



CCF 大数据教材系列丛书  
CCF 大数据专家委员会 组编

主编 杜小勇

# 大 数 据

# 管 理

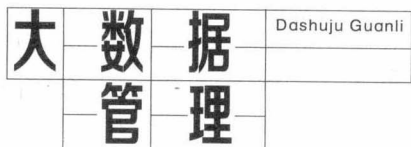
高等教育出版社



CCF 大数据教材系列丛书  
CCF 大数据专家委员会 组编

主编 杜小勇

大	数	据
管	理	



图书在版编目(CIP)数据

大数据管理 / 杜小勇主编. -- 北京: 高等教育出版社, 2019.3

ISBN 978-7-04-051561-9

I. ①大… II. ①杜… III. ①数据管理-高等学校-教材 IV. ①TP274

中国版本图书馆CIP数据核字(2019)第040956号

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。

为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话

(010) 58581999 58582371  
58582488

反盗版举报传真

(010) 82086060

反盗版举报邮箱

dd@hep.com.cn

通信地址

北京市西城区德外大街4号

高等教育出版社法律事务

与版权管理部

邮政编码 100120

防伪查询说明

用户购书后刮开封底防伪涂层，利用手机微信等软件扫描二维码，会跳转至防伪查询网页，获得所购图书详细信息。也可将防伪二维码下的20位密码按从左到右、从上到下的顺序发送短信至106695881280，免费查询所购图书真伪。

反盗版短信举报

编辑短信“JB, 图书名称, 出版社, 购买地点”发送至10669588128

防伪客服电话

(010) 58582300

策划编辑 张江漫

责任编辑 黄涵玥

书籍设计 张申申

插图绘制 于博

责任校对 张薇

责任印制 田甜

出版发行 高等教育出版社

社址 北京市西城区德外大街4号

邮政编码 100120

购书热线 010-58581118

咨询电话 400-810-0598

网址 <http://www.hep.edu.cn>

<http://www.hep.com.cn>

网上订购

<http://www.hepmall.com.cn>

<http://www.hepmall.com>

<http://www.hepmall.cn>

印刷 北京信彩瑞禾印刷厂

开本 850mm×1168mm 1/16

印张 20.25

字数 310千字

版次 2019年3月第1版

印次 2019年3月第1次印刷

定价 48.00元

本书如有缺页、倒页、脱页等质量问题，请到所购图书销售部门联系调换

版权所有 侵权必究

物料号 51561-00

## 内容提要

本书系统地全面地阐述了大数据管理系统的基础理论、基本技术和基本方法。全书分为三篇共10章。第一篇大数据管理概述,综述了数据库管理系统发展经历或正在经历的四个阶段,梳理了大数据管理系统的数据特征、系统特征、应用特征,阐述了大数据管理系统的组成,指出发展大数据管理系统是历史的必然;第二篇数据模型与语言,包括关系模型与SQL、键值对数据模型、文档模型与查询语言、图模型与类SQL查询语言,共4章;第三篇大数据管理系统,包括大数据管理系统的体系架构、数据组织与存储、分布式查询处理优化、分布式事务、故障恢复,共5章。

本书可以作为高等学校大数据专业、计算机类专业、信息管理与管理信息系统等相关专业大数据管理课程的教材,也可供从事数据库和大数据管理系统研究、开发和应用的工程技术人员和工程技术人员参考。

# 大数据教材系列丛书编委会

主任 梅 宏

成员 (按姓名拼音排序)

卜佳俊 陈宝权 陈恩红

程学旗 杜小勇 方 粮

胡 斌 黄宜华 金 海

马华东 潘柱廷 王建民

王晓阳 王元卓 袁晓如

周傲英 周 涛 周晓方

随着大数据的蓬勃发展,大数据领域人才的需求越来越大,大数据人才培养受到了各界的广泛关注。2016年,教育部开始批准设立“数据科学与大数据技术”本科专业,越来越多的高校申请开设“数据科学与大数据技术”专业或开设大数据方向的相关课程,截至2018年3月,已有近三百所高校获批建设“数据科学与大数据技术”专业。虽然大数据专业和大数据方向的课程不断开设,但是,当前我国高校的大数据教学尚处在摸索阶段,尤其缺乏成熟的、系统性和规范性的大数据教学体系和教材。

在此背景下,中国计算机学会大数据专家委员会成立了大数据教材系列丛书编委会,着手编著系列化、规范化的大数据教材。自2017年6月,经编委会多次研讨,形成丛书框架,作者们随即开始紧张的编写工作。编委会和作者间也有多轮的初稿审阅和研讨交流。数易其稿,终于付梓。

大数据教材系列丛书采用“1+3+X”的体系,即以1本《大数据导论》为基础,设置《大数据管理》《大数据处理》《大数据分析》3本关键技术教材,以及针对行业领域的X本应用教材。本套教材系列丛书既适合高校大数据专业的专科生、本科生以及研究生系统地学习大数据相关知识与技术,也适合从事大数据相关技术的企事业单位研究人员、工程师作为参考用书。

《大数据导论》是一本全面介绍大数据相关知识的专业通识教材,其系统地介绍大数据涵盖的内容,包括数据与大数据、大数据获取与感知、大数据存储与管理、大数据分析、大数据处理、大数据治理、大数据安全与隐私等,同时还介绍了部分行业中大数据的典型应用案例,反映了大数据在社会经济生活中的重要价值。

《大数据管理》首先综述数据管理系统的发展,指出发展大数据管理系统是历史的必然,并沿着数据模型和系统构件两个维度上展开。在数据模型的维度上,主要介绍关系、键值对、图和文档数据模型及其语言;在系统维度上,介绍系统结构、存储与组织、查询处理、事务管理、故障恢复等话题。

《大数据处理》包括大数据处理基础技术、大数据处理编程与典型应用处理、大数据处理系统与优化三个方面。本教材以大数据处理编程为核心,从基础、编程到优化等多个方面对大数据处理技术进行系统介绍,

使得读者能够快速入门，同时体会大数据处理系统的设计理念与优化方法本质。

《大数据分析》包括大数据分析方法和理论、典型大数据分析任务以及大数据分析系统与应用。本教材特色是理论联系实际，本书从基础理论、典型任务以及系统应用多个方面对大数据分析相关知识进行了系统而详细的介绍，使得读者能够快速入门，体会大数据分析技术的本质特征，领略大数据技术带来的创新理念。

“X”系列教材包含面向各行各业的大数据应用的知识与技术，既可面向工程实践，又可面向职业培训，且将随着产业界大数据应用的发展进行更新迭代。

大数据已成为学术界、产业界和政府共同关注的热点，正在开启信息化的新阶段。大数据人才培养也刚刚起步，还需要付出更多努力去探索。通过汇集中国计算机学会大数据专家委员会的智力资源，丛书编委会希望本系列教材能够为我国大数据人才培养尽到绵薄之力，助力我国大数据事业的蓬勃发展。尽管编委会和作者花费了很大精力规划和编写本系列教材，但是囿于对大数据的认识局限和自身能力限制，难免存在疏漏和错误，欢迎读者批评指正，以待再版时修正完善。

大数据教材系列丛书编委会

2018年7月

近年来,大数据的重要性凸显,许多国家都把大数据上升到国家战略的高度。实施国家大数据战略,离不开大数据技术的研究。回顾信息技术的发展历史,数据管理技术是信息应用技术的基础。在计算机学科分支中,数据管理是整个领域为数不多既有基础理论研究、又有系统软件研制、还有产业支撑的学科。专门从事数据管理系统软件和应用软件研制的甲骨文公司于2013年超越IBM,成为继微软公司之后全球第二大软件公司。如今,历史似乎又正在重演,大数据管理在大数据技术中表现得越来越重要。

顺应国家大数据发展战略和社会需求,根据教育部2018年3月份数据统计显示,国内共有283所高校相继获批了数据科学与大数据技术专业。在此背景下,中国计算机学会大数据专家委员会,围绕大数据专业建设和人才培养的需求,组织编写大数据“1+3+X”系列教材,该教材以梅宏院士主编的《大数据导论》为基础,并针对性地设置《大数据管理》《大数据处理》《大数据分析》三本关键技术课教材和X本领域应用教材。本书《大数据管理》就是其中的一本关键技术课教材。

本书分为三篇10章。第一篇大数据管理概述,包括第1章数据库管理系统概述;第二篇数据模型与语言,包括第2章关系数据模型与SQL、第3章键值对数据模型、第4章文档模型与查询语言、第5章图模型与类SQL查询语言,共4章;第三篇大数据管理系统,包括第6章大数据管理系统的体系架构、第7章数据组织与存储、第8章分布式查询处理优化、第9章分布式事务、第10章故障恢复,共5章。

全书内容全面丰富,既有数据库管理系统发展的历史回顾、现状综述和未来发展预测,又有大数据管理系统核心技术的深入剖析,全书每一章节分别邀请了在该领域研究多年的一线科研工作者或资深架构师执笔,教师可以针对不同专业和不同类别的学生挑选书中不同章节的内容进行讲解。

本书由杜小勇组织编写并任主编。提供本书初稿的主要有:张延松(第2、8章)、李翠平(第3、8章)、张孝(第4、8章)、卢卫(第5、8章)、张峰(第6章)、柴云鹏(第7章)、李海翔(第9章)、陈跃国(第10章)。参加一些章节部分内容初稿编写的有:卞昊穹、丁鹏傑、林玉婷、王童童。全书最后由杜小勇审定。

在本书的撰写过程中，作者阅读参考了国内外大量教材、专著、论文、技术报告、系统源代码，努力跟踪大数据管理的新方法、新技术、新系统，有选择地把它们纳入到教材中来，但因大数据管理技术和系统还在快速发展中，尚未成型，书中不足之处，敬请学术同仁与读者指正。作者邮箱：duyong@ruc.edu.cn。

杜小勇

2018年9月

■ 第 1 章 数据库管理系统概述 .....003	1.2.2 大数据管理系统的系统特征 .....011
1.1 数据管理系统的发展历史 .....003	1.2.3 大数据管理系统的系统应用特征 .....012
1.1.1 第一代：层次、网状数据库系统 .....003	1.3 大数据管理系统的组成 .....013
1.1.2 第二代：关系数据库系统 .....006	1.3.1 多引擎系统结构 .....013
1.1.3 第三代：数据仓库系统 .....008	1.3.2 混合负载系统架构 .....014
1.1.4 第四代：大数据管理系统 .....009	1.3.3 分布式系统架构 .....014
1.1.5 小结 .....010	
1.2 大数据管理系统的特征 .....010	
1.2.1 大数据管理系统的系统数据特征 .....011	

■ 第 2 章 关系数据模型与 SQL .....019	■ 第 3 章 键值对数据模型 .....049
2.1 关系数据库概述 .....019	3.1 概述 .....049
2.1.1 基本概念 .....020	3.1.1 什么是键值对模型 .....049
2.1.2 基本关系操作与实现技术 .....023	3.1.2 键值对模型应用现状 .....050
2.2 关系数据库标准语言 SQL .....024	3.2 数据结构和数据操作 .....051
2.2.1 SQL 基本语法 .....024	3.2.1 Dynamo .....051
2.2.2 SQL 扩展语法 .....026	3.2.2 Redis .....055
2.3 SQL on Hadoop .....033	3.2.3 RAMCloud .....059
2.4 NoSQL 数据库 .....036	3.2.4 BigTable .....060
2.5 代表性的关系数据库 .....038	■ 第 4 章 文档模型与查询语言 .....063
2.5.1 传统数据库技术的发展 .....039	4.1 概述 .....063
2.5.2 代表性的 MPP 数据库 .....040	4.2 文档结构 .....064
2.5.3 代表性的 NewSQL 数据库 .....044	4.2.1 XML 结构 .....064
2.5.4 基于新硬件技术的数据库 .....046	4.2.2 JSON 结构 .....069
2.6 小结 .....046	4.3 查询语言 .....073
	4.3.1 DOM 接口及应用实例 .....073
	4.3.2 XQuery 及应用实例 .....075

4.3.3	FLWOR	078	5.1.2	标签图	098
4.3.4	XPath 及应用实例	080	5.1.3	属性图	100
4.3.5	JSON API 及应用实例	082	5.2	图数据操作	102
4.4	文档数据库举例	084	5.2.1	图匹配	102
4.4.1	eXistdb	084	5.2.2	图导航	105
4.4.2	MongoDB	090	5.2.3	图与关系的复合操作	107
4.5	拓展阅读建议	095	5.3	图查询语言 Cypher	109
4.6	小结	096	5.3.1	对象创建	109
■ 第 5 章 图模型与类 SQL			5.3.2	检索	113
查询语言			5.3.3	图的更新	116
5.1	图的数据结构及其形式化定义	097	5.4	Neo4j 图数据库	118
5.1.1	简单图	097	5.4.1	Neo4j 简介	118
			5.4.2	Neo4j 应用实例	120

---

## 第三篇 大数据管理系统

■ 第 6 章 大数据管理系统的体系架构		127	6.3.4	异构与基于云的分布式数据库	145
6.1	数据库系统体系架构的发展	127	6.3.5	目录系统	147
6.1.1	集中式体系架构	127	6.4	实例分析	148
6.1.2	客户 - 服务器体系架构	129	6.5	小结	150
6.1.3	并行与分布式体系架构简述	130	■ 第 7 章 数据组织与存储		151
6.1.4	数据库系统体系架构的相关概念	132	7.1	概述	151
6.2	并行数据库体系架构	133	7.1.1	数据组织与存储的嵌套关系	152
6.2.1	并行数据库体系架构设计	134	7.1.2	数据组织实例: 文件系统	153
6.2.2	IO 并行	134	7.1.3	数据组织带来的数据映射与放大	155
6.2.3	查询间与查询内并行	136	7.2	硬件访问模型	158
6.2.4	操作间与操作内并行	137	7.2.1	内存访问模型	159
6.3	分布式数据库体系架构	140	7.2.2	磁盘访问模型	160
6.3.1	分布式事务系统结构	140	7.2.3	闪存访问模型	161
6.3.2	分布式数据库中的并发控制	141	7.2.4	瓦记录磁盘访问模型	163
6.3.3	分布式数据库设计中的折中方案	143	7.2.5	非易失内存访问模型	164

- 7.3 索引技术 .....166
  - 7.3.1 哈希索引 .....166
  - 7.3.2 有序索引 .....167
  - 7.3.3 哈希 - 有序复合索引 .....171
  - 7.3.4 存在索引 .....172
  - 7.3.5 其他索引技术 .....174
- 7.4 键值存储 .....174
  - 7.4.1 基于哈希索引的键值存储系统 .....175
  - 7.4.2 基于 LSM 树索引的键值存储系统 .....176
  - 7.4.3 基于 B/B+ 树索引的键值存储系统 .....178
- 7.5 列存储 .....179
  - 7.5.1 列存储数据库 .....181
  - 7.5.2 HDFS 列存储 .....183
- 7.6 其他类型存储 .....185
  - 7.6.1 文档存储 .....185
  - 7.6.2 无结构文档存储 .....185
  - 7.6.3 XML 文档存储 .....186
  - 7.6.4 JSON 文档存储 .....187
  - 7.6.5 图存储 .....187
- 7.7 小结 .....190
- 第 8 章 分布式查询处理优化 .....191
  - 8.1 分布式查询处理概述 .....192
    - 8.1.1 数据分布策略 .....192
    - 8.1.2 分布式查询处理 .....194
    - 8.1.3 分布式查询优化技术 .....197
  - 8.2 面向关系数据的分布式查询处理 .....198
    - 8.2.1 概述 .....198
    - 8.2.2 分布式关系数据库查询处理 .....203
    - 8.2.3 分布式关系数据库查询优化 .....210
    - 8.2.4 分布式关系数据库查询处理技术实例分析 .....212
- 第 9 章 分布式事务 .....223
  - 9.1 概述 .....223
    - 9.1.1 单机事务处理技术 .....223
    - 9.1.2 分布式事务处理技术 .....227
  - 9.2 分布式系统与事务 .....233
    - 9.2.1 分布式系统的挑战 .....233
    - 9.2.2 分布式一致性 .....235
    - 9.2.3 可用性、隔离性、一致性的关系 .....238
  - 9.3 分布式事务 .....242
    - 9.3.1 分布式提交算法 .....242
    - 9.3.2 全局可串行化保证 .....247
    - 9.3.3 去中心化的分布式事务 .....251
  - 9.4 图、键值、文档模型事务处理技术 .....260
    - 9.4.1 图模型事务处理技术 .....261
    - 9.4.2 键值、文档模型事务处理技术 .....262
  - 9.5 典型案例 .....263
    - 9.5.1 Spanner 分布式事务 .....263
    - 9.5.2 CockroachDB 分布式事务 .....268
  - 9.6 拓展阅读 .....273
    - 9.6.1 新硬件与事务 .....273
    - 9.6.2 AI 与事务 .....275
    - 9.6.3 架构与事务 .....276
    - 9.6.4 性能与事务 .....278
    - 9.6.5 并发访问控制算法 .....278
    - 9.6.6 其他 .....279
  - 9.7 小结 .....279
- 第 10 章 故障恢复 .....281
  - 10.1 传统的数据库故障恢复概述 .....281
    - 10.1.1 故障的种类 .....281
    - 10.1.2 故障恢复技术 .....284
  - 10.2 分布式数据库节点故障的终结和恢复协议 .....288
    - 10.2.1 两阶段提交协议的终结和恢复协议 .....289

10.2.2	三阶段提交协议的终结和恢复协议	.....295	10.3.3	Raft 协议与 Paxos 协议的对比分析	.....304
10.2.3	需要注意的事项	.....299	10.4	其他常见的容错与恢复技术	.....305
10.3	当前流行的分布式数据库恢复技术及应用	.....300	10.4.1	Hadoop 的存储副本容错	.....305
10.3.1	Paxos 协议	.....300	10.4.2	HA 热备	.....306
10.3.2	Raft 协议	.....302	10.4.3	键值对系统的故障恢复	.....308
			10.4.4	其他容错技术	.....309

大	数	据	第一篇
管理概述			



# 第1章 数据库管理系统概述

数据库管理系统的功能是伴随着数据库应用的扩展而不断发展起来的。第一代系统的功能主要集中在数据的组织与存储上，为了有效支持层次或者图结构的数据组织，各种链表结构被提出来。这个时期的数据库系统就是一种数据组织与数据存取的工具。

第二代系统主要围绕 OLTP 应用展开，除了关系存储技术之外，重点发展了事务处理子系统、查询优化子系统、数据访问控制子系统。第三代系统主要围绕 OLAP 应用展开，重点在提出高效支持 OLAP 复杂查询的新的数据组织技术，包括 CUBE 和列存储等技术以及 OLAP 分析前端工具的开发。第四代系统主要围绕大数据应用展开，重点围绕分布式可扩展、异地多备份高可用架构、多数据模型支持以及多应用负载类型支持上等。

## 1.1 数据管理系统的发展历史

“管理”一词的含义就是对资源进行有效的规划、组织和使用，使得系统各部分更加有效地配合以高效实现系统的目标。系统资源很少、规模很小的时候，无所谓管理。只有当系统的资源很多、规模很大、系统的复杂性很高的时候，管理才变得举足轻重。由此可见，管理一定是伴随着系统复杂性的提高而逐渐形成的，管理的内涵是随着系统目标的变化而变化的。有一句话说“世界上本来没有路、走的人多了就成了路”，数据管理系统的功能也是在人们发展信息系统的实践中，不断凝练共性而形成的。因此，回顾和总结数据管理系统的历史可以沿着数据库应用发展的主线来展开，这也是数据管理系统的目的所在；同时，回顾历史也要兼顾到系统的复杂性，包括系统所能拥有的资源种类和数据规模大小等因素。

### 1.1.1 第一代：层次、网状数据库系统

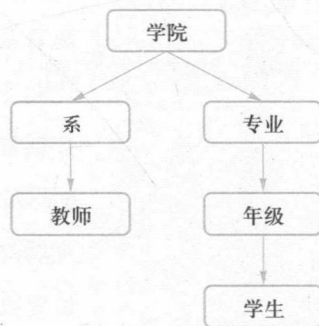
数据管理系统的发展可以上溯到20世纪60年代出现的层次数据库技术。当时计算机已经开始在商业上获得应用，文件作为数据存储的主要设施，已经无法满足对商业应用（如银行业务）中数据项之间的复杂关系进行管理的需求，主要表现在以下几个方面：第一，文件系统是面向单一应用的，也就是说根据每一个应用的需要，有针对性地设计文件

的数据结构。因此不同的应用要使用同一个文件结构会显得效率较低。第二，文件之间的数据是独立的，如果两个文件之间的数据存在内在的逻辑关系，要维护这种关系就非常困难甚至是不可能的。第三，文件的组织方式单一，难以满足不同的访问模式对数据高效率访问的需求。

因此，这个时期的信息系统对数据管理的核心需求是提供一种面向系统整体的数据组织与访问功能，简单地说就是存储和访问这两件事情。受制于当时的计算机技术限制，第一代的数据管理系统是层次型的，之后又进一步扩展为网状型数据库。这里所谓的层次型或者网状型指的是系统中的数据组织方式是按照树或者（受限）图来组织的。树由于其每一个结点最多只有一个父结点，因此可以采用更加有效的手段（例如按照树遍历的顺利）来存储数据。网状模型则通过引入“基本层次联系”的概念，将图分解为一组基本层次联系的集合。而基本层次联系实际上就是一个命名的层次联系。因此，这两类数据库本质上还是一样的，都可以用“树”结构来表达和存储数据。尽管这个时期数据管理的功能集中在数据存储组织和数据访问等，但是，这是第一次将数据管理的功能从具体的应用逻辑中分离并独立出来，在数据管理系统的发展历史上是一件里程碑的事情。

图 1-1 层次模型

举一个学校组织的例子：一所大学有不同的学院，学院有不同的系，每一个系有多位教师，学院也有不同的专业，每个专业下，每一学年都有许多学生入学。这样一个组织可以用一个层次模型表示如图 1-1 所示。



对于层次/网状数据库，数据访问的最常见的模式就是根据某一个父结点的值检索子结点的全部或者部分值。例如，查询信息学院计算机系的教师张丽的情况，数据的访问就需要从学院到系再到教师这样的路径进行访问，为了提高数据访问的效率，最有效的数据存储方式就是按照树遍历（例如中序遍历）方式访问树结点并将这些结点的数据邻近存储，兄弟结点之间则用指针进行链接。因此，这个时期的数据库看起来就是玩各种数据结构，指针、链表被大量使用。