

计算机科学与技术专业本科系列教材



11001010100110101010
1.1.10.101001010100101.1010

数据库 技术与应用

——基于SQL Server 2019与MongoDB

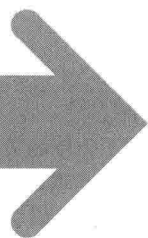
主 编 赵有恩

副主编 高珊珊 耿长欣 张 燕



重庆大学 出版社

计算机科学与技术专业本科系列教材



数据库 技术与应用

——基于SQL Server 2019与MongoDB

主 编 赵有恩

副主编 高珊珊 耿长欣 张 燕

重庆大学出版社

内 容 提 要

本书除了介绍了传统的关系数据库——SQL Server 2019 的基本理论与应用以外,也介绍了非关系数据库——MongoDB 的基本理论与应用,更符合技术发展的趋势与大数据时代下的人才培养需求。全书共 8 章内容,包括数据库基础知识、数据模型与关系数据库、关系数据库标准语言 SQL、关系数据库规范化理论、MongoDB 数据库基础、数据库的安全与维护、数据库设计和 SQL Server 2019 基础。本教材面向数据库初学者,可作为高等院校计算机科学与技术及相关专业的数据库课程教材,也可作为数据库培训班的培训教材,还可以作为数据库应用程序开发人员的参考资料。

图书在版编目(CIP)数据

数据库技术与应用:基于 SQL Server 2019 与
MongoDB / 赵有恩主编. -- 重庆:重庆大学出版社,
2022.3

计算机科学与技术专业本科系列教材

ISBN 978-7-5689-3186-1

I. ①数… II. ①赵… 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(2022)第 046678 号

数据库技术与应用——基于 SQL Server 2019 与 MongoDB SHUJUKU JISHU YU YINGYONG — JIYU SQL Server 2019 YU MongoDB

主 编 赵有恩

副主编 高珊珊 耿长欣 张 燕

特约编辑:陈 丹

责任编辑:杨粮菊 版式设计:杨粮菊

责任校对:王 倩 责任印制:张 策

*

重庆大学出版社出版发行

出版人:饶帮华

社址:重庆市沙坪坝区大学城西路 21 号

邮编:401331

电话:(023) 88617190 88617185(中小学)

传真:(023) 88617186 88617166

网址:<http://www.cqup.com.cn>

邮箱:fxk@cqup.com.cn (营销中心)

全国新华书店经销

重庆华林天美印务有限公司印刷

*

开本:787mm×1092mm 1/16 印张:18.25 字数:435千

2022 年 3 月第 1 版 2022 年 3 月第 1 次印刷

印数:1—2 000

ISBN 978-7-5689-3186-1 定价:49.00 元

本书如有印刷、装订等质量问题,本社负责调换
版权所有,请勿擅自翻印和用本书制作
各类出版物及配套用书,违者必究

前言

21 世纪已经成为“数据”的世界,无处不在的物联网和传感器,使数据/信息呈爆炸式增长。在数据/信息的快速传输、高效计算、海量存储等方面不断发展的前提条件下,数据成了未来社会发展的一种崭新的不可见资源。过去常说的“信息爆炸”“海量数据”已经成为事实,未来越来越多的 IT 基础架构将会部署在公有云、私有云或者混合云上,而数据库作为架构中最重要的基础部分,也将发挥越来越重要的作用。

数据库技术就是研究各类数据的结构、存储、设计、管理和使用的一门学科。数据库技术自 20 世纪 60 年代中期产生以来,已成为计算机领域发展最快的学科分支之一,也是应用最广泛的技术之一。数据库技术作为数据管理的最有效手段,已经成为各类信息系统的核心技术和基础,广泛应用到教育、医疗、商业、科学研究、行政管理等行业。例如各种类型的管理信息系统及 App、电子商务与电子政务、大中型网站、决策支持系统、企业资源规划、客户关系管理、数据仓库与数据挖掘等,都是以数据库技术作为重要的支撑。可以说,只要有计算机存在,就有数据库技术存在;只要有系统存在,就有数据库技术存在。因此,数据库的建设规模、应用深度成为衡量一个国家信息化程度的重要标志。在 market 需求的驱动下,数据库技术成为当前高校计算机及相关专业的必修课程以及非计算机专业选修的核心课程之一。

在云计算和大数据时代,传统的面向结构化数据的关系数据库已不再是一枝独秀,各种 NoSQL 数据库不断涌现。NoSQL 数据库面向半结构化和非结构化的数据,不进行类似关系数据库的复杂的处理,弥补了传统关系数据库的不足。基于此,本书结合目前流行的关系数据库 SQL Server 2019 和非关系数据库 MongoDB,以 21 世纪对数据人才培养的需求为突破点,着力培养以数据的观点来看待世界,描述和解释世界,进而改造世界,并且能够理解数据、组织数据、管理数据、展示数据、应用数据的复合型人才。

本书内容包括 8 章。第 1 章为数据库基础知识。从数据、大数据的基本概念入手,介绍数据管理技术发展的 5 个阶段,从而引出数据库的概念。阐述了数据库系统的组成与数据

库有关的名词术语,最后介绍了数据库系统的三级模式结构和二级映像功能。

第2章数据模型与关系数据库。从数据模型的三要素:数据结构,数据操作,完整性约束3个方面介绍了层次模型、网状模型和关系模型,最后重点介绍了关系模型的数据结构、数据操作、完整性约束。

第3章关系数据库标准语言SQL。从数据定义,数据操纵,数据检索3个方面介绍了SQL语句的使用,包括创建数据库、创建表、数据的增删改查、Transact-SQL的流程控制以及视图,索引、存储过程、触发器等数据库对象的创建与使用。

第4章关系数据库规范化理论。介绍了函数依赖、关系数据库的范式及规范化步骤。

第5章MongoDB数据库基础。介绍了NoSQL数据库管理系统MongoDB的安装、访问、管理(包括数据库的管理、集合和文档的管理)以及数据查询、索引、数据的导入与导出。

第6章数据库的安全与维护。从数据库的完整性控制、并发控制、数据库的备份与恢复介绍了SQL Server 2019和MongoDB在数据库安全性方面的实现与操作。

第7章数据库设计。从需求分析、概念结构设计、逻辑结构设计、物理设计、实施、运行与维护等方面介绍了数据库设计的整个流程。

第8章SQL Server 2019基础。介绍了Microsoft SQL Server 2019安装、登录、系统数据库以及主要管理工具的使用。

本教材的主要特色如下:

1)将关系数据库和非关系数据库的理论和操作结合进行介绍。结合目前流行的关系数据库SQL Server 2019和非关系数据库MongoDB,通过对比,强化理解关系数据库(对应结构化数据)和非关系数据库(对应于半结构化和非结构化数据)的不同本质,因本质不同导致了二者应用领域的不同。

2)数据库理论与实际应用相对照。在注重理论性、系统性和科学性的同时,兼顾理论的具体实现,以培养学生解决实际问题的能力。比如,在讲第2章数据模型与关系数据库的理论——数据模型的三要素时,以关系模型为例,讲解关系数据库在数据模型上的具体实现;在讲第6章数据库的安全与维护理论时,也讲解SQL Server和MongoDB在数据库安全理论上的具体实现。

本书可作为高等学校计算机相关专业本科生和专科生的教材,也可作为从事相关专业的工程技术人员的参考用书。

在本书的编写过程中,参考了国内外大量的数据库技术的书刊及文献资料,在此一并对资料的作者表示感谢,主要的参考书籍和学术论文在书后的参考文献列出。书中全部的Transact-SQL(T-SQL)语句、MongoDB shell 命令都已上机调试通过。

在编写过程中,编写团队开展了多次交流与研讨,对书稿进行了多次修改与完善,它的完成凝聚了所有作者的心血与智慧,凝聚了一个团队的教学成果。但由于编者水平和时间所限,书中疏漏之处在所难免,恳请同行专家和读者批评指正。

本书在编写过程中,得到了山东财经大学计算机科学与技术学院的相关领导、同事以及朋友、家人的大力支持与帮助,在此一并表示诚挚的谢意!本书的编写得到了山东省高等学校青创人才引育计划项目、山东财经大学“数据库原理及应用”课程思政项目以及山东财经大学计算机科学与技术学院教材建设专项经费的资助,特此致谢!

赵有恩

2021年7月于济南

本书微课视频清单

序号	名称	二维码图形	序号	名称	二维码图形
1	1.4 数据库系统的体系结构		12	3.3 数据操纵	
2	2.1.2 概念模型		13	3.4.1 单表查询	
3	2.1.3 数据模型		14	3.4.1 单表查询的选择表中的若干元组	
4	2.2 关系数据结构及其形式化定义		15	3.4.1 单表查询的排序分组查询	
5	2.3 关系的完整性约束		16	3.4.2 连接查询	
6	2.4.1 传统的集合运算		17	3.4.3 嵌套查询	
7	2.4.2 专门的关系运算		18	3.5.2 定义视图	
8	3.2.1 数据库的定义		19	3.5.2 定义视图之修改删除视图	
9	3.2.1 数据库的定义之修改删除数据库		20	3.6 索引的定义作用与分类	
10	3.2.2 基本表的定义		21	3.6 索引的创建删除修改	
11	3.2.2 基本表的定义之修改删除表		22	3.8 存储过程的创建执行	

续表

序号	名称	二维码图形	序号	名称	二维码图形
23	3.8 存储过程的分类和修改删除		27	4.2 函数依赖	
24	3.9 触发器的定义创建与分类		28	4.3 关系模式的规范化	
25	3.9 触发器的删除与性能		29	4.5 关系模式规范化步骤	
26	3.9 之后触发器与替代触发器		30	7.3.2 逻辑结构设计	

目 录

第 1 章 数据库基础知识	1
1.1 数据与信息、大数据	1
1.2 数据管理技术的产生与发展	5
1.3 数据库系统的组成	13
1.4 数据库系统的体系结构	16
1.5 数据库技术的研究领域及常见的数据库管理系统简介	21
练习题	24
第 2 章 数据模型与关系数据库	25
2.1 数据模型	25
2.2 关系数据结构及其形式化定义	35
2.3 关系的完整性约束	38
2.4 关系代数	40
练习题	47
第 3 章 关系数据库标准语言 SQL	50
3.1 SQL 概述	50
3.2 数据定义	53
3.3 数据操纵	77
3.4 数据查询	82
3.5 视图.....	100
3.6 索引.....	106
3.7 Transact-SQL	112
3.8 存储过程.....	137
3.9 触发器.....	144
练习题	150
第 4 章 关系数据库规范化理论	156
4.1 关系模式的规范化问题.....	156
4.2 函数依赖.....	159
4.3 关系模式的规范化.....	162

4.4	多值依赖与 4NF	167
4.5	关系模式规范化步骤	170
	练习题	171
第 5 章	MongoDB 数据库基础	173
5.1	MongoDB 数据库简介	173
5.2	MongoDB 的安装	177
5.3	用 MongoDB shell 访问 MongoDB	180
5.4	MongoDB 数据库的管理	182
5.5	MongoDB 集合的管理	183
5.6	MongoDB 文档的管理	184
5.7	MongoDB 的查询	187
5.8	MongoDB 索引的管理	189
5.9	数据的导入与导出	194
	练习题	195
第 6 章	数据库的安全与维护	197
6.1	数据库的安全性控制	197
6.2	数据库的完整性控制	214
6.3	数据库的并发控制	219
6.4	数据库的备份与恢复	231
	练习题	237
第 7 章	数据库设计	238
7.1	数据库设计概述	238
7.2	数据库需求分析	240
7.3	数据库结构设计	243
7.4	数据库行为设计	251
7.5	数据库实施	253
7.6	数据库运行和维护	255
	练习题	255
第 8 章	SQL Server 2019 基础	257
8.1	SQL Server 2019 简介	257
8.2	SQL Server 2019 的安装	258
8.3	SQL Server 2019 的管理工具和联机丛书	271
8.4	SQL Server 2019 的登录	273
8.5	SQL Server 2019 系统数据库	278
	参考文献	280

第 1 章

数据库基础知识

21 世纪是信息的世纪,信息已变成一种资源,成为人们生活中不可或缺的重要组成部分。对于企业而言,信息资源获取的多与少,资源管理的好与坏,直接决定着企业在激烈的竞争中能否成功。数据库技术作为信息系统的一个核心技术,是一种专门用于处理数据和信息的技术,它产生于 20 世纪 60 年代末,是一门应用广泛、实用性强的技术,也是计算机科学与技术的重要分支,它的出现极大地促进了计算机应用向各行各业的渗透。例如,它是信息管理系统(Management Information System, MIS)、办公自动化系统(Office Automation, OA)、企业资源规划(Enterprise Resource Planning, ERP)、决策支持系统(Decision Support System, DSS)等各类信息管理的核心部分,是进行数据资源共享、科学研究和决策管理的重要手段。

在云计算和大数据时代,传统的关系数据库已经不再是一枝独秀,各种 NoSQL(Not Only SQL)数据库也不断涌现。过去人们常说的“信息爆炸”“海量数据”等词汇已经不足以描述当今信息社会,于是出现了大数据和数据科学。未来越来越多的 IT 基础架构都会部署在“云”上,而数据库作为云架构中的重要组成部分,与云的结合将会变得越来越重要。

本章主要介绍与数据库(包括关系数据库和 NoSQL 数据库)有关的一些基本概念和术语,这些内容是学习后续各章节的基础。

1.1 数据与信息、大数据

数据、数据库、数据库管理系统、数据库系统是与数据库技术密切相关的 4 个基本又相互关联的概念,是首先要认识的。

1.1.1 数据与信息

1) 数据

数据(Data)是反映客观事物属性的记录,是用于表示客观事物的未经加工的原始资料,

是信息的载体。对客观事物属性(特征)的记录是用一定的物理符号(如数字、符号、声音、图形、图像、视频等)来表达的,所以说数据是信息的具体表现形式。

数据的概念包括数据形式和数据内容两个方面。数据形式是指数据内容存储在媒体上的具体形式(物理符号形式),即通常所说的数据的“类型”。数据内容是指所描述客观事物的具体属性,即通常所说的数据的“值”。

例如,学生的基本信息“姓名”“性别”等属性用字符型数据形式描述,“年龄”属性用数值型数据形式描述,“照片”属性用二进制型数据形式描述。

而对于一个具体的学生来讲:

其“姓名”的值为“王义”;

“性别”的值为“男”;

“年龄”的值为“19”;

“照片”的值为其照片文件。根据数据管理和数据处理的具体要求,可以选择不同的数据形式来表示。例如,性别这一数据,可以用“男”“女”文字表示,也可用 0、1 数字表示等。

(1) 数据的单位

在计算机中,衡量数据大小的单位是字节(Byte),用大写的英文字母 B 表示,依次还有 KB、MB、GB、TB、PB、EB、ZB、YB、BB、NB、DB,它们之间按照进率 $1\ 024(2^{10})$ 来计算:

$1\text{ KB} = 1\ 024\text{ B} = 2^{10}\text{ B}$

$1\text{ MB} = 1\ 024\text{ KB} = 2^{20}\text{ B} = 1\ 048\ 576\text{ B}$

$1\text{ GB} = 1\ 024\text{ MB} = 2^{30}\text{ B} = 1\ 048\ 576\text{ KB}$

$1\text{ TB} = 1\ 024\text{ GB} = 2^{40}\text{ B} = 1\ 048\ 576\text{ MB}$

$1\text{ PB} = 1\ 024\text{ TB} = 2^{50}\text{ B} = 1\ 048\ 576\text{ GB}$

$1\text{ EB} = 1\ 024\text{ PB} = 2^{60}\text{ B} = 1\ 048\ 576\text{ TB}$

...

(2) 数据的分类

从结构上来说,数据可以分为 3 大类。第一类是能够用统一的结构表示,称为结构化数据,如数字、符号等;第二类是无法用统一的结构表示,称为非结构化数据,如文本、图像、声音等;第三类是介于结构化数据和非结构化数据之间的数据,如 HTML 文档等。

①结构化数据是指数据经过分析后可以分解成多个互相关联的组成部分,各个组成部分间有明确的层次结构,其使用和维护通过数据库进行管理,并有一定的操作规范。通常接触的包括生产、业务、交易、客户信息等这类数据都属于结构化数据,可以用二维表结构来表达实现,比如企业 ERP、财务系统、医疗 HIS 数据、教育一卡通、政府行政审批等。

②非结构化数据,所谓的非结构化数据是指数据的变长记录由若干不可重复和可重复的字段组成,不方便使用二维逻辑表结构来表现。简单地说,非结构化数据就是字段可变的数据,支持重复字段、子字段以及变长字段。比如日常生活中的办公文档、图片、音频/视频等数据。

③半结构化数据是介于结构化数据和非结构化数据之间的数据,如 HTML 文档等。它一般是自描述的,数据的结构和内容混在一起,没有明显的区分。

表 1-1 从数据模型、形成过程等方面说明了 3 类数据的区别。

表 1-1 3 类数据的区别

比较项目	结构化数据	半结构化数据	非结构化数据
数据模型	二维表	树、图	无
形成过程	先有结构,再有数据	先有数据,再有结构	先有数据,再有结构
形式	数字、符号等	文本、图像、声音等	HTML 文档等

2) 信息

信息(Information)解释为可通信的事情、知识、消息等。信息是人类的一切生存活动和自然存在所传达出来的消息和知识。实际上,信息是客观事物属性(特征)的反映,所反映的是关于某一客观系统中某一事物的某些方面属性或某一时刻的表现形式,为人类带来客观世界的认识和知识,如姓名、性别、年龄等反映了一个人的基本信息。人类社会之所以如此丰富多彩,都是因为信息和信息技术一直持续进步的必然结果。

由上面定义可知,数据是信息的表达形式,信息是数据所表达的有用含义。信息是通过数据来传播的,不具有知识性和有用性的数据则不能称为信息。

如果说结构化数据产生的信息详细记录了企业的生产交易活动,那么非结构化数据产生的信息则隐性包含了提高企业效益的机会。对大多数企业来说,ERP 等业务系统所处理的结构化数据信息只占到企业全部信息的 10%左右,其他的 90%都是非结构数据信息。

1.1.2 大数据

2012 年以来,大数据(Big Data)一词越来越多地被人们提及与使用,人们用它来定义信息时代产生的海量数据,并命名与之相关的技术发展和创新。那到底什么是大数据?大数据与数据库领域的超大规模数据库(Very Large Database, VLDB)、海量数据(Massive Data)有什么区别?

“超大规模数据库”这个词是在 20 世纪 70 年代中期出现的,是指数据库中管理的数据集有数百万条记录。“海量数据”则是在 21 世纪初出现的词,用来描述更大的数据集以及更丰富的数据类型。2008 年 9 月,《科学》(Science)杂志发表了一篇名为“Big Data: Science in the Petabyte Era”的文章,“大数据”这个词开始被广泛传播。

无论是“超大规模数据库”“海量数据”,还是“大数据”,这些词都表示需要管理的数据规模很大,已经超出了当时的计算机存储和处理技术水平,需要计算机界研究和更加先进的技术才能更有效地存储、管理和分析它们。

(1) 大数据的定义

对于大数据,不同的研究机构基于不同的角度给出了不同的定义:

高德纳(Gartner)咨询有限公司给出了这样的定义:“大数据”是需要新的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的海量、高增长率和多样化的信息资产。

全球著名的管理公司麦肯锡给出的定义:一种规模大到在获取、存储、管理、分析方面极

大超出了传统数据库软件工具能力范围的集合,具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低等特点。

国际数据公司(International Data Group, IDG)给出的定义:大数据一般会涉及两种或两种以上的数据形式,它需要收集超过 100 TB 的数据,并且是高速实时数据流;或者是从小数据开始,但数据每年的增长率至少为 60%。

2015 年 8 月,国务院正式印发了《促进大数据发展行动纲要》(以下简称《行动纲要》),成为我国发展大数据产业的战略指导性文件。《行动纲要》指出:“大数据是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合,正快速发展成为对数量巨大、来源分散、格式多样的数据进行采集,存储和关联分析,从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态。”同时,中国信息通信研究院相继发布了《大数据白皮书(2014 年)》《大数据白皮书(2016 年)》和《大数据白皮书(2018 年)》等系列白皮书。《大数据白皮书 2016》称:“大数据是新资源、新技术和新理念的混合体。从资源的视角看,大数据是新资源,体现了一种全新的资源观;从技术的视角看,大数据代表了新一代数据管理和分析技术;从理念的视角看,大数据打开了一种全新的思维角度。”

总结以上对大数据的不同定义,不难发现大数据的概念具有两点共性:

①大数据的数据量标准是随着计算机软/硬件的发展而不断增长。例如 1 GB 的数据量在 20 年前可以称为大数据,而今的数据量已经达到 TB 或者 GB 量级,却不能称为大数据。

②大数据不仅体现在数据规模上,还包含了数据来自多种数据源,包括结构化数据、半结构化数据和非结构化数据,并且以实时、迭代的方式来更新。

总的来说,大数据是指所涉及的数据规模或者复杂程度超出了传统数据库技术和软件技术所能管理和处理的数据集范围。大数据通常与 Hadoop、NoSQL、数据分析与挖掘、数据仓库、商业智能以及开源云计算架构等诸多热点话题联系在一起。

(2) 大数据的特征

IBM 公司将大数据的特征归纳为 5 个 V,即 Volume(容量大)、Variety(多样性)、Velocity(存取速度快)、Value(低价值密度)、Veracity(真实性)。

①容量大:大数据的首要特征是容量大,而且在持续、急剧地增长。

②多样性:海量数据引发的危机不单纯是数据量的爆炸增长,还涉及数据的多样性,包括数据格式的多样性,不仅包含文字、数字、日期等结构化数据,还包括图形、图像、音频、视频、地理位置等非结构和半结构化的数据;同时数据来源多样,包括互联网应用、电子商务领域、电子运营商、全球定位系统、社交网络、各种传感器数据等。

③存取速度快:大数据的存取速度快(也称为实时性),一方面是指数据增长的速度特别快,另一方面是指数据处理的速度快,能实时进行分析和处理。数据处理遵循“1 秒定律”,可从各种类型的数据中快速获得高价值信息。

④低价值密度:大数据的价值是潜在的、巨大的。但在大数据中,价值密度的高低与数据总量并不存在线性关系,有价值的信息往往被淹没在海量的无用数据中。例如,在一段长达几小时的连续不断的视频监控中,可能有用的数据仅仅只有几秒。因此,如何从海量数据中洞察有价值的信息成为大数据研究的重要课题。

⑤真实性:真实性指的是当数据的来源变得多样时,这些数据本身的可靠度、质量是否足够。如果数据本身就是有问题的,那么分析后的结构也不会是正确的。真实性旨在针对大数据噪音、数据缺失、数据不确定性等问题,强调数据质量的重要性,以及保证数据质量所面临的巨大挑战。

传统数据与大数据的区别见表 1-2。

表 1-2 传统数据与大数据的区别

比较项目	传统数据	大数据
数据规模	规模小,以 MB、GB 为处理单位	规模大,以 TB、PB 为处理单位
数据增大速度	每小时,每天	每分,每秒
数据结构类型	单一的结构化数据	结构化、非结构化、半结构化数据
数据来源	集中的数据源	分布式的数据源
数据存储	关系数据库管理系统	分布式文件系统、NoSQL 数据库
模式与数据的关系	先有模式,后有数据	先有数据后有模式,且模式随着数据不断演变
处理对象	数据仅作为被处理对象	作为被处理对象或者辅助资源来解决其他领域问题
处理工具	一种或少数几种处理工具	不存在单一的处理工具

1.2 数据管理技术的产生与发展

数据管理是指对数据进行采集、分类、组织、编码、存储、检索和维护的工作,是数据处理的核心。所谓数据管理技术就是指数据管理过程中所采用的技术,其发展是随着计算机硬件、系统软件以及计算机应用的发展而不断进步。这个过程基本上可以划分为 5 个阶段,即人工管理阶段、文件系统管理阶段、数据库系统管理阶段、高级数据库系统管理阶段、新兴数据管理阶段。在计算机技术的发展和应用需求的推动下,每一阶段的发展都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。

1.2.1 人工管理阶段

20 世纪 50 年代中期之前,在计算机技术发展初级阶段,因为只有磁带、卡片、纸带等顺序外存储设备,无操作系统,也没有专门管理数据的软件,所以这一时期,程序员编写的程序 and 要处理的数据写在一起,使得一组数据只对应一个应用程序,如图 1-1 所示。

人工管理阶段的特点如下:

①数据不保存。计算机主要用于科学计算,不要求保存数据。每次启动计算机后都要将程序和数据输入主存,计算结束后将结果输出。计算机断电后,计算结果会随之消失。

②数据面向程序。每个程序都有属于自己的一组数据,程序和数据相互结合成为一体,互相依赖。各程序之间的数据不能共享,因此数据就会重复存储,冗余度大。

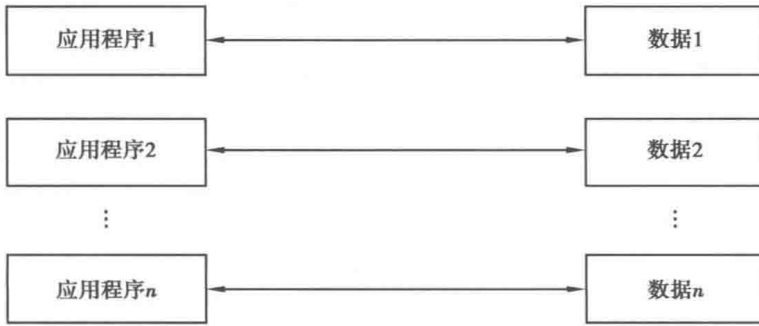


图 1-1 人工管理阶段应用程序与数据的关系示例

③编写程序时要安排数据的物理存储。程序员除了编写程序,还要安排数据的物理存储。程序和数据混合一体,一旦数据的物理存储改变,就必须重新编程,程序员的工作量大而烦琐,程序难以维护。

1.2.2 文件系统管理阶段

随着计算机技术的发展,在 20 世纪 50 年代后期到 60 年代中期,计算机大量用于数据处理等方面,数据需要长期保留在外存上反复处理,如维护、查询等。在这一时期,有了磁盘、磁鼓等大容量、读写速度快的外存储设备。操作系统也为管理数据提供了专门的软件,即文件系统。

数据与应用程序分离,数据独立存放在数据文件中,有了随机文件、链接文件、索引文件等多种高效的文件组织形式。应用程序通过文件系统与数据文件发生联系,程序和数据之间有存取方法进行转换,数据在物理结构和逻辑结构间进行转换,提高了数据的物理独立性,如图 1-2 所示。

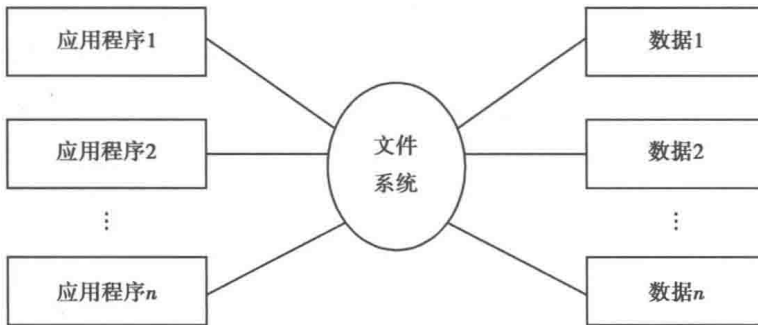


图 1-2 文件系统管理阶段应用程序与数据文件的关系

文件系统管理阶段的特点如下:

①数据以文件的形式长期保存。此阶段计算机大量用于数据处理,数据需要长期保存在外存上反复处理,即要经常对其进行查询、修改、插入和删除,因此,在文件系统中,将数据按照一定的规则组织成一个文件,长期存放在外存储器中。

②数据的物理结构和逻辑结构有了区别,但比较简单。程序员只需用文件名与数据打交道,不必关心数据的物理位置,可由文件系统提供的读写方法去读/写数据。

③文件形式多样化。为了方便数据的存储和查找,人们研究开发了许多文件类型,如索

引文件、链式文件、顺序文件、倒排文件等。

④程序与数据之间有一定的独立性。应用程序通过文件系统对数据文件中的数据进行存取和加工,文件系统充当应用程序和数据之间的一种接口。这样,可使应用程序和数据都具有一定的独立性。因此,处理数据时程序不必过多地考虑数据的物理存储细节,程序员可以集中精力于算法设计。

尽管文件系统有上述优点,但是这些数据在数据文件中只是简单地存放,文件之间并没有有机联系,仍不能表示复杂的数据结构;数据的存放仍依赖于应用程序的使用方法,基本上是一个数据文件对应一个或几个应用程序;数据面向应用,独立性较差,仍然会出现数据重复存储、冗余度大、一致性差(同一数据在不同文件中的值不一样)等问题。

1.2.3 数据库系统管理阶段

20世纪60年代后期以来,计算机广泛应用管理领域,由于管理规模逐渐庞大,数据量急剧增加,多应用、多语言程序需要共享数据集合,因此数据共享的要求越来越高,如图1-3所示。

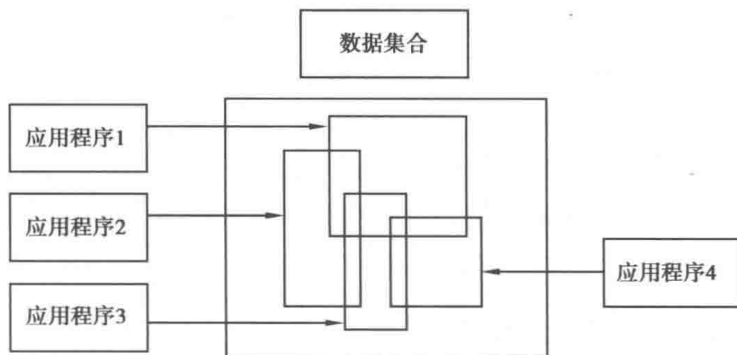


图 1-3 数据共享

此时,磁盘技术取得了重大进展,为数据库技术的发展提供了物质条件。人们开发出一种新的数据管理方法,将数据存储于数据库中,由数据库管理软件对其进行管理。

数据库系统管理方式即对所有的数据实行统一规划管理,形成一个数据中心,构成一个数据仓库,数据库中的数据能够满足所有用户的不同要求,供不同用户使用,如图1-4所示。

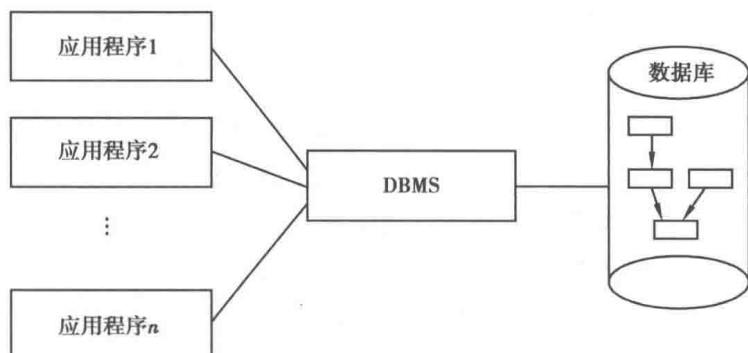


图 1-4 数据库系统管理阶段应用程序与数据文件的关系

数据库系统管理阶段的特点如下:

①数据共享。这是数据库系统区别于文件系统的最大特点之一,也是数据库系统技术先