



教育部实用型信息技术人才培养系列教材

Access

数据库基础与应用

黄崑 白雅楠 等 编著



清华大学出版社



ITAT 教育部实用型信息技术人才培养系

Access

数据库基础与应用

黄崑 白雅楠 等 编著

Go 

清华大学出版社
北京

内容简介

本书介绍了数据库原理与应用的基础知识；选取Access 2010版本，介绍了关系数据库开发的一般流程；详细介绍了Access 2010的六大数据库对象的基本概念、功能及设计方法；同时还分析了SharePoint网站构建及数据库系统的安全管理的基本问题；最后介绍了综合应用案例。本书共13章，分别是绪论、关系数据库基础、Access系统概述、数据库与数据表、查询设计、SQL查询设计、窗体设计、宏设计、报表设计、模块与VBA编程、SharePoint网站、数据安全，以及图书管理系统。

本书在介绍数据库理论知识的同时，提供了丰富的例题，对解决过程进行了充分的分析和讨论，给出了问题解决思路和具体实现步骤，深入浅出地讲解了Access数据库对象的设计方法和应用要点，并给出了相应的知识点引申和扩展。

本教材适合数据库学习的初学者，可以作为高校相关专业以及社会上开展Access数据库使用技巧教育与培训的教材，也可以作为Access国家计算机等级考试（二级）、管理信息系统、信息系统分析设计的参考用书。

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

Access数据库基础与应用 / 黄崑，白雅楠等编著. —北京：清华大学出版社，2014

教育部实用型信息技术人才培养系列教材

ISBN 978-7-302-35071-2

I. ①A… II. ①黄… ②白… 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（2014）第006767号

责任编辑：冯志强

装帧设计：吕单单

责任校对：徐俊伟

责任印制：何 芊

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

印 装 者：清华大学印刷厂

经 销：全国新华书店

开 本：185mm×260mm 印 张：26 字 数：646 千字

版 次：2014 年 8 月第 1 版 印 次：2014 年 8 月第 1 次印刷

印 数：1~3000

定 价：49.80 元

数据库技术是信息技术的重要组成部分，与计算机软硬件技术、网络与通信技术共同影响着人类的社会生产与实践活动。在人类社会信息化推进过程中，各行各业的信息建设开展得如火如荼。要深刻认识、理解、驾驭甚至自主研发信息系统和信息技术，离不开对数据库知识的学习、积累和实践。数据库技术是认识和理解信息系统运行原理的一个重要知识点，无论是专门的数据库原理课程，还是信息系统分析与设计课程、管理信息系统等课程，数据库知识都是课程讲授中不可缺少的一部分。Access是Microsoft Office系列软件之一，是一款桌面型数据库管理软件。基于Access学习数据库原理，有助于降低学习门槛。

本教材以2010版作为学习版本，对数据库相关知识以及Access使用进行系统介绍，让读者在认识、理解数据库基本概念基础上，具备一定的数据库设计和开发能力。全书共计13章，主要包括：

第1章 绪论主要介绍数据库相关的基本概念、数据库系统的结构与组成、关系数据库的概念和原理，相关数据库产品及数据库发展趋势。

第2章 关系数据库基础主要介绍关系数据库、关系运算的基本概念、关系数据库设计的基本过程，数据库设计的规范化理论。

第3章 Access系统概述主要介绍Access 2010的工作界面及新增功能、Access 2010包含的六大数据库对象的特点与功能、经典的罗斯文数据库的基本功能和界面，以及Access 2010数据库快速入门的学习路径。

第4章 数据库与数据表主要介绍数据库的创建和使用方法、数据表的创建和使用方法、数据表结构的修改和维护，以及数据的导入、链接与导出方法。

第5章 查询设计主要介绍Access 2010查询对象的基本概念，查询的类型、查询设计的一般方法，以及通过查询向导和查询设计视图完成各类查询创建与设计的方法。

第6章 SQL查询设计主要介绍SQL查询的设计流程、设计视图，以及其数据定义功能、数据查询功能和数据操作功能。

第7章 窗体设计主要介绍窗体的基本概念、功能与特点，窗体设计的流程、窗体控件，以及浏览型、登记型、编辑型、查询型和统计型等常见窗体的设计方法。

第8章 宏设计主要介绍Access宏的基本概念、宏的设计视图，建立操作系列宏、宏操作组、条件宏、数据宏等的原理，以及编辑和调试宏的方法。

第9章 报表设计主要介绍Access报表的基本概念、功能，报表的设计流程，文字型报表和图表型报表的设计方法及报表的高级操作、报表的预览和打印功能的实现。

第10章 模块与VBA编程主要介绍Access模块对象的基本概念、功能，VBA程序设计的流程与环境、基础语法、程序流程控制的设计，及子过程和函数的创建和程序调试方法。

第11章 SharePoint网站主要介绍SharePoint网站的概念、功能和组成，展示SharePoint网站与Access数据库的互联操作，包括数据迁移、链接、导入和导出以及脱机工作后的数据同步问题。

第12章 数据安全主要介绍数据库安全的基本概念、Access 2010安全机制，对数据库进行加编码及其解除的方法，以及用户安全机制、受信任位置的建立方法。

第13章 综合应用以图书管理系统为例，综合前面各章知识，分析了图书管理系统设计的基本过程，重点介绍了图书信息管理、图书借阅管理及相关报表打印功能的实现方法，以及Access与Excel、Word的综合运用。

本教材的特色有：

(1) 在知识内容安排上，注重点面结合。首先，由数据库的一般性基础知识出发，结合Access讲解关系数据库的原理；其次，以Access 2010的六大数据库对象为主体介绍了这些数据库对象的设计和实现方法，同时也介绍了SharePoint网站数据展示、数据库系统安全防范等问题；最后，以统一的案例系统进行综合的运用和演示，整体上注重前后知识的融会贯通，同时还对Access与Office其他软件的综合应用进行了举例和说明，突出了现实应用价值。

(2) 在知识讲解方式上，注重问题为导向，突出问题分析过程。每个章节的内容安排上，先以问题为导向引入，指出各章节拟解决的问题，引导读者有方向性地进行学习。在知识点讲解中，结合多个知识点设计综合操作案例，突出知识之间的逻辑联系，同时注意一题多解的分析，有助于读者举一反三。

(3) 在知识展现方面，注重形象的图形化展示方法。在一些抽象知识点的结合过程和操作案例的讲解过程中，提供框图的方式展示概念之间的关系、问题解决的思路 and 流程，阅读直观、形象，培养和锻炼读者的问题分析和解决能力。

本教材由黄崑、白雅楠主编，集体参与完成。蔡琦琪负责编写第1章，姚燕玲负责编写第2章，秦婷玉负责编写第3章，朱艳楠负责编写第4章，游杰负责编写第5章，马榕和白雅楠合作编写第6章，侯天砾负责编写第7章，吴江华和黄崑合作编写第8章，张慧云负责编写第9章，黄崑负责编写第10、13章，白雅楠负责编写第11、12章。黄崑和白雅楠负责全书统稿和校对。在此，我们对他们的辛勤工作表示衷心的感谢，也感谢为本书出版做出努力的各位老师和工作人员，感谢多年教学工作中给我们启发和帮助的学生们和朋友们。

由于水平和时间有限，写作过程中难免出现错漏之处。我们诚恳地欢迎各位专家以及每一位使用本书的读者批评指正，以便今后进一步修订和完善。

如需本教材示例文件，请发邮件至bookppt@126.com索取。

编者

2013年10月

第1章 绪论	1	2.7.1 规范化理论	48
1.1 数据、信息与商业智能	2	2.7.2 1NF	50
1.1.1 数据	2	2.7.3 2NF	51
1.1.2 信息	3	2.7.4 3NF	52
1.1.3 商业智能	6	2.7.5 BCNF	52
1.1.4 数据处理	7	2.7.6 其他范式	52
1.2 数据库与数据库管理系统	10	2.8 习题	53
1.2.1 数据库	10	第3章 Access系统概述	57
1.2.2 数据库管理系统	13	3.1 初识Access 2010	58
1.3 数据库系统的组成与结构	15	3.1.1 Access的基本功能及版本比较	58
1.3.1 数据库系统的组成	15	3.1.2 Access 2010新增功能	60
1.3.2 数据库系统的结构	16	3.1.3 Access的安装与使用	62
1.4 数据库模型	18	3.2 Access 2010工作界面	64
1.4.1 概念模型	18	3.2.1 启动Access数据库阶段	64
1.4.2 逻辑模型	21	3.2.2 设计数据库对象阶段	66
1.4.3 物理模型	22	3.3 六大数据库对象与罗斯文数据库	71
1.5 数据库产品及技术发展	23	3.3.1 六大数据库对象	71
1.5.1 主流数据库产品及供应商	23	3.3.2 罗斯文数据库	71
1.5.2 数据库发展趋势	26	3.3.3 主要业务流程操作	76
1.6 习题	27	3.4 Access数据库的学习路径	77
第2章 关系数据库基础	31	3.5 习题	78
2.1 概述	32	第4章 数据库与数据表	81
2.2 关系数据库的数据结构	32	4.1 概述	82
2.3 关系数据库的表间关联及完整性	34	4.2 数据库的创建	82
2.3.1 表间关系类型	34	4.3 数据库的使用与维护	85
2.3.2 关系的完整性	34	4.3.1 数据库的打开与关闭	85
2.4 关系运算	36	4.3.2 Access数据库文件的版本转换	86
2.4.1 关系代数	36	4.3.3 数据库的维护	87
2.4.2 传统集合运算和专门的关系运算	36	4.4 数据表的创建与维护	89
2.5 关系数据库的设计方法	39	4.4.1 数据表的创建流程与视图	89
2.5.1 数据库工程	39	4.4.2 建立表结构	92
2.5.2 软件危机与软件工程	39	4.4.3 建立表间关系	96
2.6 关系数据库设计的基本过程	41	4.4.4 字段的设置	102
2.6.1 需求分析	41	4.4.5 数据表的维护	110
2.6.2 概念设计	43	4.5 数据表的使用	110
2.6.3 逻辑设计	44	4.5.1 基于表对象的使用	110
2.6.4 物理设计	47	4.5.2 基于表数据的使用	111
2.6.5 实现、运行和维护	48	4.5.3 数据表显示方式的设置	114
2.7 关系数据库的规范化	48	4.6 数据的导入、链接与导出	115

4.6.1	链接外部数据	116	6.4.1	投影查询	166
4.6.2	数据的导出	117	6.4.2	选择查询	168
4.7	习题	118	6.4.3	排序查询	172
第5章	查询设计	121	6.4.4	使用聚合函数	173
5.1	概述	122	6.4.5	连接查询	175
5.1.1	查询的概念	122	6.4.6	子查询与相关子查询	178
5.1.2	查询的功能	122	6.5	数据操纵功能	181
5.1.3	查询的类型	123	6.5.1	数据插入	181
5.2	查询创建流程	124	6.5.2	数据更新	183
5.2.1	向导创建过程	124	6.5.3	数据删除	184
5.2.2	非向导创建过程	125	6.6	习题	184
5.3	查询相关的视图	125	第7章	窗体设计	187
5.3.1	查询设计视图	126	7.1	概述	188
5.3.2	数据表视图	127	7.1.1	什么是窗体?	188
5.3.3	SQL视图	127	7.1.2	窗体的类型与组成	188
5.4	选择查询	127	7.1.3	窗体的属性	190
5.4.1	简单选择查询	128	7.2	窗体设计流程与相关视图	197
5.4.2	条件选择查询	131	7.2.1	窗体设计流程	197
5.5	交叉表查询	141	7.2.2	窗体相关的视图	197
5.5.1	简单交叉表查询	141	7.3	窗体控件	201
5.5.2	条件交叉表查询	142	7.3.1	控件及其类型	201
5.6	参数查询	144	7.3.2	常用控件	201
5.6.1	基于选择查询的参数查询	144	7.3.3	控件属性	206
5.6.2	基于交叉表查询的参数查询	145	7.4	浏览型窗体	210
5.7	动作查询	147	7.4.1	向导型设计过程	210
5.7.1	生成表查询	148	7.4.2	非向导设计过程	212
5.7.2	追加查询	149	7.5	登记型窗体	217
5.7.3	更新查询	150	7.6	编辑型窗体	219
5.7.4	删除查询	151	7.7	查询型窗体	220
5.8	习题	152	7.7.1	非同界面查询	220
第6章	SQL查询设计	155	7.7.2	同一界面查询	222
6.1	概述	156	7.8	统计型窗体	227
6.1.1	SQL及产生背景	156	7.9	习题	231
6.1.2	SQL的功能与组成	157	第8章	宏设计	235
6.1.3	SQL的主要特点	158	8.1	概述	236
6.2	SQL查询设计流程与相关视图	158	8.1.1	什么是宏	236
6.2.1	设计流程	158	8.1.2	宏的分类	236
6.2.2	相关视图	159	8.1.3	宏的功能	237
6.3	数据定义功能	160	8.2	宏的设计流程	237
6.3.1	建立数据表	160	8.3	宏对象设计视图	239
6.3.2	修改数据表	162	8.3.1	宏设计工具栏	240
6.3.3	删除数据表	164	8.3.2	宏添加和参数设置窗口	241
6.3.4	建立索引和删除索引	164	8.3.3	操作目录	242
6.4	数据查询功能	165			

8.4	宏的建立	243	10.4	VBA基础语法	304
8.4.1	操作序列宏	243	10.4.1	数据类型	304
8.4.2	宏操作组	245	10.4.2	变量、常量和数组	306
8.4.3	条件宏	248	10.4.3	运算符和表达式	312
8.4.4	数据宏	249	10.5	程序流程控制	314
8.4.5	嵌入式宏	250	10.5.1	顺序结构	315
8.5	宏的编辑	251	10.5.2	分支结构	315
8.6	宏的执行与调试	253	10.5.3	循环结构	317
8.6.1	宏的执行	253	10.6	子过程与函数	320
8.6.2	宏的调试	253	10.6.1	子过程	321
8.7	习题	256	10.6.2	函数	328
第9章	报表设计	259	10.7	VBA程序的错误发现与调试	329
9.1	概述	260	10.7.1	错误类型	329
9.1.1	什么是报表?	260	10.7.2	错误发现	330
9.1.2	报表的组成与类型	260	10.7.3	程序调试	331
9.2	报表的设计流程与相关视图	263	10.8	习题	332
9.2.1	报表的设计流程	263	第11章	SharePoint网站	335
9.2.2	报表相关的视图	264	11.1	概述	336
9.3	文字型报表	267	11.1.1	什么是SharePoint网站	336
9.3.1	向导型设计过程	267	11.1.2	SharePoint网站的功能	336
9.3.2	非向导型设计过程	270	11.1.3	SharePoint网站的组成	337
9.4	图表型报表	274	11.2	Access数据的迁移与发布	339
9.5	报表的高级操作	276	11.2.1	Access数据的迁移	340
9.5.1	添加计算控件	276	11.2.2	Access数据的发布	341
9.5.2	排序分组	281	11.3	SharePoint网站数据的导入与导出	342
9.5.3	创建子报表	283	11.3.1	导入或链接SharePoint列表	342
9.6	报表的预览和打印	285	11.3.2	导出到SharePoint网站	344
9.7	习题	289	11.4	脱机使用链接	345
第10章	模块与VBA编程	291	11.4.1	概念	345
10.1	概述	292	11.4.2	脱机前准备和脱机工作	345
10.1.1	什么是模块?	292	11.4.3	脱机后数据更新	346
10.1.2	什么是VBA?	292	11.5	习题	346
10.1.3	模块的类型与创建方法	294	第12章	数据安全	349
10.2	VBA程序设计的流程	297	12.1	概述	350
10.2.1	VBA程序设计基本过程	297	12.2	数据库密码的设置与撤销	350
10.2.2	水仙花数编程实例	299	12.3	用户权限的分级管理	352
10.3	VBA编程环境	300	12.3.1	账号、组、权限和工作组信息	353
10.3.1	标题栏、菜单栏与标准工具栏	301		文件	353
10.3.2	工程资源管理器	302	12.3.2	管理员的安全设置	354
10.3.3	属性窗口	303	12.3.3	用户与组的安全设置	355
10.3.4	代码窗口	303	12.3.4	用户级安全机制向导	360
10.3.5	立即窗口	303	12.4	安全机制的删除	362
10.3.6	监视窗口	304	12.5	编码/解码数据库	363

12.6	信任中心	364	13.4	图书借阅与预约管理	379
12.6.1	创建受信任位置	364	13.4.1	借阅信息登记	379
12.6.2	受信任位置数据库的使用	365	13.4.2	借阅信息查看	380
12.6.3	打开数据库时启动禁用的对象	366	13.5	报表打印	381
12.7	习题	367	13.5.1	图书清单打印	381
第13章 图书管理系统		369	13.5.2	书库清单打印	386
13.1	系统概述	370	13.6	集成菜单设计	386
13.2	概念模型与数据库结构设计	370	附录	391	
13.2.1	概念模型	370	附录1	宏操作分类及其功能描述	392
13.2.2	数据库结构设计	371	附录2	常见内置函数	395
13.3	图书信息管理	372	附录3	习题答案	398
13.3.1	图书登记	372	参考文献	405	
13.3.2	图书浏览与编辑	374			
13.3.3	图书查询	376			

绪论

数据在生活中随处可见，体检中各项生理指标的测量值、学生的考试分数、工薪族的收入、超市日用品的价格等等，都是常见的数据。是否数据就是数字？其实不然，广义而言，经过数字化的文字、图像、音视频等多媒体信息也属于数据。数据的最大特点是以量化的方式描述客观事物的性质、状态和行为，便于统计和分析。要从数据中“淘金”，发掘有价值的信息，前提是要达到一定的数据量。然而，当数据量达到一定规模时，依靠随机的自由组织方式显然无法解决和应对数据的管理和分析需求。由此，数据库技术应运而生。那么，什么是数据库？数据库是否就是数据的仓库？与数据仓库是否为同一个概念？一天、一个月或者是一年的日生活消费记录是数据库吗？数据库与数据表格有什么区别？如果数据库是一个可以容纳许多数据的容器，那么这个容器有什么样的结构特点？在功能上又有什么具体作用？业界是否已经有了这样一些成熟的数据库产品？

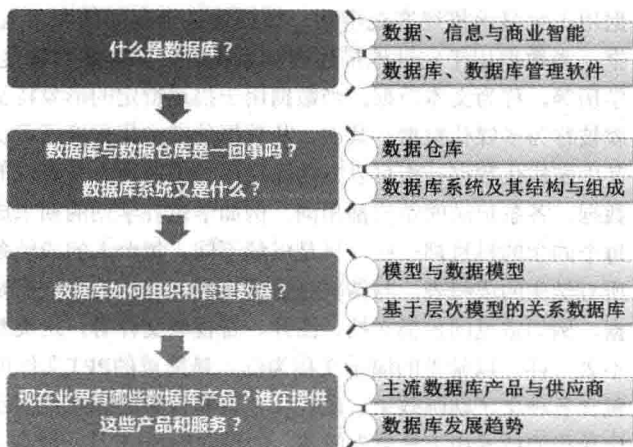
本章内容主要介绍数据库相关的基本概念、数据库系统的结构与组成、关系数据库的概念和原理，相关数据库产品及数据库发展趋势。

第1章

学习目标

1. 理解数据、数据库、数据库系统等基本概念；
2. 掌握数据库系统的结构与组成特点；
3. 了解关系数据库的基本概念、关系运算基本原理；
4. 了解主流的数据库产品和供应商及其数据库技术发展趋势。

知识结构框图



1.1 数据、信息与商业智能

1.1.1 数据

随着人类社会进入“大数据时代”，人们的一言一行、一举一动都与数据的产生、衍变与传播发生着密切的联系。诸如银行、超市每笔业务处理的各种数据、单据，重要的公共场所、交通要道及一些办公场所的摄像监控记录，医疗卫生保健机构为每位病患就诊而建立的检查结果、处方、复诊与治疗情况等信息在内的病例档案数据。特别是近年来网络社交的蓬勃兴起，也愈发加剧了人们被网络数据洪流所袭卷的情况。诸如天涯社区、强国论坛等社区论坛记录了人们广泛的言论和观点，博客、微博随时记录和传递着人们的所思所想，短信、飞信、微信等工具记录了人与人之间大量的互动内容，各种文档分享社区，如百度文库、道客巴巴、在线幻灯片分享等网站所共享的各类资源也在不断积累和增长中。现代信息技术的迅猛发展、信息交流与传播模式的变化，无不推动了“大数据时代”的来临。

大数据，可以通俗地理解为巨量庞大的数据。巨量庞大到什么程度呢？这个“大”，大到了现有主流工具都无法进行处理的地步。不过，大数据的基本构成终究还是数据，了解和掌握一般数据加工、处理、组织、存储与管理的方法和技术，有助于在未来更深刻地理解大数据所带来的机遇和挑战。那么，如何理解数据？数据与数据库之间是什么关系？

首先，数据可以是数字。比如描述一个人的身高为178cm、体重70kg，或者应聘者的面试分数是95分，电话号码是58801234，这些是最容易联想到的数据。而实际上，数据并不全是数字。比如一幅数码照片、一段对话录音或者一段影片剪辑，这些属于广义的数据。因此，数据既可以是数字，也可以是文字、图形、图像、声音、语言等多种形式，它们都经过数字化处理后被存入计算机中。

概括而言，数据可以理解为用于表示客观世界中某个实体的数量、状态和行为的符号。数据的概念包括两方面含义：一是数据内容，是指所描述客观事物的具体属性特征，也就是数据的“值”；二是数据形式，是指存储在某种媒体上的具体形式，也就是常说的数据的“类型”，包括前面提及的数字、文本、声音、图形和图像等多种形式。在计算机应用中，数据是可被计算机识别、存储和加工处理的“原料”。

按照不同的标准，数据可以分为不同的类型。首先，从数据所描述的对象来分，当数据用于定量地描述客观事物，如描述某个车间的职工人数、某季度生产量时，称为数值数据；当数据用于定性地描述客观事物时，如描述对象是一个人的籍贯、性别、政治面貌、学历等，称为文本数据；当数据用于描述特定的形象特征和过程，如图像、视频和音频等，就被称为多媒体数据。其次，从数据的结构化程度还可以分为结构化数据和非结构化数据，其中结构化数据是指可以方便地使用二维逻辑表来表现的数据，可以简单地理解成表格里的数据，各条记录的结构都相同。例如学生各学期的期末成绩可以汇总形成学生成绩表，其中每个同学的科目都一样，只是成绩不同。每个人的成绩条目依次排列到一起，就形成了班级所有学生的成绩表。与结构化数据不同，非结构化数据是指不容易以二维逻辑表来表现的数据，例如常见的各类文档、图片、音视频文件等，这类数据在大小、内容、格式、用途上都不太一样，以常见的演示文稿为例，最简单的PPT文件可以是纯文字的简洁显示，也可以是包含各种文字动画效果、图片、声音和视频的多媒体演示文件。相比而言，计算机处理结构化数据的技术和工具更为成熟，如借助Excel工具，财会、审计等工作的人员可以非常容易地进行加减乘除、汇总、统计之类的运算。



考考你

你知道“大数据”到底有多大吗？

“大数据”由巨型海量数据集组成，这些数据集大小常超出了常用软件在可接受时间内的收集、管理和处理能力。那么，究竟“大数据”有多大？一组名为“互联网上一天”的数据告诉我们，一天之中，互联网产生的全部内容可以刻满1.68亿张DVD；发出的邮件有2940亿封之多（相当于美国两年的纸质信件数量）；发出的社区帖子达200万个（相当于《时代》杂志770年的文字量）；卖出的手机为37.8万台，高于全球每天出生的婴儿数量的37.1倍……

截止到2012年，数据量已经从TB（1TB=1024GB）级别跃升到PB（1PB=1024TB）、EB（1EB=1024PB）乃至ZB（1ZB=1024EB）级别。国际数据公司（IDC）的研究结果表明，2008年全球产生的数据量为0.49ZB，2009年的数据量为0.8ZB，2010年增长为1.2ZB，2011年的数量更是高达1.82ZB，相当于全球每人产生200GB以上的数据。而到2012年为止，人类生产的所有印刷材料的数据量是200PB，全人类历史上说过的所有话的数据量大约是5EB。IBM的研究称，整个人类文明所获得的全部数据中，有90%是过去两年内产生的。而到了2020年，全世界所产生的数据规模将达到今天的44倍。

“大数据”不仅用来描述和定义信息爆炸时代产生的海量数据，也同时展示了与之相关的技术发展与创新。

来源：半月谈·大数据时代降临。 <http://im.banyuetan.org/jrt/120922/70953.shtml>

全球知名咨询公司麦肯锡曾指出：“数据，已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来。”如今人们也更加深刻地感受到，无论是在商业、经济还是政治、军事、文化、教育等领域，明智的决策都需要依赖充分可靠的数据和科学合理的分析方法，凭借个人的经验和直觉已经难以满足信息时代进行决策的需要。

为了在日益激烈的竞争中获得优势，对数据的善加利用成为商业机构、政府部门，乃至个体都不容忽视的问题。要从表面看似无关的数据中洞察规律、发现新知识，并具备对这些数据进行有效利用的能力，这些都成为成功驾驭数据的关键。数据库是用于组织与管理数据的工具，数据库的建立与维护也是进行数据分析、数据挖掘及知识发现的前提和基础。因此，数据与数据库好比对象与容器，数据库不仅装载着数据本身，还对数据及其相互联系进行维护和管理。在数据库应用基础上，数据的潜力得以发挥，数据所承载的信息得以被发现。

1.1.2 信息

“信息”一词在日常生活中使用很普遍，“你听说了今年物价调整的信息吗？”、“张三真是信息灵通的人士”、“求职信息”、“信息爆炸”、“信息家电”，各种包含“信息”的名词处处可见。那么，什么是信息？

在人类认识和改造客观世界过程中，“信息”不可或缺。据郑玄《周易注》中记载：“古者无文字，结绳为约，事大，大结其绳；事小，小结其绳。”在文字产生前，古人用结绳记事的方式记载着生活中发生的各种事情，这是早期的信息记录方式。直到有了文字之后，文字便成为了主要的信息载体。而“日出而作，日落而归”亦是人类在早期利用自然界传递的信息进行耕作的实例。下面分别从“信息”作为一个普通词汇以及专业领域中的词汇的角度来了解信息的各种概念描述，如图1-1所示。

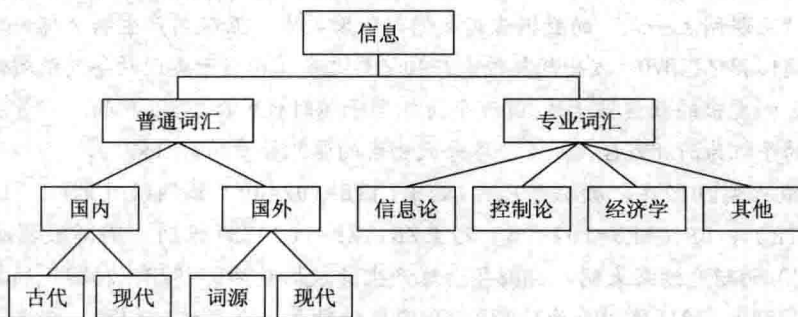


图1-1 从多角度认识“信息”

1. 普通词汇

作为一个普通词汇，国内外都对“信息”一词有过界定。

(1) 在国内

在我国古代诗词中，信息一词早已出现。例如南唐诗人李中的《暮春怀故人》诗中：“梦断美人沉信息，目穿长路倚楼台。”宋代诗人陈亮的《梅花》诗中：“欲传春信息，不怕雪埋藏。”宋代诗人李清照的《上枢密韩肖胄诗》中：“不乞隋珠与和璧，只乞乡关新信息。”大都有消息、讯息的含义。

在《现代词典》、《辞海》中关于“信息”的解释包括：

《辞海》^①对“信息”进行了两种解释：①音讯：“消息”。②通信系统传输和处理的对象，泛指消息和信号的具体内容和意义，通常须通过处理和分析来提取。

《现代汉语词典》^②中对“信息”进行了三种解释：①音信：消息。②信息论中是指用符号传送的报道，报道的内容是接收符号者预先不知道的。

(2) 在国外

“信息”一词在英文、法文、德文、西班牙文中均是“information”，追溯该词的词源可以看到，“information”源于拉丁文，原意也是指解释、陈述。而在现代英语词典中，对信息的解释包括：

韦氏在线词典（Merriam-Webster）^③中提供了关于information的多种解释。例如，①通信中接收到的知识和智能（The communication or reception of knowledge or intelligence）；②从调查、学习和指导中获得的知识（Knowledge obtained from investigation, study, or instruction）；③事实和数据（facts, data）；④通信或者计算机中表示数据的信号或者字符（a signal or character (as in a communication system or computer) representing data）。

① 夏征农，陈至立主编. 辞海（第六版）彩图本[M]. 上海：上海辞书出版社，2009.

② 吕叔湘，丁声树主编. 现代汉语词典. 北京：商务印书馆，2012.

③ <http://www.merriam-webster.com/dictionary/information>

《剑桥国际英语词典(英汉双解)》^①将information解释为信息是关于事实、新闻的知识(Information is knowledge about something, esp. facts and news)。

《柯林斯COBUILD高阶英语学习词典》^②将information解释为: ①关于某人或者某事的事实(about someone or something consists of facts about them); ②计算机存储与使用的事实和图表(consists of the facts and figures that are stored and used by a computer program); ③电话查询服务(is a service which you can telephone to find out someone's telephone number)。

可见,作为一个普通词汇,汉语中的“信息”与英语中的“information”具有相近的涵义。

2. 专业词汇

“信息”不仅是在生活中常见的词汇,在信息科学、经济学、管理学等领域也多次被作为一个专业词汇进行解释。以下列出了信息论、控制论、经济学等领域中对“信息”的解释。

信息论的创始人香农曾给出了信息的明确定义。信息论是关于信息的本质和传输规律的科学的理论,是研究信息的计量、发送、传递、交换、接收和储存的一门学科。1948年,香农在《通信的数学理论》一文中指出“信息是用以消除随机不确定性的东西”。

控制论创始人维纳在《控制论——关于在动物和机器中控制和通讯的科学》中提出“信息是我们在适应外部世界、控制外部世界的过程中同外部世界交换的内容和名称”。他认为,接收和使用信息的过程,就是我们适应外部世界环境的偶然性变化的过程,也是我们在这个环境中有效生活的过程。

我国著名的信息学专家钟义信认为“信息是事物的运动状态和过程以及关于这种状态和过程的知识”,并且,将“信息”概念分为两个层次:本体论层次和认识论层次。其中,本体论层次是指无任何约束条件来认识“信息”,认为“信息”是“事物运动的状态以及它的状态的改变”,这是一种纯客观的定义。因此,四季轮回、花开花落、虫鸣鸟啼等都属于本体论层次的信息。不同的是,认识论层次强调主体约束,“信息”是“主体所感知或者主体所描述的事物的运动状态以及状态变化的方式”。例如“周幽王烽火戏诸侯”、“梁红玉击鼓战金山”,其中的“烽火”、“战鼓”都是当事人约定好的信号,是与认识主体相关的信息。

在经济学领域中,美经济学家波拉特认为,“信息是经过组织化而加以传递的数据”。德鲁克则认为,“信息是有目的性和关联性的数据,将数据转化为信息需要知识”。

此外,电子学家、计算机科学家认为“信息是电子线路中传输的信号”。信息管理专家认为“信息是为了满足用户决策的需要而经过加工处理的数据”。

不同领域的专家从不同视角对信息的概念做出了诠释,可以从以下几方面对“信息”的概念进行理解,信息是关于事物运动状态和过程以及关于这种状态和过程的知识,文本、数值或多媒体等形式存在的数据是信息的载体,信息是数据所表达的语义。并且,根据人们认识程度的差异,对信息的接收和领悟也是具有差异性的,信息具备消除不确定性的功能。

单纯从概念界定理解“信息”显然不够全面,因此,人们会更多地从信息的类型(如隐性与显性信息;零次、一次、二次和三次信息;政务、经济、教育、军事等各领域信息)、

① [英]普洛克特编. 剑桥国际英语词典(英汉双解)(重排校订本). 上海:上海外语教育出版社, 2004.

② 柯林斯出版公司编. 柯林斯COBUILD高阶英语学习词典: 英语版=Collins COBUILD Advanced Learner's English Dictionary. 北京: 外语教学与研究出版社, 2006.

信息的特性（如客观性、普遍性、共享性、时效性、泛在性、传递性）等角度更全面地认识“信息”。总之，信息的定义和类型多种多样，在运用和理解信息时，需要注意结合其所应用的领域。例如企业的管理信息常指在企业生产、经营与管理活动中收集的各种信息，既包括来自企业内部的信息，如企业生产计划、人事档案、采购订单、库存记录、财务报表，又包括来自企业外部合作伙伴、竞争对手以及市场的信息，如供应商、分销商的联系人信息，提供同类产品与服务的竞争对手在人事变动、新产品推出方面的各种动向信息，以及顾客消费心理与行为模式的市场调研报告。

3. 数据与信息的关系

信息和数据既有联系，又有区别。数据是信息的载体，是信息的符号表示。信息是对数据加工的结果，是数据的内涵，是对数据的解释。例如，单纯地看“90, 86, 80”表示的仅仅是一组数据，但若是放在学生信息管理系统中可能就代表3个学生某科的期末成绩，此时数据被赋予了实际含义，即为信息。因此，数据是符号化的信息，信息是语义化的数据。

数据与信息的区别在于，数据可以表示信息，但是数据并不等同于信息。相同的数据可能从不同角度做出解释，即代表不同的信息。例如“半瓶水满，半瓶水空”，同样是反映半瓶水，有的人会庆幸至少还有半瓶水是满的，而另一部分人则会哀叹怎么就空了半瓶。与此同时，同一信息也可以有不同的数据表示方式。例如《三国演义》的故事情节既有小说文字版本，又有广播中的评书版本，还有电视剧、电影等不同版本。

尽管严格意义上讲，信息和数据两个术语是有区别的，但在很多场合下人们并未对二者进行严格区分。因此，在认识和理解数据与信息时，要结合具体情境来看待其所指对象。

1.1.3 商业智能

1. 智能与商业智能

智能通常是指人类智慧，是一个人智力和能力的表现。在现代数据管理应用中，数据管理的目的并不只是为了存储和保管数据，更重要的作用在于揭示和发掘其中蕴涵的一些隐性的规律，这些规律是直接服务于管理决策的知识，因此，在商业领域将其称为商业智能（Business Intelligence, BI）。比如，零售企业存储了过去三十年的销售数据，根据对过去数据的分析，预测未来五年的发展走势并制定销售战略。这与自然灾害预测的原理类似，人们同样也是根据过去地震、海啸发生时的各种地质勘探数据和征兆，进行分析和预测，得出某地某时可能的灾害类型和级别。这类新衍生的信息无法从原始数据直接得到，需要借助一定的工具和模型进行分析、处理、拟合和预测，而获得的结果正是决策人员所需要的知识，即BI。

BI最早由加特纳集团（Gartner Group）于1996年提出，当时将商业智能定义为：商业智能描述了一系列的概念和方法，通过应用基于事实的支持系统来辅助商业决策的制定。在业界，商业智能通常被理解为一组工具或者一套解决方案，其作用是可以将企业现有的数据，包括来自企业业务系统的订单、库存、交易账目、客户和供应商等数据，以及企业所处行业和竞争对手的数据、企业所处的其他外部环境中的数据进行综合的处理与分析，转化为知识，使企业的各级决策者获得知识与洞察力，辅助企业做出明智的业务经营决策。

2. 数据、信息与商业智能

商业智能是提升企业竞争优势的重要来源，但商业智能的获得并非无源之水、无本之

本，其基础依然是数据与信息。

以客户基本信息管理为例，某商场在为顾客办理会员卡时，记录了每一位顾客的性别、年龄、收入、受教育程度等基本信息，如图1-2所示，其中，年收入根据不同收入范围分为：① 10000元以下、② 10000~20000元、③ 20000~50000元、④ 50000元以上；受教育程度根据学历高低分为：① 大专及以下、② 大学本科、③ 研究生。每位顾客的各项登记信息即为初始的数据，如顾客“黄刚”的性别是“男”、顾客“顾城”的年收入为第四档，50000元以上。

	A	B	C	D	E	F
1	编号	姓名	生日	性别	年收入	受教育程度
2	1	孟美娟	1980/1/1	女	1	2
3	2	黄刚	1978/5/9	男	3	3
4	3	李毅	1981/12/16	男	3	2
5	4	胡馨	1980/7/9	女	2	3
6	5	顾城	1975/3/25	男	4	3

图1-2 顾客基本信息

根据顾客的生日信息可知顾客的年龄，按照图1-3所示的年龄分段区间，计算各年龄段的顾客人数及人数百分比，通过双坐标的柱形图、折线图进行展示。可见，该商场超过一半的顾客年龄在30~40岁之间。图1-3中所列出的顾客年龄段统计属于在顾客原始登记信息上的处理、分析成果，已经不再是原始的数据，而成为信息，揭示了商场顾客的年龄分布特征。

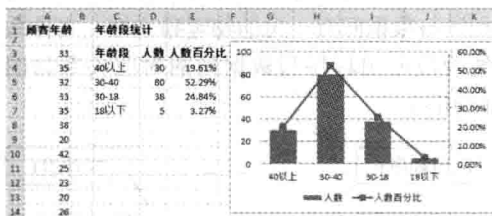


图1-3 顾客年龄段的人数统计图表

要对商场最终的营销计划发挥更大辅助作用，仅靠图1-3这类的描述性统计图表提供的信息还不够深入，需要从更多维度对数据进行处理和分析，如图1-4所示。

	A	B	C	D	E	F	G	H
1	顾客编号	年龄	性别	商品	货架	消费额		
2	gk201310001	33	女	A	hj1	300		
3	gk201310002	35	男	B	hj2	100		
4								
5	所有顾客在A商品上消费的总额：4125元							
6	从货架hj2购买商品顾客平均年龄：33.5岁							
7	从货架hj1购买了A商品的顾客中多少人会购买hj2的B商品：20%							
8	女性顾客平均每次购物的购买额：206.23元							
9	男性顾客平均每次购物的购买额：315.78元							

图1-4 基于顾客消费数据获得的BI

根据顾客消费数据，图1-4从顾客的性别、年龄、商品、货架等维度进行分析，得出了诸如“从货架hj1购买了A商品的顾客中20%的人会购买hj2的B商品”、“女性顾客平均每次购物的购买额为206.23元”的发现，这些发现即为商业智能，是决策制定过程中重要的依据。

由上述说明可知，数据是人们使用各种工具和手段观察外部世界所得到的原始材料，只描述发生了什么事情，并不能构成决策或行动的可靠基础。通过对数据的分析，找出其中包含的关系，赋予数据以某种意义和关联，便形成了信息。虽然信息展现出数据中一些有意义的东西，但其往往和需要完成的任务没有直接联系，也不能作为判断、决策和行动的依据。这便需要对信息进行再加工，进行更深入的归纳分析，从信息中理解其模式，以获得更有用的信息，即智能。商业智能则是以商业决策相关的变量作为分析维度，从数据中挖掘和产生的规律性认识。

商业智能的实现，依赖于数据仓库（Data Warehousing）、联机分析处理（On-Line Analytical Processing, OLAP）工具和数据挖掘（Data Mining）等技术。提供商业智能解决方案的著名IT厂商有微软、IBM、Oracle、SAP、Informatica、Microstrategy、SAS、Royalsoft等。

1.1.4 数据处理

由数据到智能的转变离不开数据处理过程。数据处理是指将数据转化成信息的一系列数据搜集、整理、组织、加工、存储、

维护与传递的活动总和。数据处理的目的是从大量的、无序的、难以理解的数据中抽取并衍生出有意义、有价值的信息。数据处理贯穿于社会生产和生活的各个阶段、各个领域，包括人工、半自动化及全自动化等多种处理方式。在数据处理活动中，数字运算的计算过程相对容易，而一旦涉及复杂的数理统计模型、数据量大并且数据之间的逻辑关系复杂时，数据处理过程变得更困难，此时成功的数据处理不仅在于快速高效的运算性能，很大程度上还依赖良好规范的数据管理机制和平台。数据管理技术的发展及其应用的广度和深度极大地影响着人类社会的发展进程。数据管理技术的发展历经人工管理、文件管理、数据库系统等阶段。

1. 人工管理阶段（20世纪50年代中期以前）

人工管理阶段的计算机主要用于科学计算，数据处理的方式是批处理。这一阶段数据管理的主要形式如图1-5所示。



图1-5 人工管理阶段应用程序与数据之间的对应关系

人工管理阶段的数据管理特点如下：

(1) 数据不作保留。因为计算机的主要工作是科学计算，数据只在计算时进行输入，使用后一般不进行专门保存。

(2) 通过应用程序管理数据。需要开发应用程序进行数据的管理，没有相应的软件系统负责数据的管理工作。针对各数据集的特点，专门编写应用程序进行读取、操作和管理，没有软件系统负责统一的数据管理。应用程序与数据之间的关系如图1-5所示。

(3) 数据不共享。数据面向应用程序，一组数据对应专门的应用程序，应用程序与应用程序之间存在大量冗余。

(4) 数据不具有独立性。数据与应用程序之间具有紧密的依赖性，当数据的逻辑结构或物理结构发生变化时，应用程序必须做出相应修改，加大了应用程序维护成本。

2. 文件系统阶段（20世纪50年代后期到60年代中期）

在文件系统阶段，计算机硬件方面出现了磁盘、磁鼓等直接存取存储设备，在软件方面基于操作系统开发了专门的数据管理软件，一般称为文件系统；数据处理方式上不仅有了批处理，而且能够联机实时处理。文件系统阶段的数据管理特点如下：

(1) 数据可以长期保存。数据处理具有更复杂、多样化的要求，数据需要长期保留在外存上反复进行查询、修改、插入和删除等操作。

(2) 通过文件系统管理数据。文件系统把数据组织成相互独立的文件，文件内部数据按照记录的方式进行组织。尽管这一方式实现了文件内的结构化组织，在整体上还属于无结构，即多个文件之间相互独立，无法建立全局的结构化数据管理模式。

(3) 有一定的物理独立性。应用程序与数据之间由文件系统提供的存取方法进行转换，程序员不必过多考虑物理细节，应用程序与数据间有一定的物理独立性。文件系统阶段应用程序与数据间的对应关系如图1-6所示。

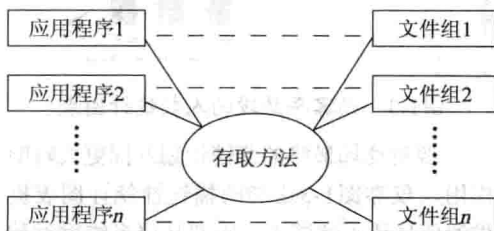


图1-6 文件系统阶段应用程序与数据之间的对应关系