



# 面向时空信息数据的大数据平台设计

耿 晴<sup>1</sup>, 李 兵<sup>1</sup>, 詹 伟<sup>1</sup>

(1. 湖北省测绘成果档案馆, 湖北 武汉 430071)

**摘 要：**设计了面向时空信息数据的大数据平台，以 GISCloud 的云计算框架作为基础资源平台，通过两层服务架构为平台提供计算和存储资源的管理。以数据融合思想构建大数据平台的基础数据支持，依托 GISCloud 建立各种资源池，对数据进行抽取和分析，建立核心数据集与共享数据集，实现数据资源的差异化共享。通过面向时空信息数据的大数据平台设计，构建时空信息数据中心，利用公共服务时空信息和共享交换平台提供数据服务，实现海量时空信息数据的高效管理和服务。

**关键词：**时空信息数据；时空标识；大数据平台；GIS 云

**中图分类号：**P208

**文献标志码：**B

**文章编号：**1672-4623 (2017) 10-0052-03

地理信息系统 (GIS) 是通过对各类地理信息数据进行处理来提供不同的服务<sup>[1-3]</sup>。通过时空信息数据，GIS 可以表达出目标对象的时间和空间属性，在相关领域具有重要的应用价值。随着对时空信息数据的需求越来越深入，时空标识成为地理信息数据的关键属性<sup>[4-6]</sup>。时空信息数据量庞大，传统的数据组织方式不能有效应对。因此，对海量数据的处理和存储成为必须面对的挑战。本文所提出的面向时空信息数据的大数据平台设计，是以大数据为基础的针对性时空信息数据的组织和处理平台。

## 1 地理信息数据的时空标识

传统的基础地理信息数据以二维平面化为主，强调静态地理信息表述，动态信息的表达则受限制。随着时代的发展，卫星导航、国情普查、土地调查、环境监测等地理信息数据应用中，对历史数据进行重建、分析，对未来进行预测的需求日益旺盛。因此，以时空数据模型为基础，通过多种方式对具有时空标识的地理信息数据的组织和应用提供基本方法，提高时空信息数据的利用效率成为重要的研究方向<sup>[7-9]</sup>。当海量的具有时空标识的地理信息数据积聚时，就会产生新问题，即缺少时空信息数据公共基础整合管理平台，各类时空信息专题数据库尚未形成有效的共享，缺少面向公众服务的时空信息服务数据资源等。这就需要新的技术来对时空信息数据的高效利用进行支持，大数据平台设计是解决这一问题的关键。

## 2 时空信息数据与大数据技术

大数据的出现为高效利用具有时空标识的地理信

息数据提供了新方向<sup>[10]</sup>，它具有 3 个明显特征，分别是规模性、多样性和高速性<sup>[11]</sup>。大数据的处理，是在软件和工具的辅助下，对异构的大数据源进行分析、抽取、处理和集成，构建出合理的数据组织形式，提供不同类型的应用，实现定制化、个性化的数据服务。

时空信息数据的数据源丰富，大量具有时空特征的数据积累，速度非常惊人<sup>[12]</sup>。时空信息数据具有大数据的全部特征，是一种面向特定领域的大数据集合。这就需要以大数据技术为支撑，来进行基础数据平台设计。在既有工作基础上<sup>[13-15]</sup>，提出面向时空信息数据的大数据平台设计，通过构建时空信息数据的大数据平台 (GISBDPlatform, GIS Big Data Platform)，利用大数据技术进行数据融合，为时空信息数据的高效使用提供基础的物理载体，从而实现时空信息数据的整合与共享。

## 3 GISBDPlatform 大数据平台设计

### 3.1 基础云平台设计

云计算通过将大量的服务器构建为具有统一逻辑架构的云来提供计算和存储资源，以便进行高效安全的数据存储和计算资源分配。通过对底层硬件资源的抽象化，云计算技术再向上层提供计算和存储资源的支撑。通过云计算所提供的统一管理，底层基础设施的细节被屏蔽，对上层的应用和用户透明，实现资源的高效利用。这正是 GISBDPlatform 基础云平台需要达到的设计目标。

GISBDPlatform 的基础云平台以 GISCloud<sup>[13]</sup> 为基础进行构建，通过 GISCloud 来对核心资源进行整合和集成，构建出一个统一的基础环境，为用户和应用提

供可定制的个性化服务<sup>[14,15]</sup>。GISCloud 的架构如图 1 所示。

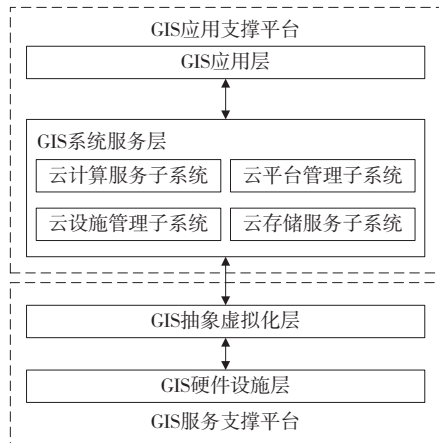


图 1 GISBDPlatform 基础云架构

GISCloud 包括 GIS 服务支撑平台和 GIS 应用支撑平台两个部分。GIS 服务支撑平台构建出 GISCloud 云端。底层具有大量计算与存储资源，包括用于提供计算能力的服务器和用于提供存储能力的存储服务器等。底层资源以异构形式存在，通过 GIS 抽象虚拟化将大量服务器所提供的计算资源和存储资源从逻辑上组织为具有高可用性、高可靠性和高扩展性的基础资源平台。GIS 服务支撑平台实现了对底层硬件资源细节的隔离，同时也有利于提高基础设施的扩展性。

GIS 应用支撑平台则以 GIS 服务支撑平台为基础，提供 GIS 系统服务。这些 GIS 系统服务既为设施和平台的管理提供支持，也为上层的用户和应用提供应用服务接口。GIS 应用支撑平台提供了两个层次的服务。其一是面向管理的 GIS 系统服务，其二是面向应用的 GIS 应用层。GIS 系统服务层的子系统共同为系统管理和上层应用提供了云设施管理、云存储服务、云计算服务和云平台管理。

### 3.2 GISBDPlatform 架构设计

GISBDPlatform 面对的数据对象是多源异构的全要素地理信息数据，具有共享接入、动态更新、实时接入等不同接入方式。基础数据均需要经过清洗、抽取、融合等方式来进行处理，再以用户和应用需要为中心进行数据的按需提取，对上层提供数据应用服务。以此为依据，对 GISBDPlatform 进行分层设计，架构如图 2 所示。从底向上依次是资源层、数据层、服务层和应用层。

资源层的核心是计算资源、存储资源、为计算和存储资源服务的其他基础性硬件设施以及在基础硬件设施之上的系统性软件资源。资源层中的资源构建为 3 个资源池，分别是计算资源池、存储资源池和网络资源池。所有的资源将通过虚拟化的方式整合为统一的

资源平台，向上提供资源服务。而资源的虚拟化，则是通过 GISCloud 的服务支撑平台来完成。GISCloud 的抽象虚拟化层完成所有资源的虚拟化，并将其封装为 GISBDPlatform 的基础资源抽象虚拟化服务。

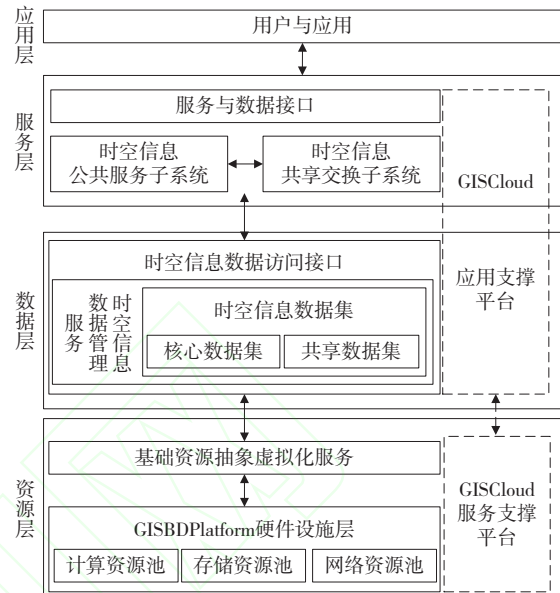


图 2 GISBDPlatform 架构

GISBDPlatform 的核心资源是具备时空标识的地理信息数据集，它包括两个部分的子数据集，分别是核心数据集和共享数据集。时空信息数据集以地理空间基础数据为基础，汇集来自不同来源的各类数据，作为时空信息数据的基础。所有的时空信息数据需要进行数据的融合，主要包括了基础数据的数据整合、行业交换数据的数据融合、非空间数据的数据融合和动态实时传感器数据融合。所有数据通过时空信息数据访问接口向上层提供数据和服务。

服务层包含 2 个重要的子系统，分别是时空信息公共服务子系统和时空信息共享交换子系统，用于支撑与业务逻辑和应用请求相关的时空信息数据的抽取、整理、分析、管理、共享、交换、集成、定制等具体操作。时空信息公共服务子系统是以数据层的共享数据集为支撑，面向公众提供各种时空信息数据和服务。时空信息共享交换子系统则是与相关行业和平台进行数据共享和交换的子系统，实现以地理信息数据为中心、各关联数据融合的服务。该层具有数据与服务接口，将具体与业务和应用有关的数据和服务封装起来，通过统一的接口向上层提供。

应用层则是直接面向用户和公众的层次。在该层利用 GISBDPlatform 所提供的数据和服务接口，向用户提供友好的地理信息服务。数据在 GISBDPlatform 中的流转由其系统管理等部分来提供和控制。而涉及到云相关的部分，则由 GISCloud 中的 GIS 系统服务层来进

行具体的管理。

### 3.3 分层接口设计

在 GISBDPlatform 架构中, 各层之间逻辑独立, 功能清晰, 通过高效的接口交互数据。接口设计采用 CDMI (Cloud Data Management Interface) 协议。CDMI 接口定义了云平台上对数据的访问及管理方法, 如数据的产生、获取、更新、删除等操作, 突破了各层之间异构访问的局限性。在 GISBDPlatform 架构中, 各层通过 CDMI 协议去提供数据、接收数据, 并且各类元数据直接通过 CDMI 存储在数据容器 DataContainer 中, 管理员可通过 CDMI 管理各类数据容器、账户以及访问权限。

CDMI 采用基于对象的管理方式, 主要分为数据对象、容器对象、域对象、队列对象、功能对象。数据对象用来存储数值, 提供文件管理服务。容器对象提供目录服务, 不直接存储数值。域对象用作用户(组)认证手段。功能对象描述 CDMI 服务所提供的功能, 以及能被用户直接调用的功能列表。CDMI 使用不同类型的元数据, 有 HTTP 元数据、数据系统元数据、用户元数据以及存储系统元数据。不同类型的元数据使用不同的协议, 如 HTTP 元数据使用 HTTP 协议。

CDMI 遵从 ISO 各类规范, 灵活高效的数据管理方式、安全的认证机制使得其在云存储平台上获得了广泛的认可和使用。GISBDPlatform 采用 CDMI 接口不仅使各层之间数据交互更为高效和安全, 还兼容了平台在发展过程中的升级操作。

应用 GISBDPlatform 大数据平台进行湖北省时空信息数据中心总体架构的搭建, 如图 3 所示。

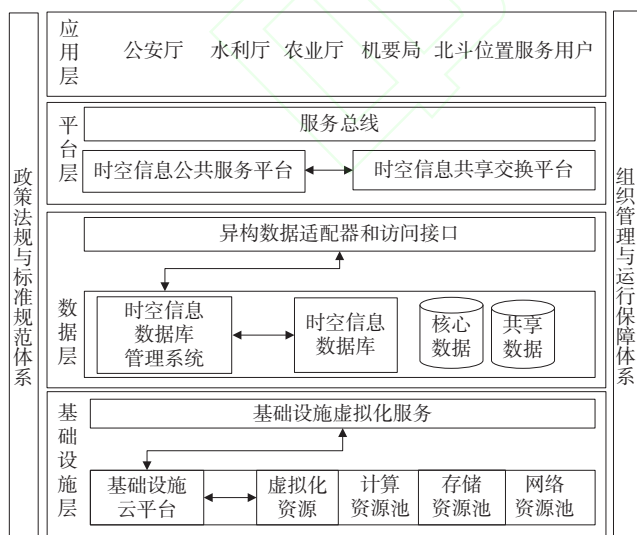


图 3 湖北省时空信息数据中心总体架构图

框架体系的最底层是基础设施层, 逐渐向上展开。政策法规与标准规范体系、组织管理与运行保障体系贯穿于框架体系的各个层次。

1) 基础设施层: 基础设施层的核心内容主要包括服务器集群、网络设施、存储设施、机房环境设施等硬件资源和操作系统、数据库、中间件、GIS 平台等软件资源, 由一定规模的同构或异构基础设施资源组成。基础设施通过基础设施云平台的虚拟化技术将解决方案资源池化, 建立统一的服务器资源池、数据存储池和网络资源池, 通过虚拟化技术将基础设施层的各种异构资源进行整合, 形成一个整体, 对外提供计算、存储、网络等基础设施的通用服务接口。

2) 数据层: 数据层的核心资源是时空信息数据库, 时空信息数据库着眼于满足按需定制“一张图”需求, 依托地理空间基础数据, 逐级汇集各个部门现有的行业专题、公众服务数据等, 逐步实现覆盖全省范围的、统一、唯一、权威的一体化数据库, 为时空信息数据关联奠定数据基础。时空数据库的数据资源预处理、入库、调用、更新等一系列流程化数据管理工作统一由时空信息数据库管理系统实现。

3) 平台层: 平台层包含时空信息数据共享交换平台、时空信息公共服务平台, 用于支撑各类时空信息资源的整合、集成、管理、交换、共享以及应用服务。

4) 应用层: 基于北斗时空基准, 为公安厅、水利厅、农业厅、机要局、北斗位置服务用户建立相应的应用服务工程。

## 4 结 语

本文以大数据技术为导向, 提出面向时空信息数据的平台设计方案 (GISBDPlatform)。GISBDPlatform 以 GISCloud 作为基础的云平台设施, 以资源层、数据层、服务层和应用层贯穿, 采用 CDMI 标准数据交互接口, 形成一个完整的大数据平台结构。大数据平台设计 (GISBDPlatform) 在湖北省时空信息数据中心中的应用, 为湖北省时空信息大数据的流转、处理、融合和共享等提供了新的设计思路, 推动时空信息数据服务的提升。

## 参考文献

- [1] 陈康, 郑纬民. 云计算: 系统实例与研究现状 [J]. 软件学报, 2009, 20(5): 1337-1348
- [2] 姜晓轶, 周云轩. 从空间到时间: 时空数据模型研究 [J]. 吉林大学学报 (地球科学版), 2006, 36(3): 480-485
- [3] 余晓敏, 李兵, 詹庆明, 等. 基于地理国情时空数据的城市建成区动态监测 [J]. 地理空间信息, 2015, 13(6): 8-10
- [4] Peuquet D J. Making Space for Time: Issues in Space-Time Data Representation [J]. GeoInformatica, 2001, 5(1): 11-32
- [5] Dykes J A, Mountain D M. Seeking Structure in Records of Spatio-temporal Behaviour: Visualization Issues, Efforts and Applications [J]. 2003, 43(4): 581-603

(下转第 56 页)



雅，符合游客消费水平，交通情况最佳。南湖至周边宾馆的权重分配表如表 3、4。其中人均消费占总权重比例的 30%，环境及食品评价占总权重比例的 30%，交通情况占总权重比例的 40%，根据这一原则得到总权重表如表 4。

表 3 南湖周边的宾馆权重表  
(其中交通状况主要考虑的是到达时间长短)

饭店名称	人均消费	环境及评价	南湖至此处的交通情况
南湖大酒店	1	1	8
如家快捷宾馆	7	8	7
金桔酒店	5	3	3
7 天快捷酒店	3	5	5
暖暖主题宾馆	9	9	2

表 4 南湖周边的宾馆总权重表

饭店名称	总权重
南湖大酒店	3.8
如家快捷宾馆	7.3
金桔酒店	3.6
7 天快捷酒店	4.4
暖暖主题宾馆	6.2

根据上表得出，从南湖至周边宾馆的最短线路是南湖至如家快捷宾馆。此线路符合游客的综合要求，符合游客消费水平，交通情况最佳。

2 结 语

传统的旅游线路设计是从旅行社的角度或者行为学的角度出发，运用一些数学模型，通过分析景点之间的关系，总结得出旅游线路的具体安排。传统的旅

游线路设计比较注重线路主体，通常会忽略旅游活动中游客的住宿及饮食情况<sup>[1]</sup>。发挥 GIS 技术的优势，使用 GIS 技术设计路线，能够在复杂的道路网络中，通过进行点到点分析，从而制定出最优的景点至饭店及宾馆的路线，避免了人为设计的主观性，增大了实际操作的可行性。另外，这样通过科学分析得到的旅游线路不仅在时间安排方面具有准确性，而且在旅游活动的策划上也比较丰富、舒适，旅游活动主题突出<sup>[7]</sup>。

参考文献

[1] 邹时林, 阮见, 刘波, 等. 最短路径算法在旅游线路规划中的应用: 以庐山为例 [J]. 测绘科学, 2008(5):190-192

[2] 苏迎春, 周廷刚. 地理信息学的形成与发展 [J]. 安徽农业科学, 2008(34):15 269-15 271

[3] 卫伟, 陈存根, 彭鸿, 等. 旅游信息管理现状及其发展趋势 [J]. 资源开发与市场, 2003(5):341-342

[4] 芦佳硕. 考虑路径转移系数的电网黑启动分区恢复策略的研究 [D]. 北京: 华北电力大学, 2013

[5] 穆艳玲. Dijkstra 算法在路由选择中的应用 [J]. 电子制作, 2013(11):281

[6] 徐光著. 动态场景拼接技术的研究与实现 [D]. 重庆: 重庆大学, 2011

[7] 付晶, 郑中霖, 高峻. GIS 技术在旅游线路设计中的应用 [J]. 上海师范大学学报 (自然科学版), 2006(3):92-97

第一作者简介: 吴风华, 副教授, 主要从事地理信息系统的研究和教学工作。

(上接第 54 页)

[6] Yu Hongbo, Shaw Shih Lung. Exploring Potential Human Activities in Physical and Virtual Spaces: ASpatio - Temporal GIS Approach[J]. International Journal of Geographical Information Science, 2008, 22(4): 409-430

[7] Strötgen J, Gertz M. Proximity-Aware Ranking for Textual, Temporal, and Geographic Queries[C]. 22nd ACM International Conference on Information & Knowledge Management (CIKM '13), ACM, New York, 2013

[8] Gey F C, Kando N, Larson R R. The Crucial Role of Semantic Discovery and Markup in Geo-temporal Search[C]. 3rd Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR '10), ACM, New York, 2010

[9] Larson R R. Geographic IR and Visualization in Time and Space[C]. 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR

'08), ACM, New York, 2008

[10] 吴芬芳, 熊卿. 大数据时代地理信息获取与服务方式的嬗变 [J]. 测绘工程, 2015, 24(10): 14-18

[11] 孟晓峰, 慈祥. 大数据管理: 概念、技术与挑战 [J]. 计算机研究与发展, 2013, 50(1):146-169

[12] 马振刚, 李黎黎, 许学工. 自然地理学的大数据研究 [J]. 地理与地理信息科学, 2015, 31(3): 54-58

[13] 赵薇, 耿晴. 云计算在 GIS 系统模型中的应用 [J]. 地理空间信息, 2010, 8(6):8-11

[14] 耿晴, 范伟, 孙续锦, 等. 面向智能移动设备的 GIS 服务 [J]. 地理空间信息, 2012, 10(3): 64-66

[15] 耿晴, 邢光成, 陈涛, 等. 湖北省测绘成果分发服务模式的初步建设 [J]. 地理空间信息, 2013(增刊):46-48

第一作者简介: 耿晴, 硕士, 高级工程师, 主要从事测绘地理信息技术、测绘资料档案管理、信息化服务等相关工作。