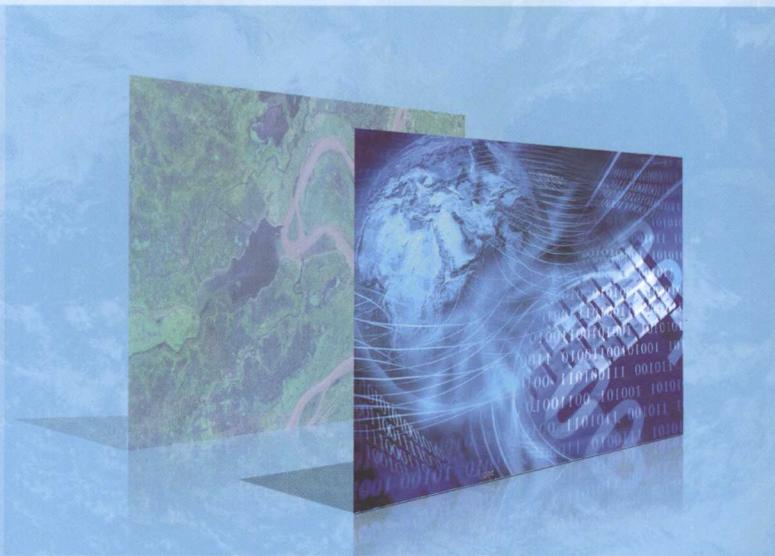


空间数据库教程

毕硕本 编著



科学出版社

013052468

P208
225

内容简介

空间数据库教程

毕硕本 编著



科学出版社

北京



北航

C1656253

P208
225

013025488

内 容 简 介

本书以高等院校培养综合型交叉学科本科人才为发展目标而编写,介绍了空间数据库的原理与方法。全书共 13 章,主要内容包括:数据库系统概论、空间数据库、数据模型与空间信息模型、关系数据库模型、结构化查询语言、空间查询语言、关系模式的规范化理论、数据库设计与实施、空间数据存储和空间数据索引、数据查询优化、空间查询处理与优化、数据库技术新发展、空间数据库发展趋势。

本书内容丰富、结构合理、针对性强,理论叙述严谨、能力培养目标明确。读者学完本门课程后,应掌握空间数据库系统的基本知识、基本原理与方法。

本书可作为各高等院校地理信息类专业学生的教科书,也可作为遥感科学与技术、测绘工程、计算机科学与技术等相关专业学生的教科书,同时也适合于从事空间数据库技术、地理信息系统软件开发的人员参考。

图书在版编目(CIP)数据

空间数据库教程/毕硕本编著. —北京:科学出版社,2013

ISBN 978-7-03-037192-8

I. ①空… II. ①毕… III. ①地理信息系统-高等学校-教材 IV. ①P208

中国版本图书馆 CIP 数据核字(2013)第 054184 号

责任编辑:杨 红 / 责任校对:张小霞
责任印制:阎 磊 / 封面设计:迷底书装

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>

铭浩印刷厂印刷

科学出版社出版 各地新华书店经销

*

2013 年 6 月第 一 版 开本:720×1000 B5

2013 年 6 月第一次印刷 印张:14

字数:276 000

定价:30.00 元

(如有印装质量问题,我社负责调换)

前 言

空间数据库技术是地理信息系统数据组织的核心技术,也是地理科学、测绘科学、计算机科学和信息科学相结合的产物。空间数据库技术已经成为空间数据管理的主流技术。由于空间数据的特殊性,空间数据管理在为计算机和信息科学作贡献的同时,也吸取计算机主流技术的各项最新成果,成为计算机领域中应用研究技术之一。

空间数据库技术的主要任务是研究空间物体的计算机数据表示方法、数据模型及计算机内的数据存储结构和建立空间索引的方法,研究如何以最小的代价高效存储和处理空间数据,正确维护空间数据的现实性、一致性和完整性,为用户提供现实性好、准确性高、完备、开放和易用的空间数据。空间数据库技术是理论性和实践性很强的学科,理解起来也较为抽象。编写本书的目的正是帮助读者更好地理解空间数据库的基本概念和构建方法。

本书作者是多年讲授数据库系统概论与空间数据库原理的一线教师。本书是作者结合自己的教学经验和教学体会,参考国内外的有关数据库和空间数据库资料编写而成。本书的特点在于能够把数据库系统原理和空间数据库原理有机结合起来,理论叙述严谨,培养目标明确,使读者在学习过程中做到一般数据库的基本原理与空间数据库原理相结合;内容叙述力求简单明了、深入浅出,在数据库原理相关章节尽量避免冗长的理论叙述,同时更侧重于学生空间知识的学习和空间思维能力的培养,使学生学完本门课程后能掌握一般数据库和空间数据库的基本知识,快速适应实际工作的需要。

全书共 13 章,参考学时为 51 学时。使用者可以根据实际情况进行适当的取舍。

第 1 章数据库系统概论,主要介绍了数据库的基本术语、数据库系统的组成与结构,概述了数据库管理系统。

第 2 章空间数据库,主要介绍了空间数据库概述、空间数据库的发展现状、空间数据库与传统数据库的比较。

第 3 章数据模型与空间信息模型,主要介绍了数据模型、空间信息模型。

第 4 章关系数据库模型,主要介绍了关系数据结构及形式化定义、关系的完整性,以及关系代数。

第 5 章结构化查询语言,主要介绍了 SQL 的基本知识、SQL 的数据定义功能、SQL 的数据操作功能、SQL 的数据控制功能。

第 6 章空间查询语言,主要介绍了处理空间数据的扩展 SQL。

第 7 章关系模式的规范化理论,主要介绍了关系模式的函数依赖、码和关系模式

的各个范式。

第 8 章数据库设计与实施,主要介绍了数据库系统设计的概念和特点,数据库的设计步骤、需求分析、概念结构设计、逻辑结构设计、数据库物理结构设计,以及数据库的实施和维护等。

第 9 章空间数据存储和空间数据索引,主要介绍了空间数据存储、空间数据组织、栅格数据管理,以及空间数据索引等。

第 10 章数据查询优化,主要介绍了查询优化的一般策略、关系代数的等价公式和查询优化的一般步骤。

第 11 章空间查询处理与优化,主要介绍了空间操作计算、空间查询优化和空间索引结构分析。

第 12 章数据库技术新发展,主要介绍了面向对象数据库技术中的面向对象数据模型和面向对象数据库管理系统,以及分布式数据库的特点、模式结构和分类等。

第 13 章空间数据库发展趋势,主要介绍了支持场实体的数据库,基于内容的检索等。

本书在编写过程中得到了单位领导和同仁的热情帮助和支持,在此表示衷心的感谢!本书的编写参考了广大同行专家的著作和成果,在此对他们表示衷心的感谢!

由于时间仓促,加之作者水平有限,书中难免有疏漏和不足之处,恳请同行专家和广大读者批评指正。

编著者

2013年3月30日

目 录

前言	1
第 1 章 数据库系统概论	1
1.1 数据库的基本概念	1
1.1.1 数据	1
1.1.2 数据库	1
1.1.3 数据库管理系统	2
1.1.4 数据库系统	2
1.2 数据库系统的组成与结构	3
1.2.1 数据库系统的组成	3
1.2.2 数据库系统结构	4
本章小结	5
习题	6
第 2 章 空间数据库	7
2.1 空间数据库概述	7
2.1.1 空间数据库基本概念	7
2.1.2 空间数据类型	9
2.1.3 空间数据特征	10
2.1.4 空间数据库作用	12
2.2 空间数据库的发展现状	14
2.3 空间数据库与传统数据库的比较	16
2.3.1 空间数据库特征	16
2.3.2 空间数据库与传统数据库的差异	17
本章小结	19
习题	19
第 3 章 数据模型与空间信息模型	20
3.1 数据模型	20
3.1.1 数据模型的概念	20
3.1.2 概念模型	20
3.1.3 层次模型	22
3.1.4 网状模型	23
3.1.5 关系模型	23

3.2	空间信息模型	24
3.2.1	基于场的模型	26
3.2.2	基于对象的模型	27
3.2.3	空间数据类型	27
3.2.4	空间对象的操作	28
3.2.5	动态空间操作	32
	本章小结	32
	习题	33
第4章	关系数据库模型	34
4.1	关系模型概述	34
4.2	关系数据结构及形式化定义	34
4.2.1	关系的数学定义	34
4.2.2	关系的性质	36
4.2.3	关系模式	36
4.3	关系的完整性	36
4.3.1	实体完整性	37
4.3.2	参照完整性	37
4.3.3	用户定义的完整性	38
4.4	关系代数	38
4.4.1	传统的集合运算	38
4.4.2	专门的关系运算	39
	本章小结	45
	习题	45
第5章	结构化查询语言	46
5.1	SQL 语言基本知识	46
5.2	数据定义命令	47
5.3	数据查询语言	50
5.4	数据更新语言	60
5.5	视图	63
5.6	数据控制	68
	本章小结	70
	习题	70
第6章	空间查询语言	71
6.1	World 实例数据库	71
6.2	扩展 SQL 以处理空间数据	73
6.3	强调空间的查询示例	76

本章小结	80
习题	80
第7章 关系模式的规范化理论	82
7.1 问题的提出	82
7.2 关系的规范化	83
本章小结	89
习题	89
第8章 数据库设计与实施	90
8.1 数据库设计概述	90
8.1.1 数据库设计内容	90
8.1.2 数据库设计的特点	90
8.1.3 数据库设计的步骤	91
8.2 需求分析	92
8.2.1 需求分析的任务	92
8.2.2 需求分析的步骤	93
8.2.3 需求分析的方法	93
8.2.4 需求分析的结果	93
8.3 概念结构设计	97
8.3.1 概念结构设计的目标和任务	97
8.3.2 概念结构的设计方法	98
8.3.3 概念结构设计的步骤	98
8.3.4 概念设计的结果	104
8.4 逻辑结构设计	104
8.4.1 逻辑结构设计的目标和任务	104
8.4.2 逻辑结构设计的方法和步骤	104
8.5 物理结构设计	108
8.5.1 物理设计的内容和方法	108
8.5.2 物理设计的评价	110
8.6 数据库的实施和维护	110
8.6.1 数据库的实施	110
8.6.2 数据库的试运行	112
8.6.3 数据库的运行和维护	112
本章小结	113
习题	113
第9章 空间数据存储和空间数据索引	115
9.1 引言	115

08	9.2 空间数据存储	116
08	9.2.1 磁盘的存储结构	117
88	9.2.2 缓冲区管理器	118
88	9.2.3 域、记录和文件	118
88	9.2.4 文件结构	120
08	9.2.5 空间填充曲线	121
08	9.3 空间数据组织	127
00	9.3.1 纵向分层组织	127
00	9.3.2 横向分块组织	131
00	9.3.3 分层分块索引	132
00	9.3.4 三维空间数据组织	133
10	9.4 栅格数据管理	134
80	9.4.1 管理方案	134
80	9.4.2 组织形式	135
80	9.4.3 存储结构	135
00	9.5 空间数据索引	136
80	9.5.1 网格文件	138
70	9.5.2 R树	140
70	本章小结	144
80	习题	145
	第10章 数据查询优化	147
101	10.1 查询优化概述	147
101	10.2 查询实例分析	148
101	10.3 查询优化的一般策略	150
101	10.4 关系代数的等价公式	151
801	10.5 查询优化的一般步骤	152
801	本章小结	154
011	习题	154
	第11章 空间查询处理与优化	155
011	11.1 空间操作计算	155
811	11.1.1 概述	155
811	11.1.2 空间操作	155
811	11.1.3 对象操作的两步查询处理	157
811	11.1.4 空间选择技术	158
811	11.1.5 一般的空间选择	159
811	11.1.6 空间连接操作算法	159

11.1.7 空间聚集操作策略:最近邻居	162
11.2 空间查询优化	163
11.2.1 逻辑转换	164
11.2.2 基于代价的优化:动态规划	167
11.3 空间索引结构分析	168
11.3.1 枚举可选的计划	170
11.3.2 混合体系结构中的分解与归并	170
本章小结	172
习题	173
第 12 章 数据库技术新发展	175
12.1 数据库技术概述	175
12.2 面向对象数据库技术	176
12.2.1 面向对象方法学简介	176
12.2.2 面向对象数据模型	178
12.2.3 面向对象数据库管理系统	182
12.3 分布式数据库	188
12.3.1 分布式数据库系统产生	189
12.3.2 分布式数据库系统的定义	190
12.3.3 分布式数据库系统的主要特点	192
12.3.4 分布式数据库系统的模式结构	194
12.3.5 分布式数据库管理系统及其分类	195
12.3.6 分布式数据库系统的优缺点	197
本章小结	199
习题	199
第 13 章 空间数据库发展趋势	200
13.1 支持场实体的数据库	200
13.1.1 栅格与图像操作	200
13.1.2 存储和索引	204
13.2 基于内容的检索	205
13.2.1 拓扑相似性	206
13.2.2 方位相似性	207
13.2.3 距离相似性	207
13.2.4 属性关系图	208
13.2.5 检索步骤	209
本章小结	209
习题	210
主要参考文献	212

第 1 章 数据库系统概论

在计算机的三大主要应用领域(科学计算、数据处理和过程控制)中,数据处理是其中的重要方面。数据库技术作为数据处理中的一门技术,是于 20 世纪 60 年代末 70 年代初产生和发展起来的。

数据库技术是计算机数据管理技术发展的新阶段,其所研究的问题就是如何科学地组织和存储数据,如何高效地获取和处理数据。

作为数据管理的主要技术,数据库技术目前已广泛应用于各个领域,在事务处理、情报检索、人工智能、专家系统、计算机辅助设计等方面都表现出了强大的功能,成为了计算机系统的重要组成部分。近年来,数据库技术还和计算机网络技术的发展相互渗透、相互促进,使得它成为了当今计算机领域发展迅速、应用广泛的一门技术。

1.1 数据库的基本概念

数据库技术涉及许多基本概念,主要包括数据、数据处理、数据库、数据库管理系统及数据库系统等。

1.1.1 数据

数据是指人们用来反映客观世界而记录下来的可以鉴别的数字、字母或符号,可以存储在某一种媒体上。所以数据的概念包括两方面的含义:一是描述事物特性的数据内容,也就是我们常说的信息;二是存储在某一种媒体上的数据形式,即符号。数据是承载信息的物理符号或载体,而信息是数据的内涵。用一定的符号表示信息,符号可以是多种多样的,而采用什么符号完全是人为规定的。同一信息可以有不同的数据表示方式,而同一数据也可以有不同的解释。

现在数据已经有了更广泛的含义,除了数字、字母、文字和其他特殊字符组成的文本形式数据,图形、图像、动画、影像、声音(包括语音、音乐)等多媒体数据也已经成为计算机的处理对象。

1.1.2 数据库

在社会飞速发展的今天,人们需要处理的数据量急剧增加,过去的手工管理已经远远不能满足人们对于信息的渴求,借助计算机软件技术来保存和管理复杂、大量的数据正是数据库技术应运而生的原因。

数据库是指长期存储在计算机内、有组织、可共享、可以表现为多种形式的数据集。

数据库中的数据具有如下特点。

- (1) 按一定的数据模型组织、描述和存储。
- (2) 具有较小的冗余度。
- (3) 具有较高的数据独立性和易扩展性。
- (4) 可被用户共享。

1.1.3 数据库管理系统

数据库管理系统(Data Base Management System, DBMS)是位于用户与操作系统之间的一层数据管理软件,正如使用高级语言需要解释/编译程序的支持一样,使用数据库语言也需要一个特定的支持软件,这就是“数据库管理系统”。

数据库管理系统的主要任务是科学有效地组织和存储数据、高效地获取和管理数据、接受和完成用户提出的访问数据的各种请求。

数据库管理系统的主要功能包括以下几个方面。

(1) 数据定义功能。提供数据定义语言(Data Definition Language, DDL),用户通过它可以方便地对数据库中的数据对象进行定义。例如对数据库、表、索引进行定义。

(2) 数据操作功能。提供数据操作语言(Data Manipulation Language, DML),用户通过它可以实现对数据库的基本操作。例如对表中数据的查询、插入、删除和修改等。

在微机数据库管理系统中,DDL 和 DML 通常合二为一,构成一体化的语言。

(3) 数据库运行控制功能。包括并发控制(即处理多个用户同时使用某些数据时可能产生的问题)、安全性检查、完整性约束条件的检查和执行、数据库的内部维护(如索引的自动维护)等。这是数据库管理的核心部分。数据库在建立、运用和维护时所有操作都要由这些控制程序统一管理、统一控制,以保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

(4) 数据库的建立和维护功能。包括数据库初始数据的输入、转换功能,数据库的转储恢复功能,数据库的重新组织功能和性能监视、分析功能等。这些功能通常是由一些实用程序完成的。它是数据库管理的一个重要组成部分。

1.1.4 数据库系统

数据库系统(Data Base System, DBS)是指拥有数据库技术支持的计算机系统,一般由数据库、数据库管理系统及其开发工具、应用系统、数据库管理员和用户构成。数据库系统可以实现有组织地、动态地存储大量相关数据,提供数据处理和信息资源共享服务。

与文件系统相比,数据库系统具有以下特点。

(1) 数据的结构化。同一数据库中的数据整体上按照一定的结构形式存储,但数据文件之间相互有联系。

(2) 最小的冗余度。数据库系统中的数据是集成化的,只是在逻辑存储上存在重复,而在物理存储上不存在。实现数据共享后,将消除不必要的重复,但有时可保留少量冗余以提高查询效率。

(3) 数据的共享。不同的用户可以使用同一数据库资源,这样可以节约存储空间。

(4) 数据与程序独立。数据按照某种规则,以能反映数据之间内在联系的形式组织在库文件中,数据不会受到应用程序变化的影响,数据的变动也不会影响应用程序。

(5) 数据的安全性和完整性。数据的安全性是指数据的保密性,数据的完整性是指数据的正确性、有效性和相容性。数据库系统提供了管理和控制数据的各种操作命令及程序设计语言,使用户控制数据库,例如设置口令来保证数据安全,防止数据被破坏或窃取。

1.2 数据库系统的组成与结构

1.2.1 数据库系统的组成

数据库系统是指引进数据库技术后的计算机系统。数据库系统一般由支持数据库运行的软硬件、数据库、数据库管理系统、数据库管理员和用户等部分组成,如图1-1所示。



图 1-1 数据库系统的组成

(1) 硬件与软件。与数据库系统密切相关的硬件主要有 CPU 及存储设备等,要求系统有较高的通道能力,以提高数据传输率。数据库系统的软件主要包括 DBMS,支持 DBMS 运行的操作系统,具有数据库接口的机器语言编译系统,以 DBMS 为核心的应用开发工具以及为特定应用环境开发的数据库应用系统。

(2) 数据库管理系统。一个以统一的方式管理、维护数据库中数据的一系列软件的集合。

(3) 数据库系统。指在计算机系统中引进数据库后的系统构成。

(4) 数据库管理员(DataBase Administrator, DBA)。负责建立、维护和管理数

据系统的人员。具体职责包括决定数据库中的信息内容和结构,决定数据库的存储结构和存取策略,定义数据的安全性要求和完整性约束条件,监控数据库的使用、运行和数据库的改进、重组重构。大型数据库通常由专业人员设计,还要有专职的数据库管理员进行管理。

(5) 用户。数据库系统的用户分两类:最终用户和专业用户。最终用户包括偶然用户、简单用户和复杂用户。专业用户主要是应用系统开发人员。

1.2.2 数据库系统结构

从数据库最终用户的角度看,数据库结构分为集中式、分布式、客户/服务器和并行结构等,但从数据库管理系统的角度看,数据库系统通常采用三级模式结构,这是数据库管理系统内部的系统结构。

数据库的三级模式组织结构,是美国国家标准学会(ANSI)所属的标准计划和要求委员会(Standards Planning And Requirements Committee, SPARC)在 1975 年公布的关于数据库标准报告中提出的。数据库的三级模式组织结构也称为 SPARC 分级结构。

1. 模式的概念

模式(Schema)是数据库中全体数据的逻辑结构和特征的描述,它仅仅涉及“型”的描述,不涉及具体的“值”。其中,型是数据模型中对某一类数据结构和属性的说明,而值是型的一个具体赋值。模式具有如下特点。

- (1) 模式的一个具体描述称为模式的一个实例。
- (2) 模式是相对稳定的,而实例是相对变动的。
- (3) 模式反映的是数据的结构及其联系,而实例反映的是数据库某一时刻的状态。

2. 数据库系统的三级模式结构

数据库系统的三级模式结构是指数据库系统由外模式、模式、内模式三级构成,如图 1-2 所示。

(1) 模式。模式也称逻辑模式或概念模式,是数据库中全体数据的逻辑结构和特征的描述,是所有用户的公共数据视图,是数据库系统模式结构的中间层。一个数据库只有一个模式。模式是数据项值的框架。数据库系统模式通常还包含访问控制、保密定义、完整性检查等方面的内容。

(2) 外模式。外模式也称为子模式或用户模式,它是数据库用户(程序员和最终用户)能够看见和使用的局部数据的逻辑结构和特征的描述,是数据和用户的数据视图,是与某一应用有关的数据的逻辑表示。

外模式一般是模式的子集。一个模式可以有多个外模式。一个应用程序只能使用一个外模式。外模式是保证数据库安全性的一个有力措施。

(3) 内模式。内模式也称存储模式,是数据库在物理存储器上具体实现的描述,是数据在数据库内部的表示方法,也是对数据物理结构和存储方式的描述。一个数

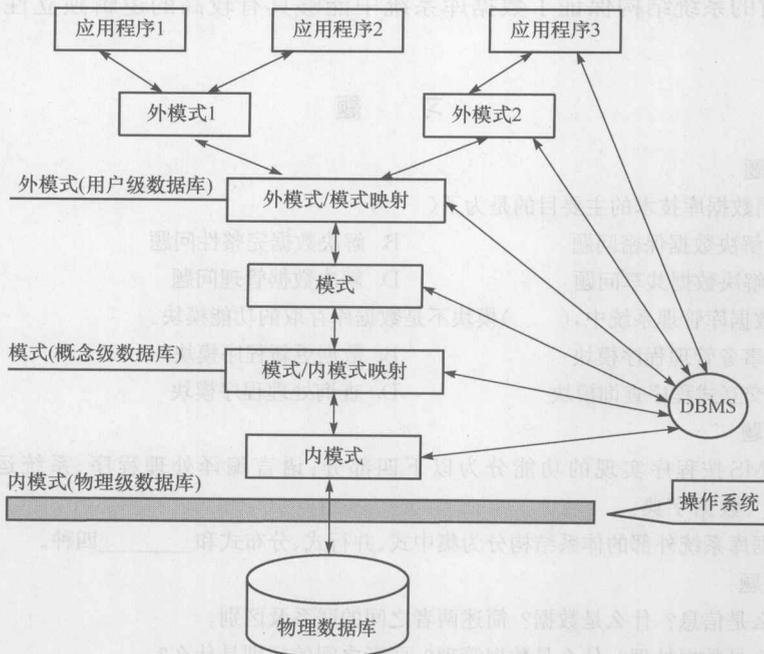


图 1-2 数据库系统的三级模式结构

数据库只有一个内模式。

3. 数据库的二级映射功能与数据独立性

数据库的三级模式是对数据的三个抽象层次。为了能够在内部实现这三个抽象层次的联系和转换,数据库管理系统在这三级模式之间提供了两层映射,即外模式/模式映射和模式/内模式映射,使得数据具有了较高的逻辑独立性和物理独立性。

(1) 外模式/模式映射。对于每一个外模式,数据库管理系统都有一个外模式/模式映射,它定义了外模式与模式之间的对应关系。当模式改变时,由数据库管理员对各个外模式/模式映射做相应的改变,这里的映射是把用户数据库与概念数据库联系起来,从而使外模式保持不变,这就保证了数据库的逻辑独立性。

(2) 模式/内模式映射。数据库只有一个模式,也只有一个内模式,所以模式/内模式的映射是唯一的,这种映射把概念数据库与物理数据库联系起来,定义了数据全局逻辑结构与存储结构之间的对应关系。当数据库的存储结构改变了,由数据库管理员对模式/内模式的映射作相应的改变,可以使模式保持不变,从而应用程序也不必改变。这就保证了数据与程序的物理独立性。

本章小结

本章首先通过介绍数据库的基本概念,说明了数据库的优点。然后阐述了数据库系统的组成与结构,尤其强调了人-机系统理念,并且说明了数据库系统三级模式

第2章 空间数据库

2.1 空间数据库概述

2.1.1 空间数据库基本概念

1. 数据

数据是指客观事物的属性、数量、位置及其相互关系等的符号描述。

在空间数据库中,数据可以是一个数,如某一点的高程值、一个多边形的面积等,也可以是一组符号组成的字符串,如一个地名、一个河流注记、一幅图像。空间数据库中的数据大多与地理位置有关,一般称为空间数据(Spatial Data)。空间数据不同于普通的数据,它具有空间性、时间性、多维性和大数据量等特点,而且数据之间不仅有传统的关联关系,还有空间关系,这就给数据的处理和利用带来了更多的难度。

2. 空间

空间(Space)是一个应用很广泛的名词。空间是客观存在的物质空间,是一个复杂的概念,具有多义性,既有与时间对应的含义,也有“宇宙空间”的含义。空间可以定义为一系列结构化物体及其相互间联系的集合。日常语义上的“空间”是指事物之间的距离或间隔。从感观角度将空间看做目标或物体所存在的容器或框架,因此空间更倾向于被理解为物理空间。空间知识的本质问题是一个古老的研究领域,关于空间的概念,不同的学科有不同的解释。哲学家、天文学家、物理学家对空间的论述众说纷纭。

从物理学角度看,空间为宇宙在三个相互垂直的方向上所具有的广延性;天文学中的空间指宇宙空间,是时空连续体系的一部分;行星物理和相关地球物理中的空间常指地球高层空间和行星际空间;在数学中空间概念的范围很广,一般指某种对象(现象、状况、图形、函数等)的任意集合,其中要求说明“距离”或“邻域”的概念;从地理学的意义上讲,空间是人类赖以生存的地球表层具有一定厚度的连续空间域,是一个定义在地球表层空间实体集上的关系。GIS领域的空间是指地理环境或地球表层空间,是地理信息系统表达和研究的对象。为了在GIS中对地理空间(Geospace)进行描述,常常需要借助于抽象的数学空间表达方法。

3. 地理空间

地球表面上的一切地理现象、地理事件、地理效应、地理过程都发生在以地理空间为背景的基础之上。地理学的空间定义了地球表层空间实体集上的关系,在空间实体之间有无数种关系,定义一种关系自然就定义了一种空间,而这个空间又是和几