先后历经了实验、理论和计算三种范式。当数据量不断增长和累积到今天,传统的三种范式在科学研究,特别是一些新的研究领域已经无法很好地发挥作用,需要有一种全新的第4种范式来指导新形势下的科学研究。基于这种考虑,James Gray提出了一种新的数据探索型研究方式,被他自己称为科学研究的“第4种范式”(The Fourth Paradigm)。第4种范式的实质就是从以计算为中心转变到以数据处理为中心,也就是我们所说的数据思维。这种方式需要从根本上转变思维,正如前面提到的“捕鱼”,在大数据时代,数据不再仅仅是“捕捞”的对象,而应当转变成一种基础资源,用数据这种资源来协同解决其他诸多领域的问题。计算社会科学基于特定社会需求,在特定的社会理论指导下收集、整理和分析数据足迹,以便进行社会解释、监控、预测与规划的过程和活动。计算社会科学是一种典型的需要采用第4种范式来作指导的科学研究领域。

1.3 大数据时代

大数据(Big Data)是指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合,是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产。

在维克托·迈尔-舍恩伯格及肯尼思·库克耶编写的《大数据时代》中,大数据是指不用随机分析法,如抽样调查这样的方法来分析处理,而是采用所有数据来进行分析处理的