

数据库应用理论系列图书

# 数据库理论 研究方法解析

郝忠孝 著



科学出版社

数据库应用理论系列图书

# 数据库理论研究方法解析

郝忠孝 著

科学出版社

北 京

## 内 容 简 介

本书是基于作者三十多年对数据库理论研究过程中使用方法的经验总结的基础上撰写而成的。本书以发现、确定到解决问题为主线,系统阐述了数据库理论研究方法并以实例进行解析。

本书共 13 章。主要内容包括:哲学的方法论原理、课题选择方法、系统思维划分产生命题、演绎中增强(削弱)条件限制确定命题、具体到抽象方法确定命题、类比推理方法确定命题、文献阅读产生命题、学术论文评价产生命题、专著产生项目和课题、各种命题正确性证明、算法性质和证明及实验、各种算法复杂性分析技术和设计、学术论文写法和严守道德规范等。

本书可供从事计算机数据库、网络安全理论和计算机其他分支领域理论学习研究的本科生、硕士生、博士生、科研人员使用,也可供从事其他学科理论研究人员参考。

图书在版编目(CIP)数据

数据库理论研究方法解析/郝思学等. —北京:科学出版社,2015.5

(数据库应用理论系列图书)

ISBN 978-7-03-044260-2

I. ①数… II. ①郝… III. ①数据库系统 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2015)第 098585 号

责任编辑:余 丁 闫 悦 / 责任校对:郭瑞芝

责任印制:张 倩 / 封面设计:耕者设计工作室

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>

文林印务有限公司印刷

科学出版社发行 各地新华书店经销

\*

2015 年 5 月第 一 版 开本:720×1 000 1/16

2015 年 5 月第一次印刷 印张:20 3/4

字数:418 000

定价:98.00 元

(如有印装质量问题,我社负责调换)

## 作者简介



郝忠孝,教授,山东蓬莱人,1940 年 12 月生,曾任原东北重型机械学院副院长,齐齐哈尔大学副校长,哈尔滨理工大学学术委员会主席。现任哈尔滨工业大学博士生导师(兼)、哈尔滨理工大学博士生导师。原机械电子工业部有突出贡献专家,享受国务院政府特殊津贴,全国优秀教师,黑龙江省共享人才专家,黑龙江省

省级学科带头人,黑龙江省计算机学会副理事长。

主要研究领域:①空值数据库理论。在国内外首次提出了空值数据库数据模型,完成一系列相关研究,形成了比较完整的理论体系,著有国内外第一部该方面的论著《空值环境下数据库导论》。②数据库 NP-完全问题的求解问题。首次基本解决了求全部候选关键字、主属性,基数为  $M$  的候选关键字,最小候选关键字等问题,著有《关系数据库数据理论新进展》。③数据库数据组织的无环性理论研究。在无  $\alpha$  环、无  $\beta$  环、无  $\gamma$  环的分解条件与规范化理论研究方面取得突破性进展,著有《数据库数据组织的无环性理论》。④时态数据库理论研究。系统提出并完成了时态数据库中基于全序、偏序、多粒度环境下的各种时态理论问题研究,著有《时态数据库设计理论》。⑤主动数据库理论研究。著有国内外第一部该方面的论著《主动数据库系统理论基础》。⑥不完全信息下 XML、概率 XML 数据库理论研究。首次解决了不完全信息下 XML 数据库部分理论研究问题,著有《不完全信息下 XML 数据库基础》。⑦空间、时空数据库理论研究。首次解决了空间数据库线段最近邻查询和其他多个问题,著有《时空数据库查询与推理》、《时空数据库新理论》、《移动对象数据库理论基础》、《空间数据库理论基础》等著作。

作为负责人完成了国家、省部级项目 10 项,获省、部级科技进步奖一、二、三等奖 7 项。发表学术论文 230 余篇,其中,国家一级论文 160 余篇;在《计算机研究与发展》上发表个人学术论文专辑两部,被 SCI、EI 等检索 140 余篇。1991 年发表学术论文数居中国科技界第五位(并列)。著书 11 部。

## 前 言

计算机数据库科学理论研究方法是科学研究中的一个组成部分。科学研究方法是发现、确定研究问题和解决问题的有力工具,是指导正确进行研究工作的保证。即它是为计算机数据库理论研究提供方法、原则、手段和途径的。计算机数据库科学理论作为一种高级复杂的知识形态和认识形式,是在人类已有知识的基础上,通过正确的思维方式、研究方法、研究手段和一定的实践活动而获得的。因此,在计算机数据库科学理论研究和发展的过程中,是否拥有正确的科学方法,是对计算机数据库理论研究做出贡献的关键。正确的科学方法可以使研究者根据计算机数据库理论科学研究发展的客观规律,确定正确的研究方向;可以为研究者提供研究的具体方法;可以为计算机数据库理论的新发现、新发展提供支撑。

哲学的方法论原理是指导人类科学研究的重要指南,因此计算机数据库理论研究方法也离不开哲学方法论原理的指导,计算机数据库理论研究方法的实质是哲学方法论原理在计算科学中的具体应用。因此,必须强调在计算机科学理论研究中特别需要注重科学方法的研究和利用。数据库理论研究方法是解决数据库理论问题的一把钥匙。

可以说,计算机运算的基础是二进制数,计算机的出现是数学计算的产物。计算机技术的发展与数学的发展密不可分。数学理论研究是演绎性质的科学,因此,计算机数据库理论中命题结论的正确性,一般情况下必须遵循数学逻辑证明,并且相当大的一部分也是通过演绎得到的。此外,数据库还和数学计算有很大不同,大多数理论研究属于应用理论研究范畴,因此和有些数学研究方法也不同。

对于任何自然科学,正确的思维方式和推理形式是产生正确研究方法的源泉。本书是作者 38 年来对数据库理论研究过程中使用方法的经验总结,本书的出版可能对致力于数据库理论研究的人员提供一些可借鉴的研究方法,这也是作者出版本书的目的。

本书以发现、确定到解决问题为主线,包括①哲学的方法论原理:需求和想象、思维和抽象、演绎推理、合情推理。②科学分析问题、选择课题的原则、选题的禁忌、课题的选择方法和提出科学假设的方法。③课题划分:对象目标逆向推理划分策略、对象目标逆向静态分析和动态分析。④课题和命题选择与确定方法:演绎中增强条件限制确定命题、演绎中削弱条件限制确定命题、具体到抽象的方法确定命题、类比推理的方法确定命题、文献阅读产生命题、学术论文评价产生命题、专著产生项目和课题。⑤命题正确性证明:命题证明基础、证明方法和实验、算法性质和

证明。⑥算法复杂性分析基础、分析技术、设计方法和程序设计等。本书对以上相关方法和技术进行了详细地讨论并以实例进行解析,力求做到条理清晰、逻辑性强,易于理解、掌握和使用。最后讨论了学术论文写法和严守道德规范的重要性。

本书的出版得到了科学出版社耿建业、闫悦等同志的大力支持和帮助,在此表示衷心感谢。

李松博士和张丽平审阅了全书,提出了许多宝贵意见。本书所有图表的绘制分别由张丽平、孙冬璞、万静和郝晓红等完成。在此书出版之际,向他们表示诚挚的谢意!

作 者

2015年2月于哈尔滨

# 目 录

## 前言

|                          |    |
|--------------------------|----|
| 第 1 章 客观世界需求和想象          | 1  |
| 1.1 客观世界需求               | 1  |
| 1.1.1 客观世界需求和数据库产生与发展    | 1  |
| 1.1.2 客观世界需求和网络安全技术产生与发展 | 7  |
| 1.2 想象                   | 8  |
| 第 2 章 思维和抽象              | 11 |
| 2.1 形式逻辑                 | 11 |
| 2.2 逻辑思维                 | 12 |
| 2.3 非逻辑思维                | 15 |
| 2.4 辩证逻辑思维               | 16 |
| 2.5 灵感思维                 | 17 |
| 2.5.1 灵感思维特征和产生的基础       | 17 |
| 2.5.2 灵感和机遇              | 19 |
| 2.6 创新性思维                | 20 |
| 2.6.1 创新性思维表现方式          | 20 |
| 2.6.2 创新性思维的特点           | 29 |
| 2.6.3 如何应用创新思维           | 30 |
| 2.7 抽象                   | 32 |
| 2.7.1 抽象概述               | 32 |
| 2.7.2 科学抽象               | 32 |
| 2.7.3 科学抽象与抽象的区别         | 33 |
| 2.7.4 科学抽象意义             | 34 |
| 2.7.5 科学抽象思维的基本过程        | 35 |
| 第 3 章 演绎推理               | 37 |
| 3.1 推理类型和作用              | 37 |
| 3.1.1 推理概述               | 37 |
| 3.1.2 推理的逻辑性和结论正确的条件     | 38 |
| 3.2 演绎推理的一般模式——三段论       | 41 |
| 3.3 直言命题和直接推理            | 46 |

|            |                     |           |
|------------|---------------------|-----------|
| 3.3.1      | 直言命题                | 46        |
| 3.3.2      | 性质命题的种类             | 47        |
| 3.4        | 假言(条件)命题及推理         | 48        |
| 3.4.1      | 条件命题                | 48        |
| 3.4.2      | 条件命题推理              | 52        |
| 3.5        | 选言推理和联言推理           | 54        |
| 3.5.1      | 选言(析取式)推理           | 54        |
| 3.5.2      | 联言(合取式)推理           | 57        |
| <b>第4章</b> | <b>合情推理和因果关系推理</b>  | <b>60</b> |
| 4.1        | 归纳推理概述              | 60        |
| 4.2        | 不完全归纳推理和完全归纳推理      | 62        |
| 4.2.1      | 不完全归纳推理             | 62        |
| 4.2.2      | 完全归纳推理              | 63        |
| 4.3        | 归纳推理和演绎推理           | 64        |
| 4.3.1      | 总结出规律和结论的两个阶段       | 64        |
| 4.3.2      | 归纳推理和演绎推理的关系        | 65        |
| 4.4        | 类比推理及特点             | 67        |
| 4.5        | 类比推理的种类及结构          | 69        |
| 4.5.1      | 类比推理的种类             | 69        |
| 4.5.2      | 类比推理的结构             | 70        |
| 4.6        | 归纳和类比推理的联系与区别       | 76        |
| 4.7        | 因果关系推理              | 77        |
| 4.7.1      | 因果关系及性质             | 78        |
| 4.7.2      | 逻辑推理与因果关系的区别        | 80        |
| <b>第5章</b> | <b>科研课题的选择和确定方法</b> | <b>82</b> |
| 5.1        | 问题和课题               | 82        |
| 5.2        | 选择课题的原则             | 85        |
| 5.2.1      | 选择课题是科学研究的第一步       | 85        |
| 5.2.2      | 课题选择的原则             | 85        |
| 5.2.3      | 选题的可行性原则            | 88        |
| 5.3        | 选题的禁忌               | 89        |
| 5.4        | 科学分析问题              | 91        |
| 5.5        | 提出科学假设的意义及方法        | 94        |
| 5.5.1      | 提出科学假设的意义           | 94        |
| 5.5.2      | 科学假设产生的客观基础         | 96        |



|              |                     |            |
|--------------|---------------------|------------|
| 5.5.3        | 要发现和认识冲突与矛盾         | 96         |
| 5.5.4        | 以批判的精神冲破传统观念的束缚     | 97         |
| 5.5.5        | 产生科学假设的环境           | 98         |
| 5.5.6        | 产生科学假设的方法           | 100        |
| 5.5.7        | 建立形式不同但本质相同的理论体系    | 101        |
| <b>第 6 章</b> | <b>系统思维产生子系统和命题</b> | <b>103</b> |
| 6.1          | 系统思维的特点和原则          | 104        |
| 6.1.1        | 产生子系统的思维特点和原则       | 104        |
| 6.1.2        | 系统的整体性特点和原则         | 106        |
| 6.1.3        | 系统的综合性特点和原则         | 107        |
| 6.1.4        | 系统的结构性特点和优化原则       | 107        |
| 6.1.5        | 系统的动态性特点和原则         | 108        |
| 6.1.6        | 系统的多维思维特点和原则        | 109        |
| 6.1.7        | 系统的模型化特点和原则         | 110        |
| 6.2          | 系统思维划分产生课题          | 110        |
| 6.2.1        | 目标逆向推理划分策略          | 111        |
| 6.2.2        | 目标逆向静态分析和动态分析       | 112        |
| 6.3          | 系统思维产生命题            | 115        |
| 6.3.1        | 演绎中增强条件限制确定命题       | 115        |
| 6.3.2        | 演绎中削弱条件限制确定命题       | 116        |
| 6.3.3        | 具体到抽象的方法确定命题        | 119        |
| 6.3.4        | 类比推理的方法确定命题         | 119        |
| <b>第 7 章</b> | <b>文献研究确定课题和命题</b>  | <b>121</b> |
| 7.1          | 文献研究的基本知识           | 121        |
| 7.2          | 阅读文献打好研究基础          | 123        |
| 7.3          | 批判地阅读和吸收            | 126        |
| 7.4          | 阅读坚持的原则和课题及命题选择     | 130        |
| 7.4.1        | 阅读文献和论文时的原则及方法      | 130        |
| 7.4.2        | 学术论文评价与课题及命题选择      | 133        |
| 7.4.3        | 专著与课题选择             | 134        |
| 7.5          | 阅文整理和科研方案的制订        | 135        |
| <b>第 8 章</b> | <b>命题证明基础</b>       | <b>137</b> |
| 8.1          | 概念                  | 137        |
| 8.2          | 命题                  | 139        |
| 8.2.1        | 命题及真假性              | 139        |

---

|       |                                 |     |
|-------|---------------------------------|-----|
| 8.2.2 | 命题的种类和逻辑性 .....                 | 140 |
| 8.2.3 | 集合论中的“交”“并”“补”与逻辑联结词的对应关系 ..... | 142 |
| 8.2.4 | 全称命题和特称命题 .....                 | 143 |
| 8.2.5 | 含有一个量词的命题的否定 .....              | 144 |
| 8.3   | 几种不同的命题 .....                   | 144 |
| 8.3.1 | 定义 .....                        | 144 |
| 8.3.2 | 公理 .....                        | 148 |
| 8.3.3 | 公理系统的有效性和完备性证明 .....            | 150 |
| 8.3.4 | 定律和原理 .....                     | 153 |
| 8.3.5 | 定理 .....                        | 154 |
| 第9章   | 证明方法和实验 .....                   | 157 |
| 9.1   | 证明 .....                        | 157 |
| 9.1.1 | 证明概述 .....                      | 157 |
| 9.1.2 | 证明和推理的联系与区别 .....               | 159 |
| 9.1.3 | 证明的规则和步骤 .....                  | 161 |
| 9.2   | 综合法和分析法 .....                   | 165 |
| 9.2.1 | 综合法 .....                       | 165 |
| 9.2.2 | 分析法 .....                       | 166 |
| 9.2.3 | 综合法和分析法的特点 .....                | 169 |
| 9.3   | 条件关系证明法 .....                   | 171 |
| 9.4   | 反证法和同一法 .....                   | 176 |
| 9.4.1 | 间接证明方法种类 .....                  | 176 |
| 9.4.2 | 反证法 .....                       | 177 |
| 9.4.3 | 同一法 .....                       | 179 |
| 9.5   | 构造性证明法和存在性证明法 .....             | 180 |
| 9.5.1 | 构造性证明法 .....                    | 181 |
| 9.5.2 | 存在性证明法 .....                    | 182 |
| 9.6   | 数学归纳法 .....                     | 184 |
| 9.6.1 | 完全数学归纳法 .....                   | 184 |
| 9.6.2 | 不完全数学归纳法 .....                  | 187 |
| 9.7   | 因果证明 .....                      | 188 |
| 9.8   | 计算机理论研究中的科学实验 .....             | 189 |
| 9.8.1 | 计算机科学理论实验的作用 .....              | 189 |
| 9.8.2 | 计算机理论研究中实验的种类 .....             | 191 |
| 9.8.3 | 科学模拟实验 .....                    | 195 |

|                             |     |
|-----------------------------|-----|
| 9.8.4 整理经验材料的方法 .....       | 196 |
| <b>第 10 章 算法复杂性</b> .....   | 198 |
| 10.1 类 PASCAL 语言 .....      | 198 |
| 10.2 算法的性质和证明 .....         | 207 |
| 10.2.1 基本概念 .....           | 207 |
| 10.2.2 算法应具有的性质 .....       | 208 |
| 10.2.3 算法正确性证明 .....        | 208 |
| 10.2.4 算法工作量 .....          | 212 |
| 10.3 算法复杂性 .....            | 213 |
| 10.3.1 空间复杂性 .....          | 213 |
| 10.3.2 时间复杂性 .....          | 214 |
| 10.4 算法复杂性分析基础 .....        | 216 |
| 10.4.1 多项式时间算法与指数时间算法 ..... | 216 |
| 10.4.2 算法分析的三种情况和表示方法 ..... | 219 |
| 10.4.3 总结和说明 .....          | 224 |
| 10.4.4 难解性或为难解问题 .....      | 225 |
| <b>第 11 章 算法设计方法</b> .....  | 227 |
| 11.1 问题的模型 .....            | 228 |
| 11.1.1 数学模型 .....           | 228 |
| 11.1.2 数据模型 .....           | 229 |
| 11.2 算法设计概述 .....           | 232 |
| 11.2.1 算法设计的步骤 .....        | 232 |
| 11.2.2 算法的有效性 .....         | 233 |
| 11.3 递归方法和算法递归设计 .....      | 234 |
| 11.3.1 递归技术和递归算法概述 .....    | 234 |
| 11.3.2 递归函数和递归过程 .....      | 242 |
| 11.3.3 递归过程实现 .....         | 243 |
| 11.3.4 尾递归和线性递归 .....       | 245 |
| 11.3.5 递归设计 .....           | 247 |
| 11.3.6 消除递归 .....           | 248 |
| 11.4 穷举法和贪心法 .....          | 253 |
| 11.4.1 穷举法 .....            | 253 |
| 11.4.2 贪心法 .....            | 255 |
| 11.5 治类方法 .....             | 258 |
| 11.5.1 分治法 .....            | 259 |

|                           |            |
|---------------------------|------------|
| 11.5.2 减治法                | 264        |
| 11.5.3 变治法                | 265        |
| 11.6 时空权衡法和动态规划           | 268        |
| 11.6.1 时空权衡法              | 268        |
| 11.6.2 动态规划               | 271        |
| 11.7 回溯法和分支限界法            | 277        |
| 11.7.1 回溯法                | 278        |
| 11.7.2 分支限界法              | 280        |
| <b>第12章 算法复杂性分析技术</b>     | <b>282</b> |
| 12.1 几种常用的比较两个函数阶的方法      | 282        |
| 12.1.1 几种常用的参照法           | 282        |
| 12.1.2 比较两个函数阶的方法         | 284        |
| 12.1.3 常用的和式估计上界法         | 285        |
| 12.2 递归算法的复杂度分析技术         | 286        |
| 12.2.1 递归算法的复杂度分析方法       | 287        |
| 12.2.2 递归过程分析             | 290        |
| 12.2.3 递归方程求解             | 292        |
| 12.3 生成函数与求和              | 294        |
| 12.4 算法实现和程序设计            | 298        |
| <b>第13章 学术论文写法和严守道德规范</b> | <b>300</b> |
| 13.1 学术论文及写法规范            | 300        |
| 13.1.1 基础知识               | 300        |
| 13.1.2 学术性研究和学术论文选题       | 300        |
| 13.1.3 学术论文的写法规范          | 302        |
| 13.2 严守道德规范               | 309        |
| 13.2.1 学术不端               | 309        |
| 13.2.2 违背学术道德的讨论          | 310        |
| 13.3 学术论文和著作投稿            | 314        |
| 13.3.1 学术论文投稿             | 314        |
| 13.3.2 著作和投稿              | 315        |
| 13.4 坚守国家利益高于一切           | 317        |
| <b>参考文献</b>               | <b>319</b> |

# 第 1 章 客观世界需求和想象

## 1.1 客观世界需求

### 1.1.1 客观世界需求和数据库产生与发展

数据库技术是在 20 世纪 60 年代末作为数据管理的最新技术登上数据处理舞台的。半个世纪以来,随着计算机应用的不断扩大,计算机硬件的快速发展,计算机数据库技术、网络技术也得到了迅速的发展。它们已经成为当今世界计算机应用中两个最重要的基础领域。

本书所讨论的客观世界需求包含以下方面。

(1) 人类的面向实际工程的直接需求,即面向实际的一些应用领域。对于计算机学科中的数据库来说,包括航天航空、火箭发射、军事系统、导弹系统、军事情报、卫星监测及图像处理、移动通信、多媒体数据库、地理信息系统和气象云图分析等多种领域。

(2) 人类的面向非实际工程应用的间接需求。①广泛应用于计算机学科中人工智能、知识发现、数据挖掘、模式识别、CAD/CAM 系统等。②计算机数据库、网络安全、数学和计算机各门学科理论研究的需求。这些学科理论的不断研究,理论体系的不断推进都是客观世界的需求,这些理论的研究大多是因为实际的需求而产生和发展的。

20 世纪 80 年代以来,由于非传统应用领域的不断扩大,针对一些特殊领域的应用提出了许多新的数据管理需求功能,关系数据库已经不具备这种能力。在计算机数据库的理论研究中,许多研究成果,如面向对象数据库、空值数据库、时态数据库、主动数据库、空间数据库、时空数据库、移动数据库、分布式数据库和网络数据库便产生和发展起来;单一的关系数据库已经无法满足科学发展的客观需要。

以上数据库的产生和发展有两大原因:①由于人类科学的不断发展和实际应用的需求而产生;②由于人类想象力和科学思维能力的提升,使计算机中各类数据库、网络安全技术的科学理论不断得到发展。

本节只讨论和分析由于人类科学的不断发展和实际应用的需求而产生的各种数据库、网络安全技术,即第一个原因。

(1) 由于科学的不断发展,数电信号出现及使用为计算机的产生奠定了基础。

(2) 在科学的不断发展中,仅以计数的运算是无法适应人类科学研究不断发

展的需求。

(3) 生产实践或实际应用是现代科学发展的主因,如工程设计(一种生产实践)的一个重要手段是应用计算机上专门设计的数据库,现代人类活动的方方面面都是如此。

(4) 在科学理论不断发展的过程中,为使新的理论向前发展,只有在解决了阻碍新理论发展瓶颈的前提下才能得以发展,才能满足实际应用的需求。

下面对各种数据库进行解析。

#### 1) 客观世界需求和关系数据库

关系数据库(郝忠孝,1998;厄尔曼,1988)是以关系模型为基础的数据库,它利用关系描述客观世界。关系实质上是一个二维表,表的一行称为一个元组,对应于客观世界的一个事物;表的一列称为一个属性;一个元组在某一属性上的值称为该元组的一个分量,它对应于客观世界中一个事物的某一特征。通常,关系数据库是建立在假定数据库中每个数据都是实实在在的数据的基础上,它在未出现其他各种数据库之前,为处理许多实际问题做出了重要贡献。不仅如此,由于这种数据库有严格的理论体系,还为以后各种数据库的发展奠定了基础,在各种数据库的发展中功不可没。在关系数据库范式理论研究中,第一范式是关系数据库对模式的基本要求(要求属性值必须是原子属性即不可以再分);第二范式是为了消除关系数据库模式中非主属性对候选关键字的部分依赖;第三范式是为了消除关系数据库模式中非主属性对候选关键字的传递依赖;BC范式是为了消除关系数据库模式中一切属性(主属性和非主属性)对候选关键字的传递依赖;第四范式是为了消除关系数据库模式中非平凡的和非候选关键字所隐含的多值依赖;第五范式是为了消除关系数据库模式中非平凡的和非候选关键字所隐含的连接依赖。范式的级别越低,冗余与更新异常就越容易产生。但是,模式级别越高或最高并不表明是最好,因为模式级别越高,被划分得越精细,颗粒度越小,在查询过程中需要模式连接时,就会付出很高的时间与操作复杂性的代价。更为严重的是数据之间的依赖关系会因模式的细分而受到损坏。在进行数据库模式设计时,究竟将数据库模式划分到哪一级,需要根据应用的具体情况权衡之后做出适当的选择。关系数据库范式理论研究既是理论研究的需求,也是客观设计的实际需求。

#### 2) 客观世界需求和面向对象数据库

在实际生活中,需要把数据库技术和面向对象软件开发技术及方法相结合,利于把常规管理信息系统(MIS)等数据库应用系统推广到较为复杂的应用领域,此时由于通常的关系数据库适用于管理和操作日常的数据(二维表),对于那些更为复杂的数据和信息,如工程设计图纸、卫星遥感得到的信息和数据影像、医疗诊断影像等,通常的关系数据库是没有能力管理和处理的。在这个背景下,不仅要有效地存储这些信息和数据,而且还需要开发新的信息查询技术,由此便产生了面向对

象数据库(汪成为等,1992)。

### 3) 客观世界需求和空值数据库

通常,关系数据库是建立在假定数据库中不存在任何未知信息的基础之上的。这种数据库描述只反映了客观世界的已知信息部分,它与客观世界的客观存在性有一定差距。客观世界告诉我们,有些信息暂时未知,有些信息不存在,有些信息甚至连是否存在都不知道。但是,这些信息是普遍存在的,有时是大量的。如在客观世界中,常常出现某些学生因故缺考而暂时无成绩;历史档案中有时出现“生日不详”;会议发言人有时宣布议事日程“有待公布”;警方记录常常出现“目前下落不明”等,这些在通常的关系数据库中是无法描述的。

用通常的关系数据库的相关理论去研究含有空值的数据库是无法进行的。当一个新的元组被插入关系时,若它的个别属性值尚未确定,则无法插入。想要使这种元组在连接过程中不丢失,这在没有引入空值之前是无法达到的。当一个新的属性被填入关系模式时,原模式上各个关系中所有元组在此属性列上的值,显然是暂时填入“空值”最为恰当,这在通常的关系数据库中是不允许的。在这种情况下便产生了空值数据库(郝忠孝,1996)。由于“空值”的出现,虽然通过类比推理方式可能知道要解决什么问题,但是,概念、理论推理过程,大多数情况下都是不同的。其主要原因是,处理“空值”是空值数据库的根本特性。

### 4) 客观世界需求和时态数据库

由于现实世界是不断演变进化的,时间是反映现实世界信息的基本组成部分,因而大多数数据库应用程序都有时态的特性。例如,地震资料分析应用程序、天气监测应用程序、天气预报分析程序、资源管理应用程序、银行等财经类应用程序、项目管理等记录性应用程序等都与信息、数据的时态性密切相关。关系数据库管理系统对时态信息的存储、处理和操作都十分有限,缺乏对时态数据的支持,因而在很多方面产生了问题。例如,它把时间数据作为一个字段值进行存储和管理,只反映了对象某一个时刻或当前时刻的信息和状态,没有联系对象的历史、现在和将来,无法将对象的历史、现在和将来作为一个发展过程来看待,无助于解释事物发展的本质规律。抓住事物发展趋势这一点对于决策支持系统这类应用程序来说是最基本、最重要的,管理数据库系统中元事件的时态信息也是重要的,例如数据库被查询修改的时刻、时间区间。多用户系统中对锁定排队以及资源竞争协调的时标等这些时态数据也有助于提高数据库系统的可靠性和效率。人们开始逐渐意识到必须为时态数据建立时态数据库的模型,或者在现有的关系数据库模型上加以改造,于是提出了时态数据库(郝忠孝,2009b)的概念,时态数据库系统便产生了。所谓时态数据库,是指能够处理时间信息的数据库。关系数据库只记录了数据的当前状态,在现实情况改变时,数据库也发生变化。而时态数据库不仅存放对象的现状,而且存放对象过去的一切状态,并且可以根据对象现在和过去的状态推测其

未来可能的状态。

#### 5) 客观世界需求和主动数据库

关系数据库系统只能按照用户明确给出的请求执行相应的操作,完成某个事务,因此关系数据库系统是被动的。数据库状态的改变是外界或用户程序影响的结果,也就是所有的查询和数据处理必须通过人工操作完成。为了实现数据完整性和一致性的自动维护以及满足实时信息处理的需要,要求数据库系统通过主动规则或触发器的形式扩充关系数据库,做出实时响应。在这种情况下便产生了主动数据库系统(郝忠孝,2008)。

#### 6) 客观世界需求和空间数据库

由于遥感等空间信息获取技术的不断进步,现代社会对位置服务和分析决策的需求也日益迫切,因此深入研究和掌握空间信息技术的理论和方法的重要性也日益突显出来。空间信息,也就是在某个空间参考系(例如地球表面)中对象的位置信息,长期以来被视为特殊的计算问题。

尽管在 20 世纪 70 年代开始出现的每一个经典数据库管理模型都对空间应用领域有所考虑,但不管是关系模型还是面向对象的建模都不能完全适用于这一领域。关系模型能够较好地处理拓扑关系,但对表示横跨空间区域的复杂层次关系却无能为力;而面向对象模型能够处理拓扑和层次关系,但难以处理空间中重要的连续性现象。为此,空间数据库系统(郝忠孝,2013)便产生了。

#### 7) 客观世界需求和时空数据库

针对一些特殊领域的应用,空间数据库和时态数据库已经成为现代数据库的两个重要分支。但是,随着越来越高的数据库应用要求,单独的时态数据库和空间数据库已经无法满足需求。

空间数据库的研究侧重于对数据库中对象的几何模型和空间查询方面的支持,仅能存储空间信息的当前状态;时态数据库的研究则主要关注数据的当前信息和历史信息的处理和扩展,即有效时间和事务时间的表达,不能处理空间信息,这些不足限制了对时空对象的有效管理和处理。空间数据库一般不保存历史变化或只保留若干典型时间点的全局状态快照序列,具有较弱的时空语义建模能力,无法提供时态分析功能。

时态数据类型根据时间的功能和结构有两种分类方式:有效时间类型和事物时间类型,有效时间类型可以是时间点和时间区间;事物时间类型只能是时间区间。时间间隔是指用一系列时间点组成的时间片。

如果要求一种数据库管理两类空间对象,一类是静态的空间对象,如山脉、道路、河流以及城市等;另一类则是移动对象,就是指随时间的变化位置也在不断变化的物体,移动对象的特点是在任意时刻都同时具有时间和空间特性,尽管关系数据库技术为移动对象的管理提供了基础,但要在数据库中表示移动对象的信息,还



需要考虑移动对象所独有的特性,即移动性。

在这种情况下,能够同时处理时态数据和空间数据的时空数据库系统(郝忠孝,2010,2011a)的产生成为必然,时空数据库系统是能够同时处理时态数据和空间数据的数据库系统。

#### 8) 客观世界需求和移动数据库

在客观世界中,如铁路中运行的火车、公路中运行的车辆、空中飞行的飞机等,它们都是移动的、动态的。移动对象与空间对象数据的不同之处就在于它的任何一个数据项的时间戳的帧都是单调递增的,其数据处于不停地变化与更新中,这也是移动对象与空间对象数据的不同之处。针对这种情况,便出现了专门处理移动对象信息和数据的移动数据库系统(郝忠孝,2012)。

#### 9) 客观世界需求和数据组织的无环性理论研究

在科学理论的不断发展过程中,为使新的理论向前发展,只有在解决了阻碍新理论发展的“瓶颈的理论”的前提下才能得以发展,才能满足实际应用的需求。

众所周知,凡是编制程序的人在编制过程中都小心翼翼的避免程序出现死循环,一个具有死循环的程序是无法得到结果的,出现死循环的重要原因之一,就是由于数据组织的结构中存在环。在这种情况下,怎样避免出现环就成为避免程序出现死循环的一种关键技术,为达到此目的所要研究的关系数据库理论中最重要的组成部分是数据依赖及关系规范化理论。数据依赖作为对数据固有语义的一种描述形式被引入,一些重要的数据依赖类型,如 FD、MVD、JD 等,描述数据依赖之间的逻辑蕴涵关系的形式公理系统被提出。在对数据库模式性能与数据依赖之间的联系进行研究的基础上,形成了一系列不同的数据库范式的概念以及以实现这些范式为目标的关系数据库设计和实现查询算法。使数据库研究人员更加清楚地意识到对于数据库模式分解不仅要达到保持函数依赖性、无损连接性以及  $4NF$  为主要特性的分解问题,还必须深入的研究和探讨如何同时达到无  $\alpha$  环的分解问题。而把无  $\alpha$  环性作为继保持依赖性、无损连接性、较好的规范化等性质之后的又一优良特性。对于数据库模式中的无环性(郝忠孝,2009a),存在不同的级别,无  $\alpha$  环数据库消除了许多有  $\alpha$  环数据库出现的不良现象;而无  $\beta$  环除具有无  $\alpha$  环所具有的优良特性外,还具有子图的无环性,因此无  $\beta$  环数据库有效消除了子模式的数据异常;无  $\gamma$  环数据库模式是无  $\alpha$  环数据库模式的一个真子集,除具有无  $\alpha$  环数据库模式所有特性外,无  $\gamma$  环数据库模式还具有特殊的性质。无论是无  $\alpha$  环、无  $\beta$  环或无  $\gamma$  环,统称为无环。反之,统称其为有环。

#### 10) 客观世界需求和不完全信息下 XML 数据库

在客观世界中,常常存在某些不确定的信息在经典数据库中是无法描述和处理的,这些信息称为不完全信息。为使对不完全信息的描述更接近现实世界,可根据不完全信息本身所固有的重要语义信息类型:不存在型不完全信息、存在型不完全