

第 1 章 MapGIS 地理数据库

1.1 数据模型

1.1.1 模型的概念层次

MapGIS 空间数据模型的概念分 6 个层次：地理数据库、数据集、类、几何元素、几何实体、坐标点，如图 1.1-1 所示。

非空间实体被抽象为对象，空间实体被抽象为要素；相同类型的要素构成要素类；相同类型的对象构成对象类；若干对象类或要素类组成要素数据集；若干要素数据集构成地理数据库。

要素在某个空间参照系中的几何特征被抽象为图形信息，图形信息由任意的点状、线状或面状几何实体组成，几何实体通过几何坐标点表达。

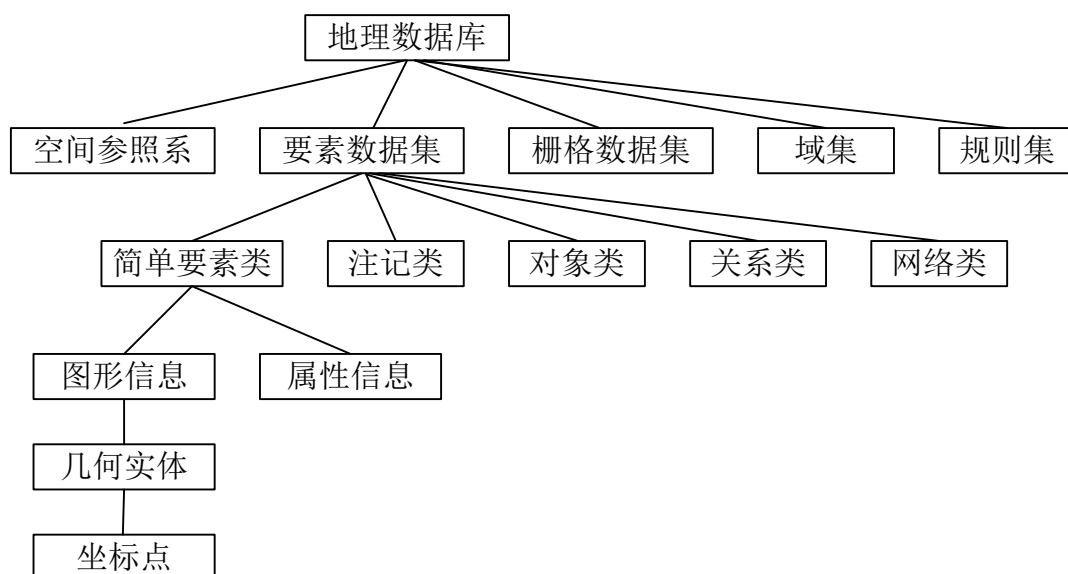


图 1.1-1 MapGIS 空间数据模型概念层次

1.1.2 模型的特点

MapGIS 的空间数据模型将现实世界中的各种现象抽象为对象、关系和规则，各种行为（操作）基于对象、关系和规则，模型更接近人类面向实体的思维方式。该模型还综合了面向图形的空间数据模型的特点，使得模型表达能力强，广泛适应 GIS 的各种应用。该模型具有以下特点：

真正的面向地理实体，全面支持对象、类、子类型、关系、有效性规则、数据集、地理数据库等概念。

对象类型覆盖 GIS 和 CAD 对模型的双重要求，包括：要素类、对象类、关系类、注记类、几何网络。

简单要素类可描述任意几何复杂度的实体，如水系、道路等。

完善的关系定义，可表达实体间的空间关系、拓扑关系和非空间关系。拓扑关系支持结构表达方式和空间规则表达方式；完整地支持 3 类非空间关系，包括关联关系、继承关系(完全继承或部分继承)、组合关系（聚集关系或组成关系）。

支持关系多重性，包括 1-1、1-M、N-M。

支持有效性规则的定义和维护，包括定义属性规则、关系规则、拓扑规则、边-边规则、边-结点连接规则。

支持多层次数据组织，包括地理数据库、数据集、数据包、类、图形元素，如图 1.1-2 所示。

图形元素支持向量表示法和解析表示法，包括折线、圆、椭圆、弧、矩形、样条、bezier 曲线等形态，能够支持规划设计等应用领域。

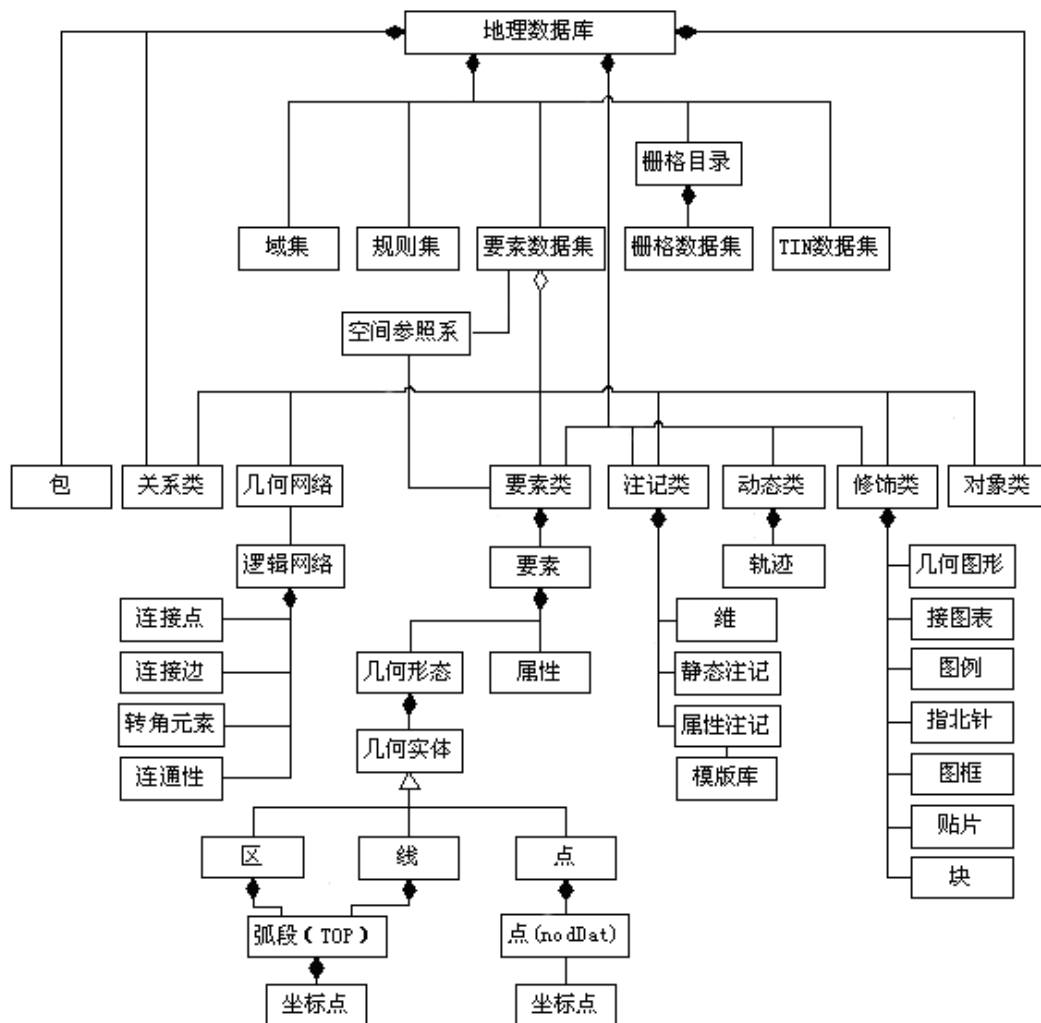


图 1.1-2MapGIS 面向实体的空间数据模

1.2 系统特点

1、分布式跨平台可拆卸的多层多级体系结构：

最新的第四代多层结构体系

具备完全支持“全球空间网格”能力

.net 和 j2ee 架构

分布式全组件化的跨平台系统

面向互连网的系统设计

面向“服务”的最新思想

基于 GML 的开放式接口

适应异构数据库的多级服务器协同工作环境

2、面向地理实体的空间数据模型：

面向地理实体的抽象模型

可描述任意复杂度的空间特征和非空间特征的地理实体特征

完全表达空间、非空间、非空间的多重性、实体的空间共生性的关系

面向实体语义关系的操作

统一了 GIS 与 CAD 对模型要求的面向实体的信息可视化

3、海量空间数据存储与管理：

TB 级的空间数据存储与处理能力

矢量、栅格、三维、影像四位一体的海量数据存储

异构数据库的多级服务器数据更新与同步

完全一致的存储无关的概念模型（文件系统或 RDBMS）

基于版本和数据锁的长事务解决机制

高效的空间索引（矩形索引、R 树索引、聚集索引、格网拼合索引、四叉树索引）

4、时空处理：

采用“元组级基态+增量修正法”实施方案

版本与增量相结合的时空数据模型

元组级的时空数据控制粒度

可实现单个实体时态演变

“事件”作为时态追踪的参考点

通过时态数据索引管理

任意时刻的历史回溯

多用户并发的历史事件的控制

5、真三维建模与可视化：

三维海量数据的有效存储和管理

三维模型数据一体化管理（TIN、三维景观、三维地质）

三维数据的 LOD_RTree 索引组织技术

面向实体和拓扑的数据组织

三维数据专业模型的快速建立

高程数据 TIN/GRD 模型的建立、处理等基本功能

三维地质构造建模、断层处理技术

地质体内属性三维分布建模技术

三维数码景观动态建模技术

三维数据的综合可视化和融合分析

基于拓扑的三维剖切分析

基于拓扑的等值面提取

三维体数据的面绘制技术

三维体数据直接体绘制技术

6、空间信息应用服务：

提供基于 SOAP 和 XML 的空间信息 Web Services

遵循 OpenGIS 规范，支持 WMS、WFS、WCS 等标准，以及 XML 和 GML3 标准

支持互联网和无线互联网，支持各种智能移动终端

提供各类高速缓存、无状态的负载平衡策略，满足高速度访问的需要

提供用户权限的控制和安全策略

提供空间分析、以及应用逻辑分析等服务，满足对空间数据库的专业查询和分析

7、版本与长事务处理：

长事务期间，可以自由地编辑要素、执行地理分析、编辑地图

长事务完时，如被实施，则更新到地理数据库中，否则丢弃

使用乐观的并发访问控制技术，实现长事务机制，没有对要素加锁

允许产生编辑冲突，当提交事务时，检测冲突，并协调解决冲突

版本控制使多个用户可直接编辑数据而不用锁定要素或复制数据

版本管理具有版本创建、删除、归并、冲突解决等功能和机制

8、工作流管理：

基于网络拓扑数据模型的工作流控制引擎

实现了业务的灵活调整和定制，解决了 GIS 和 OA 的无缝集成

符合国际工作流联盟制定的规范

不同业务流程之间的交叉、融合

历史案件的办理过程不受模板变化的影响

通过拓扑关系能够自动实现条件判断、循环、会签等功能

工作流“可扩充”性与动态表单可“自定义”性

支持多级子表和数据字典

9、空间元数据：

元数据采集、编辑和录入

元数据读取、查询和共享发布；

面向 Web 的客户端操作界面；

支持 SRW 协议（新一代 Z3950 协议），分布式检索能力强

1.3 地理数据库

MapGIS 地理数据库（GeoDatabase，简称 GDB）新概念，它集成了地理数据库创建、管理、浏览等多种功能。GDBCatalog 目录树，记录了 GDB 的信息，包括各种类，如：简单要素类、对象类、注记类、网络类、栅格数据集、元数据库、地图集等。

1.3.1 地理数据库的数据组织

MapGIS 按照“地理数据库—数据集—类”这几个层次组织数据，以满足不同应用领域对不同专题数据的组织和管理需要。

1. 地理数据库

地理数据库是面向实体空间数据模型的全局视图，完整地、一致地表达了被描述区域的地理模型。一个地理数据库包括 1 个全局的空间参照系、1 个域集、1 个规则集、多个数据集、元数据库、地图集和各种对象类。

2. 数据集

数据集是地理数据库中若干不同对象类的集合，通过命名数据集提供了一种数据分类视图，便于数据组织、管理和授权。根据不同的用途，数据集分为：要素数据集、栅格目录、栅格数据集。

3. 要素数据集

要素数据集是地理数据库中具有相同空间参照系的要素类、对象类、关系类、注记类、几何网络的集合；

在一个要素数据集下存放不同的类，有利于以下情况的数据组织：

专题归类：当不同的要素类属于同一范畴时，可将其组织到一个要素数据集中。例如：全国范围内某种比例尺的地质图数据。

共享几何特征：如用地、水系、行政区界等不同的要素类需要共享公共几何边界时，将

其组织到一个要素数据集中，当移动其中的一个要素时，公共的部分也要求一起移动，并可保持这种公共边的关系不变。

创建几何网络：同一个几何网络中作为边要素和点要素的各种要素类，必须组织到一个要素数据集中，以便于管理和建立网络模型。例如：通讯网络中，各种交换机、交接箱、路由器、电缆、光缆等，分别对应点要素类和线要素类，在通讯网络建模时,将其全部考虑到通讯网络对应的几何网络模型中。

4. 栅格目录

栅格目录用于管理有相同空间参照系的多幅栅格数据，各栅格数据在物理上独立存储，易于更新，常用于管理更新周期快、数据量较大的影像数据。同时，栅格目录也可实现栅格数据和栅格数据集的混合管理，其中目录项既可以是单幅栅格数据，也可以是地理数据库中已经存在的栅格数据集，具有数据组织灵活、层次清晰的特点。

5. 栅格数据集

栅格数据集用于管理具有相同空间参照系的一幅或多幅镶嵌而成的栅格影像数据，物理上真正实现数据的无缝存储，适合管理 Dem 等空间连续分布、频繁用于分析的栅格数据类型。由于物理上的无缝拼接，因此以栅格数据集为基础的各种栅格数据空间分析具有速度快、精度较高的特点。

6. 存储策略

地理数据库存储和管理地理数据，它采取基于文件和基于商业数据库两种存储策略。由于这两种存储策略支持相同的空间数据模型，因此在文件和数据库之间能够实现无损的平滑的数据迁移；同时，两种策略具有共同的平台，这使得上层软件不需要因为数据迁移而改变。

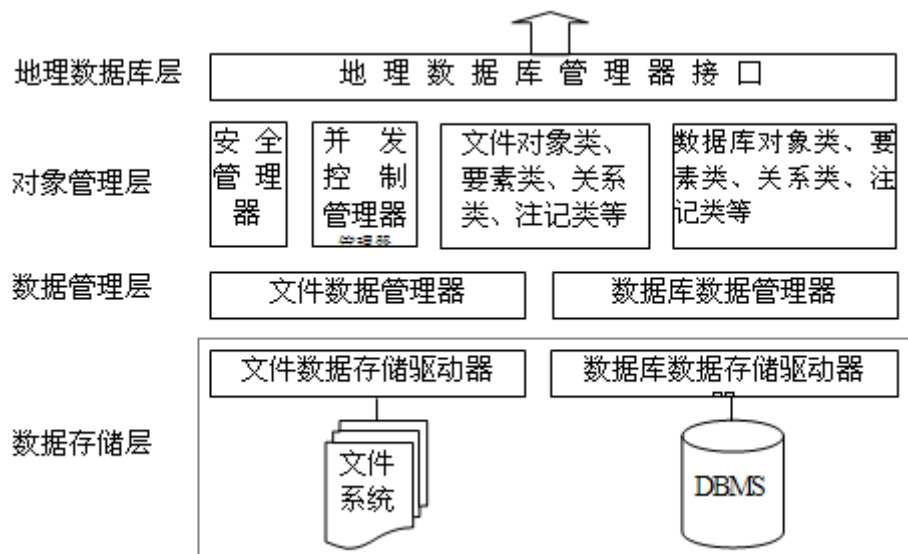


图 1.3-1 MapGIS 存储策略概念图

针对不同的应用规模和应用阶段，给用户提供了多种最佳的性价比和最大的投资收益率选择方案。例如：

应用规模小的用户、二次开发团体、教学单位、数据累积规模较小的用户都可选择基于文件的存储策略，以节省昂贵的商业数据库费用；

大型、超大型应用可选择基于商业数据库的存储策略；

分多个阶段进行开发的应用，在前期阶段，数据规模较小，用户不多，在后期阶段数据规模大，用户多，则可先采用文件存储策略，再购买适当许可数的商业数据库和服务器设备，以后根据数据规模和业务情况再增加数据库许可数和服务器等软硬件设备。这不仅提高了用户的资金利用率，而且软硬件性能迅速提高，让用户享受到多重好处。

1.3.2 地理数据库的数据管理

MapGIS 的分布式数据管理体系是采取“纵向多级、横向网格”的组网方案，如下图所示，在级与级之间，节点与节点之间的连接是采用一种“松耦合”方式。“松耦合”方式是互连网的最佳耦合方式，它受网络环境影响最小。分布式数据的存取操作采取面向“服务”方式进行，就是把“进行数据存取操作”变为“请求数据存取服务”，“数据存取服务”是所有“服务”的特例，充分体现“面向服务”的最新设计思想。

MapGIS 分布式数据管理的面向“地理实体”的增量式订阅和发布技术有效地支持分布式数据的增量更新与同步。由于采用面向“服务”设计思想和面向“地理实体”的数据模型

相结合，克服了传统分布式数据库面向“记录”的增量式订阅和发布只能用于“同构数据库”的缺点，使网格节点之间、父节点与子节点之间，因不同操作系统、不同数据库平台、不同数据大小而产生的“异构数据库”可实现增量更新与同步。空间数据的增量更新与同步解决了“海量空间数据在互联网上调用速度问题”。

MapGIS 分布式数据管理是跨平台的，按照“面向服务”的思想，每个节点上的数据“管理者”必须提供“服务”，在“谁管数据谁提供服务”的原则基础下，可解决网格节点之间、父节点与子节点之间、不同平台不同系统之间数据不通问题，因为从用户角度来说，它不考虑数据只要服务。因此，管理数据的软件必须提供数据服务，“应用端请求服务而不是直接操作数据”；“服务端提供服务而不是提供数据内部结构”。各个站点管理数据的软件、提供应用服务功能的软件都可以由不同的厂商提供。

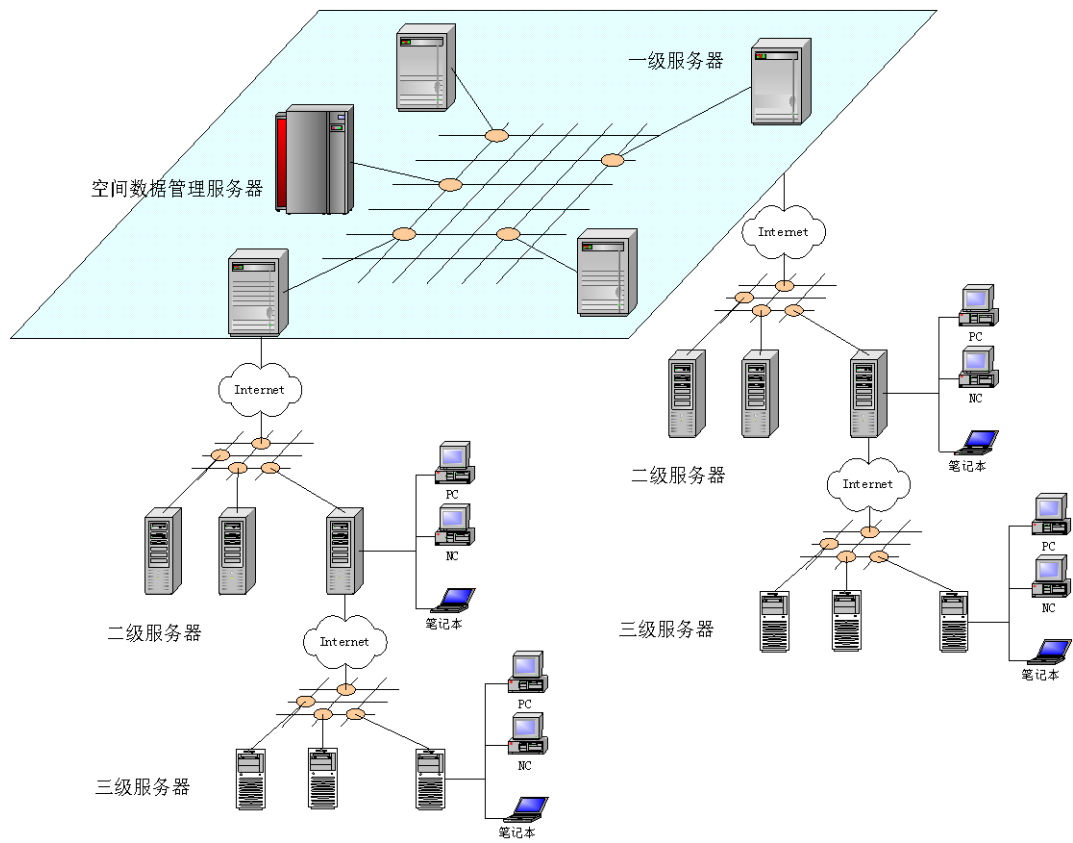


图 1.3-2 MapGIS 分布式数据管理原理图

第 2 章 空间数据

空间数据是描述地球表层一定厚度、一定范围内的地理事物及其关系的数据。

2.1 空间数据的基本特性

2.1.1 数据的基本特性

数据是对事物的复写和描述，可以文字、数字、图形、影像、声音等多种方式存在，但是数据不是事物本身，这是因为：数据只能从有限的方面去描述事物，而不可能也没必要全面、详尽、保真的复制事物本身。因此，一般地说，数据所反映的只是事物的某些特性。

数据是以诸如人工统计、仪器测量、社会调查等多种方式获得的，这必然导致甚至是必然存在的各种误差，如人为差错、仪器的系统误差等等。此外，数据与其描述对象——事物是以不同方式存在的，从一种存在方式转化为另一种存在方式，其间的转换误差肯定是存在的，而且还有可能是极为复杂的。如地图是关于地理事物的数据的图形表达，地图是以二维平面方式存在的，而地理事物是存在于三维空间中的，其间要经过地图投影这一数学转换，这种转换必然会导致误差，并且还可能很大。

尽管数据不等于事物本身，但因为它可以从某些方面反映事物的特性，因此它是分析认识事物的重要途径。经过对数据的分析处理，使我们更清楚地对事物有所了解，掌握了事物的某些特性，因此获得了事物的有关信息。

一般地说，数据具有以下基本特性：

（1）选择性数据只是从某一（些）侧面描述事物本身，这些侧面就是被选择的事物的属性和角度，数据的选择性还不仅仅指这种“侧面”的取舍，同时，还有存在方式的选择，因为存在方式的选择将直接影响“侧面”的选择。用文字来描述事物与用数字描述事物显然直接涉及到“描述什么”的问题。而用图像来描述事物的选择性则直接就是视觉和视角的选择。

（2）可靠性如果不是直接获取的第一手资料，数据就必然存在是否可靠的问题。有些数据不是通过直接量测统计得到的，而是分析估算而来，则更存在可靠性问题。即使采用的是精密的仪器、规范的统计方法，仍会有误差，影响数据的可靠性。数据在存储、管理、传播过程中发生差错也可能影响其可靠性。可靠性亦可被理解为正确性。

（3）时间性数据是静态的，而事物本身则可能是动态的、不断变化的，任何数据只能反映事物的某个特定时刻或特定时间段内的情况，在空间数据处理中，数据的时间性已构成一个重要的研究分支。

（4）完备性是指数据是否系统，是否有缺失。不完备的数据往往对分析造成困难，尤

其是数据的数值分析，没有完备的数据是不可能进行的。完备性有空间的完备性、时间的完备性和主题的完备性。空间的完备性指统计对象要全面，如对一个地区降水的观测，观测点要不满足整个地区。时间的完备性指每一个时间段和时间点上的数据都要具备。

（5）详细性指数据的分辨率，也就是可描述最细微差异的程度及最微小物体的大小。详细性根据数据的属性可分为空间详细性、时间详细性、质量详细性和数量详细性。空间详细性是指统计单元的级别高低，在非空间数据中，空间详细性可指统计（抽样）的基本单位，如关于高等学校科研成果的统计是以学校为单位、以院（系）为单位，还是以区为单位。

在空间数据中，空间详细性指空间统计单元的大小、地图比例尺、图像的分辨率、格网和抽样点的密度等等。时间详细性是指统计的周期、时间间隔的大小。质量的详细性是分类指标体系的详细性，一般地说，各类指标的分类体系会形成一个分类树，那么质量详细性意味着分类树中的哪一级分类被采用，如关于人口的统计，可以只统计总人口，也可以以区分性别、职业、学历等，不同的质量详细性将决定分析的可能性。数量详细性指计量的单位、分级情况、使用的量，如人口可以按百万、万、人等不同单位进行统计。

2.1.2 空间数据的基本特性

空间数据作为数据的一类而具有数据的一般特性，但它由具备自身的一些特性。空间数据的自身特性构成了空间分析的条件和任务。

1. 抽样性

空间物体以连续的模拟方式存在于地理空间，为了能以数字的方式对其进行描述，必须将其离散化，即以有限的抽样数据表述无限的连续物体。空间物体的抽样不是对空间物体的随机选取，而是对物体形态特征点的有目的的选取，其抽样方法根据物体的形态特征的不同而不同。抽样的基本准则就是能够力求准确地描述物体的全局和局部的形态特征。

2. 概括性

概括是地图数据处理的一种手段，是对地理物体的简化和综合。空间物体的概括性区别于前面所述的数据的详细性。空间数据的空间详细性反映人为规定的系统的数据分辨力，而空间物体的概括性指对物体形态的化简综合以及对物体的取舍。

3. 多态性

空间数据的多态性具有两层含义：一是同样地物在不同情况下的形态差异，二是不同地

物占据同样的空间位置。就形态而言，任何城市在地理空间都占据一定范围的地域，因此可以认为其是面状地物，但在比例尺较小的空间数据中，或者在相对宏观的分析中，城市是点状地物处理的。关于后者，大多表现为社会经济人文数据与自然环境数据在空间位置上的重叠，如长江是水系要素，但同时在不同的地段上，长江又与省界、县界相重叠。

4. 空间性

空间性是空间数据的最主要特性，它是指空间物体的位置、形态以及由此产生的系列特性。

2.2 空间数据的表示

空间数据表示的基本任务，就是将以图形模拟的空间物体表示成计算机能够接受的数字形式，因而空间数据的表示必然涉及到空间数据模式和数据结构的问题。

空间数据有两种基本的表示模型：栅格数据模型和矢量数据模型。

在栅格数据模型中，地理空间被划分为规则的小单元（像元），空间位置由像元的行、列号表示。像元的大小反映了数据的分辨率即精度，空间物体由若干像元隐含描述。例如，一条道路由其值为道路编码值的一系列相连的像元表示，要从数据库中删除这条道路，则必须将所有有关像元的值改变成该道路邻域的背景值。栅格模型的设计思想是将地理空间看成一个连续的整体，在这个空间中处处有定义。

矢量模型将地理空间看成是一个空域，地理要素存在其间。在矢量模型中，各类地理要素根据其空间形态特征分为点、线、面三类。点状要素用坐标点对表示其位置，线状要素用其中心轴线（或侧边线）上的抽样点的坐标串来表示它的位置和形状，面状要素用范围轮廓线上的抽样点坐标串表示其位置和范围，因此在矢量模型中，地物是显式描述的。

2.2.1 栅格数据模型

在栅格数据模型中，它的基本单元是一个格网，每个格网称为一个栅格（像元），被赋予一个特定值。这种规则格网通常采用三种基本形式：正方形、三角形、六边形（如图 2.2-1）。每种形状具有不同的几何特性。其一是方向性：正方形和六边形栅格数据模型中的所有格网都具有相同的方向，而三角形栅格数据模型中的格网却具有不同的方向；其二是可再分性：正方形和三角形格网都可以无循环地再细分成相同形状的子格网，而六边形格网则不能进行

相同形状的无限循环再分；其三是对称性：每个六边形格网的邻居与该六边形格网等距，也就是说该六边形格网的中心点到其周围的邻居格网的中心点的距离相等，而三角形和正方形格网就不具备这一特征。

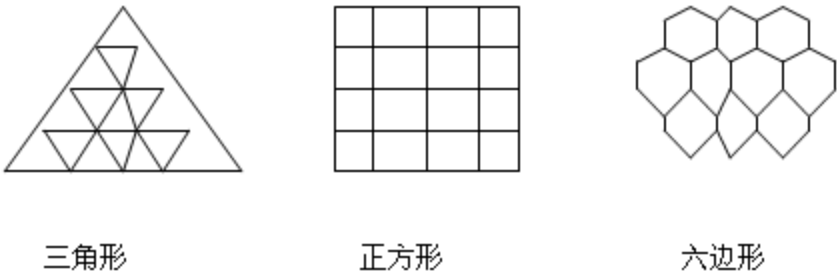


图 2.2-1 栅格数据模型中常用的三种基本单元

栅格模型中最常用的也是最简单的是正方形格网，除了因为它具有上述方向性和可再分性外，还因为大多数栅格地图和数字图像都采用了这种栅格数据模型。

栅格模型的缺点在于难以表示不同要素占据相同位置的情况，这是因为一个栅格被赋予一个特定的值。因而一幅栅格地图仅适宜表示一个主题（土地貌、土地利用等）。

在栅格模型中，栅格大小的确定是一个关键，根据抽样原理，当一个地物的面积小于 $1/4$ 个栅格时就无法予以描述，而只有面积大于一个栅格时才能确保被反映出来。从中可以发现很多栅格具有相同的值，数据冗余非常大。在地图数据库中，为了节约存储空间，通常不是直接存储每个像元的值，而是采用一定的数据压缩方法，常用的有“行程编码”和“四叉树”方法。

1. 行程编码

行程编码方法是逐行将具有相同取值的栅格用两个数值（ L, V ）表示， L 表示栅格数， V 表示栅格的值。行程编码法也可以是按列压缩，其压缩倍率与按行压缩一般不相等。行程编码的数据在处理时需要还原，因此研究数据压缩还应考虑到数据还原的可能性与方便性。

2. 四叉树

四叉树是一个分层的多分辨率的栅格地图表示方法。它首先将整幅图以 2×2 的方式四等分，如果其中任一栅格内属性不唯一，则将该栅格再进行 2×2 四等分，如此循环直至每个栅格内的值唯一，图 2.2-2 就是一个对多边形表示的面状地物的四叉树描述，每个像元的大小并不相等，各像元根据其所处层次的编号逐层记录，如图左上角的像元处于第一层，右上角的像元处于第二层左下角的像元处于第三层，以此类推。四叉树就其本质而言不仅是一种数

据压缩方法,还是一种数据结构,用四叉树表示的栅格地图一般无需还原即可进行数据分析。
四叉树曾经是一个相当活跃的研究领域,基于四叉树的算法非常丰富。

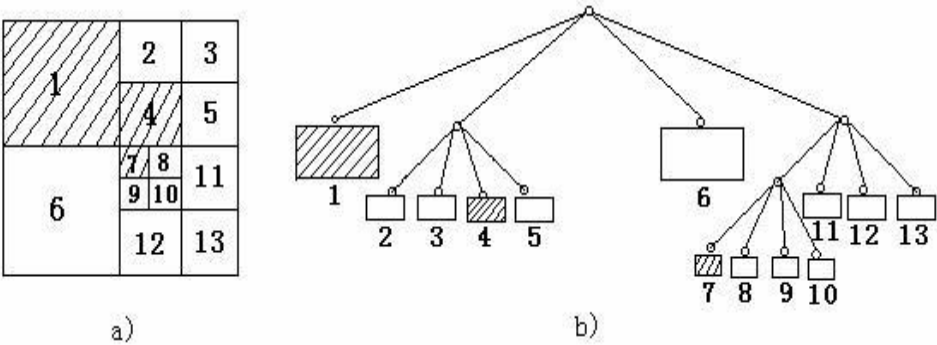


图 2.2-2 四叉树压缩编码原理图

2.2.2 矢量数据模型

矢量数据模型是以坐标点对来描述点、线、面三类地理实体。描述地理实体的矢量方法有许多, 这些不同的矢量数据模型间的一个最重要的差别是采用路径拓扑 (Path Topology) 模型, 还是采用网络拓扑 (Graph Topology) 模型。这两种模型间的主要区别在于前者将二维要素的边界作为独立的一维要素来单独处理, 而不考虑要素之间的相互关系; 而后者则是在一个关于边界的关系网络模型中来考查二维要素。

路径拓扑常用的有“面条模型” (Spaghetti Model)、“多边形模型” (Polygon Model)、 “点字典模型” (Point Dictionary Model)、 “链/点字典模型” (Chain/Point Dictionary Model)。

2.2.3 栅格数据模型和矢量数据模型比较

栅格模型与矢量模型似乎是两种截然不同的空间数据结构, 栅格模型“属性明显、位置隐含”, 而矢量模型“位置明显、属性隐含”, 栅格数据操作总的来说比较容易实现, 尤其是作为斑块图件的表示更易于为人们接受; 而矢量数据操作则比较复杂, 许多分析操作 (如两张地图的覆盖操作, 点或线状地物的邻域搜索等) 用矢量模型实现十分困难, 矢量模型表达线状地物是比较直观的, 而面状地物则是通过对边界的描述而表达。无论哪种结构, 数据精度和数据量都是一对矛盾, 要提高精度, 栅格模型需要更多的栅格单元, 而矢量模型则需记录更多的线段结点。一般来说, 栅格模型只是矢量模型在某种程度上的一种近似, 如果要使栅格模型描述的图件取得与矢量模型同样的精度, 甚至仅仅在量值上接近, 则数据也要比后

者大得多。

栅格模型在某些操作上比矢量模型更有效更易于实现，如按空间坐标位置的搜索，对于栅格模型是极为方便的，而对矢量模型则搜索时间要长得多；在给定区域内的统计指标运算，包括计算多边形形状、面积、线密度、点密度，栅格模型可以很快算得结果，而采用矢量模型则由于所在区域边界限制条件难以提取而降低效率，对于给定范围的开窗、缩放栅格模型也比矢量模型优越；另一方面，矢量模型用于拓扑关系的搜索则更为高效，即诸如计算多边形形状搜索邻域、层次信息等；对于网络信息只有矢量模型才能完全描述；矢量模型在计算精度与数据量方面的优势也是矢量模型比栅格模型受到欢迎的原因之一，采用矢量模型记录数据量大大少于栅格模型的数据量。

栅格模型除了可使大量的空间分析模型得以容易实现之外，还具有以下两个特点：（1）易于与遥感相结合。遥感影像是以象元为单位的栅格模型，可以直接将原始数据或经过处理的影像数据纳入栅格模型的地理信息系统。（2）易于信息共享。目前还没有一种公认的矢量模型地图数据记录格式，而不经压缩编码的栅格格式即整数型数据库阵列则易于为大多数程序设计人员和用户理解和使用，因此以栅格数据为基础进行信息共享的数据交流较为实用。

许多实践证明，栅格模型和矢量模型在表示空间数据上可以是同样有效的，对于一个GIS软件，较为理想的方案是采用两种数据结构，即栅格模型与矢量模型并存，对于提高地理信息系统的空间分辨率、数据压缩率和增强系统分析、输入输出的灵活性十分重要。两种格式的比较见表 2.2-1。

表 2.2-1 矢量格式与栅格格式的比较

数据类型	优点	缺点
矢量数据	1. 数据结构紧凑、冗余度低 2. 有利于网络和检索分析 3. 图形显示质量好、精度高	1. 数据结构复杂 2. 多边形叠加分析比较困难
栅格数据	1. 数据结构简单 2. 便于空间分析和地表模拟 3. 现势性较强	1. 数据量大 2. 投影转换比较复杂