

云 GIS 及其关键技术^①

彭义春^{1,2,3}, 王云鹏²

¹(东莞理工学院城市学院 计算机与信息科学系, 东莞 523106)

²(中国科学院 广州地球化学研究所, 广州 510640)

³(中国科学院 研究生院, 北京 100049)

摘要: GIS 与云计算的结合为 GIS 的数据存储、管理、处理及其应用提供了一个新的发展前景. 分别给出了云计算和云 GIS 的定义, 提出了云 GIS 的四种应用模式和一个多层次架构模型, 重点研究了云 GIS 的关键技术: 虚拟化技术、分布式数据存储技术(GFS)、虚拟化数据管理技术(Bigtable)、并行空间分析技术、数据和功能互操作技术和部署模式, 最后给出了一个云 GIS 实际应用框架.

关键词: 云计算; 云 GIS; 数据存储; 数据管理; 并行编程模型

Cloud GIS and its Key Techniques

PENG Yi-Chun^{1,2,3}, WANG Yun-Peng²

¹(City College of Dongguan University of Technology, Dongguan 523106, China)

²(Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, China)

³(Graduate University of the Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The combination of GIS and clouding computing provides an extensive development prospect for GIS information storage, management, processing and application. The definition of cloud computing and cloud GIS are presented, four application mode and a multi-level architecture model of cloud GIS are proposed. The key techniques, such as virtual technology, distributed data storage technology (Google File System), virtual data management technology (BigTable), parallel spatial analysis technology, data and functional interoperability technology as well as deployment model are mainly studied. Finally, an actual application framework of cloud GIS is given.

Key words: cloud computing; cloud GIS; data storage; data management; programming model

1 引言

GIS 即地理信息系统, 是用于采集、存储、管理、处理、检索、分析和表达地理空间数据的计算机系统, 是一种分析和处理海量地理数据的技术系统. 一个完整的 GIS 是由计算机软硬件、空间数据和用户组成, 其中, 空间数据是 GIS 的操作对象也是其核心, 而 GIS 的技术优势在于它的空间数据处理和分析能力.

当前, GIS 具有以下特点: 1)空间数据生产单位相对较少, 而数据使用者众多且多样化; 2)基础数据多, 数据量庞大, 更新频度低; 3)并发用户数很大, 但每次使用量较小; 4)需要海量数据存储, 进行数据处理和数据挖掘. 这些特点已不能完全满足当前海量的并发用户对海量数据的数据分析、处理的需求, 同时, 现有的

数据管理和处理的体系结构、方法和技术也不能完全满足上述要求. 云计算的出现给 GIS 新的发展带来契机, 云计算非常适合当今 GIS 的需求. 首先, 在地理数据存储方面, 云计算的高可靠、高吞吐和可伸缩的新型地理数据存储技术可为 GIS 提供可靠、安全的数据存储中心, 也可满足海量地理数据在高并发下的地理信息服务要求; 其次, 在地理数据计算方面, 云计算基于虚拟集群这一新型云计算设施的可伸缩处理数据计算技术, 可满足海量地理数据高性能计算; 最后, 在地理信息服务方面, 云计算具有松耦合, 数据、功能、服务的全共享, 服务可聚合、可迁移等特点, 可解决传统 GIS 缺乏统一的平台、地理信息资源难以共享以及“信息孤岛”等问题, 同时, 重复投入和资源浪费

^①收稿时间:2013-12-25;收到修改稿时间:2014-02-17

也将大大减少并可快速搭建和部署 GIS 应用。

2 云计算概述

中国电子学会云计算专家委员会给出的云计算的定义^[1]: 云计算就是一种基于互联网的、大众参与的计算模式, 其计算资源(包括计算能力、存储能力、交互能力等)是动态、可伸缩、被虚拟化的, 以服务的方式提供. 提供资源的网络被称为“云”, “云”中的资源在使用者看来是可以无限扩展的, 并且可以随时获取, 按需使用, 随时扩展, 按使用付费.

加利福尼亚大学圣塔芭芭拉分校的博士生 Lamia Youseff 和纽约 IBM T.J. Watson 研究中心的研究员 Maria Butrico、Dilma Da Silva 在 2008 年发表了一份题为《Toward a Unified Ontology of Cloud Computing》的研究报告, 该报告建立了一个五层模型^[2], 主要包括以下服务: SaaS(Software-as-a-Service: 软件即服务)、PaaS(Platform-as-a-Service: 平台即服务)、IaaS(Infrastructure-as-a-Service: 基础架构即服务)、DaaS(Data-as-a-Service: 数据即服务)和 HaaS(Hardware-as-a-Service: 硬件即服务). 其中, IaaS 是云计算的基础; DaaS 基于 IaaS 构建; SaaS 可基于 PaaS 部署或直接部署于 IaaS 之上, 以便高效、便捷的对外提供服务; PaaS 可基于 IaaS 构建或直接构建于物理资源之上, 可直接向用户提供开发和应用程序部署服务. 将 SaaS、PaaS、IaaS 结合 DaaS 就可以构建一个完整的云计算环境. 除了这几种常见模式之外, 云计算又有很多新的模式: NaaS(Network as a Service: 网络即服务)、STaaS(Storage as a Service: 存储即服务)、SECaaS(Security as a Service: 安全即服务)、DBaaS(Database as a Service: 数据库即服务)、TEaaS(Test Environment as a Service: 测试环境即服务)、APIaaS(API as a Service: API 即服务)、BaaS(Backend as a Service: 后端即服务, 为应用开发提供后台的云服务)等.

3 云GIS应用模式和架构模型

所谓云 GIS^[3], 就是将云计算的各种特征用于支撑地理空间信息的各要素, 包括建模、存储、处理等等, 从而改变用户传统的 GIS 应用方法和建设模式, 以一种更加友好的方式, 高效率、低成本的使用地理信息资源.

3.1 云 GIS 的四种应用模式

当前, 云 GIS 还处于发展阶段, 云计算的四种服务模式在云 GIS 都有体现. “云”的规模可按行业、规模、流程、业务和不同层次的用户可以动态伸缩; “云”之间也可根据不同需要、根据不同业务属性进行聚合; 服务提供商可重新定制其应用, 为各种用户提供无限多的千变万化的应用. 根据云计算的四种服务模式, 云 GIS 也提出了四种 GIS 应用模式, 即: 地理信息软件即服务(也即 SaaS)、地理信息平台即服务(也即 PaaS)、地理信息基础设施即服务(也即 IaaS)、地理信息内容即服务(也即 DaaS). 云 GIS 的四种应用模式如图 1 所示.

3.2 云 GIS 的多层架构模型

内容即服务	面向最终用户的各类数据、地图, 以服务的形式提供, 用户按需访问, 无需自己建立和维护数据。
软件即服务	面向最终用户的应用, 以服务的形式提供, 而不是部署在本地的软件。
平台即服务	应用平台或中间件, 以服务的形式提供, 开发人员可基于此类服务构建和部署定制的应用系统。
基础设施即服务	网络、计算、存储等一体化基础设施, 以服务的形式提供而不是专用的能力。

图 1 云 GIS 的应用模式

按照云计算平台的架构模型, 云 GIS 体系结构也采用了一个多层次模型如图 2 所示. 自下向上分为物理层、虚拟层、数据资源层、云计算支持平台及服务组件层、服务层、应用层以及涉及多个层的服务发现、服务监控、资源调度、计量(计费)统计等多种服务.

应用层	桌面、平板、移动终端等客户端、服务管理门户 业务流程(可选)	服务发现
服务层	企业级、特点业务、特定项目等组件及其接口 GIS 服务器、Web 服务器、监控服务、注册服务	
云计算支持平台及服务组件层	GIS 运行环境、HPC 平台、Dryad/DryadLINQ 平台 GRID 计算、并行计算、分布式计算、消息中间件、 服务组件、通用开发工具	服务监控
数据资源层	分布式存储系统、空间数据引擎、数据访问接口、 服务包装 数据库服务器(空间数据库、属性数据库、 文件型数据库)	
虚拟层	虚拟机操作系统 虚拟机	虚拟
物理层	物理机操作系统 物理硬件设施(机房/计算机/网络...)	

图 2 云 GIS 平台体系结构

4 云GIS的关键技术

4.1 虚拟化技术

虚拟化技术以前在计算机体系结构、操作系统、编译器和编程语言等领域得到了广泛应用.虚拟化技术实现了资源的逻辑抽象和统一表示,它从逻辑上对资源进行重新组织,从而使资源能实现共享.虚拟化技术在服务器、网络及存储管理等方面发挥了巨大的作用,大大降低了管理复杂度,提高了资源利用率,提高了运营效率,从而有效地控制了成本,同时,在大规模数据中心管理、基于 Internet 等解决方案交付运营方面也有着巨大的价值,不过,虚拟化技术虽然可以有效地简化数据中心管理,但仍然不能消除企业为了使用 IT 系统而进行的数据中心构建、硬件采购、软件安装、系统维护等环节,与云计算的结合使得两者相得益彰.云计算借助虚拟化技术的伸缩性和灵活性,提高了资源利用率,简化了资源和服务的管理和维护,而在大规模数据中心管理和解决方案交付方面因云存储使得数据更加安全.

当前,虚拟化已成为云计算的关键技术之一,虚拟化技术包括计算虚拟化、存储虚拟化、网络虚拟化等.在云 GIS 中,虚拟化主要包括资源的虚拟化和应用的虚拟化:

(1)资源虚拟化.为了实现软件应用与底层硬件相隔离,在云 GIS 平台体系结构中专门设计了一层虚拟层,它采用将单个资源划(硬件和软件)分成多个虚拟资源的裂分模式和将多个资源整合成一个虚拟资源的聚合模式.它通过抽象化,可为系统提供多台可用的“虚拟机”,这些虚拟机之间在逻辑上是独立可用的硬件资源集合,当多种操作系统运行在虚拟机上时,资源组织仍然是独立的,这样也就保证了系统运行的可靠性;

(2)应用虚拟化.分布式的 GIS 应用服务器集群是整个云 GIS 建立的核心,主要完成包括人机交互、数据编辑、拓扑关系生成、投影、格式转换、影像处理与信息提取、地图生成和发布等 GIS 专业应用,为了便于在各个服务器之间进行动态切换、相互替代,服务器集群应具有统一的标准.一个服务器实例可以运行一个或者多个应用实例,具体运行的是哪个实例需要通过“服务监控”进行分配,服务器实例也是虚拟的计算资源.另一方面,虚拟化将应用程序与操作界面迁移到用户本地来实现交互和显示操作,而数据的处

理、应用程序的运行则在远端的“云”中来完成.

4.2 分布式数据存储技术

云 GIS 的空间数据具有海量、异构、多源、多尺度和多时空等特点,这导致了系统的空间数据通常是无序、杂乱、动态的,因此云 GIS 数据采用分布式存储方式,通过冗余存储的方式来保证数据的可靠性.云存储技术具有分布式、高吞吐率和高传输率的特点,完全适合空间数据的上述特点.当前,大多数云 GIS 采用基于 BigTable、HBase、NoSQL 等数据库技术来存储和管理空间数据,并通过标准的空间数据库连接技术(SDE)和 REST 接口来统一访问,同时,分布式缓存技术(如 Memcached)的应用也能有效地降低后台服务器的压力和加快响应速度.目前数据存储技术主要有 Google 的非开源的 GFS(Google File System)和 Hadoop 的开源的 HDFS(Hadoop Distributed File System).下面对 GFS 作一个概述.

GFS 是 Google 公司为了存储海量搜索数据而设计的专用文件系统,是一个适用于大规模分布式数据处理、可扩展的分布式文件系统.它对硬件设备要求不高,包括了容错技术且为大量客户端提供极高的聚合处理性能.GFS 的体系结构(GFS 集群)由单个的 master 和多个 chunkserver(块服务器)组成,还有相应的很多客户端 client. GFS 的体系结构如图 3 所示:

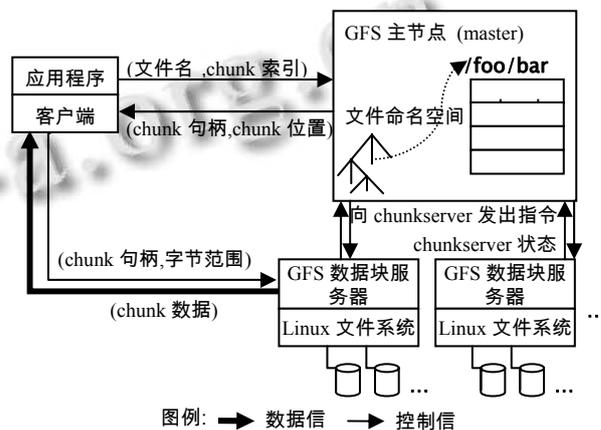


图 3 GFS 的体系结构

(1) chunkserver. 在 GFS 下, 每一个文件都拆成固定大小的 chunk(块). 每一个块都由 master 根据块创建的时间产生一个全局唯一的以后不会改变的 64 位的 chunk handle 标志. chunkservers 在本地磁盘上用 Linux 文件系统保存这些块, 并且根据 chunk handle 和字节

范围, 通过 Linux 文件系统读写这些块的数据. 为了保证数据的可靠性, 每一个块都会在不同的 chunkserver 上备份.

(2)master. 负责管理所有的文件系统的元数据(metadata), 包括文件和 chunk 的命名空间、访问控制信息、文件到 chunk 的映射关系、当前 chunk 的位置等. 它也控制系统范围的活动, 如块租约(lease)管理, 孤儿块的垃圾收集, chunkserver 间的块迁移. master 定期通过 HeartBeat 消息与每一个 chunkserver 通信, 给 chunkserver 传递指令并收集它的状态.

(3)client. client 代码包含了文件系统的 API 并与 master 和 chunkserver 通信以实现应用程序读、写数据. client 与 master 的交换只限于对元数据的操作, 而所有数据通信都直接和 chunkserver 联系, client 和 chunkserver 都不缓存文件数据, 这是由于数据太多或工作集太大而无法缓存, 同时, 因不必考虑缓存的一致性问题的简化了客户程序和整个系统. client 缓存元数据, chunkserver 也不缓存文件, 因为 chunk 是作为本地文件存储的.

Client 读取数据的操作顺序:

(1)client 把应用要读取的文件名和偏移量, 根据固定的 chunk 大小, 转换成为文件的 chunk index.

(2)向 master 发送这个包含了文件名和 chunkindex 的请求.

(3)master 返回相关的 chunk handle 以及对应的位置, client 缓存这些信息.

(4)client 就向最近的对应位置的 chunkserver 发起请求, 请求包含 chunk handle 以及一个在这个 chunk 内需要读取得字节区间.

(5)chunkserver 返回给 client 要读取的 chunk data.

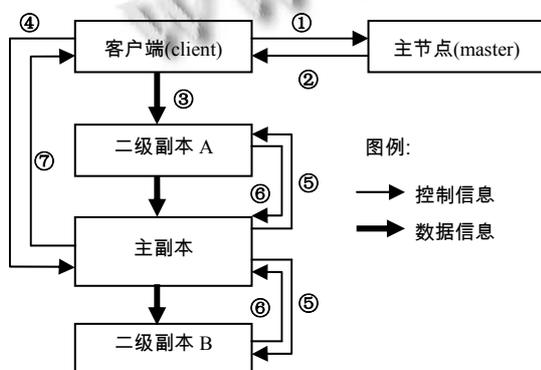


图 4 租约和修改顺序

(1)client 向 master 询问哪一个 chunkserver 保存了当前的租约 lease, 以及其他副本的位置.

(2)master 回复主块的 chunk handle 以及其他副本的位置.

(3)client 把数据推送到所有的副本上.

(4)所有的副本都被确认已经得到数据后, client 发送写请求到主块.

(5)主块把写请求传递到所有的二级副本.

(6)所有二级副本回复主块说明他们已经完成操作.

(7)主块回复 client.

4.3 虚拟化数据管理技术

云 GIS 由大量分布式服务器组成, 要并行地为大量用户提供服务, 要对分布的、海量的空间数据进行处理和分析, 在传统的 GIS 的“RDBMS+SDE”结构和“Extended ORDBMS”的结构中, 空间数据库管理系统则会成为系统性能的瓶颈. 如何提高数据的更新速率以及进一步提高随机访问速度是云 GIS 数据管理技术必须解决的问题.

一方面, 虚拟化技术的空间数据管理, 具有离线应用和在线更新技术, 支持在系统之间、部门之间、层级之间快速迁移和自动同步数据, 实现了分布式、多级别、支持多终端的空间数据保障流程; 另一方面, 云计算系统的平台管理技术, 能将分布式服务器群协同工作, 为业务的部署和开通提供了极大的便利, 也能快速发现系统故障并能及时恢复系统, 同时, 运用自动化、智能化方法保障大规模系统运营的可靠性. 当前主要有 Google 的 BigTable 数据管理技术和 Hadoop 团队开发的开源数据管理模块 HBase. 下面以 BigTable 为例介绍如何实现云计算的大规模数据集管理.

Bigtable 是一个分布式的结构化数据存储系统, 它被设计用来处理海量数据: 通常是分布在数千台普通服务器上的 PB 级的数据. Bigtable 是一个稀疏的、分布式的、持久化存储的多维度排序 Map. Map 的索引是行关键字、列关键字以及时间戳; Map 中的每个 value 都是一个未经解析的 byte 数组. BigTable 管理的数据的存储结构为: `<row: string, column: string, time: int64>->string`. BigTable 中的数据项按照行关键字的字典序排列, 每行动态地划分到记录板中. 每个节点管理大约 100 个记录板. 时间戳是一个 64 位的整数, 表示数据的不同版本.

BigTable 整体结构如图 5 所示, 包括: 主控服务器 (master server), 子表服务器(tablet server)和客户端程序:

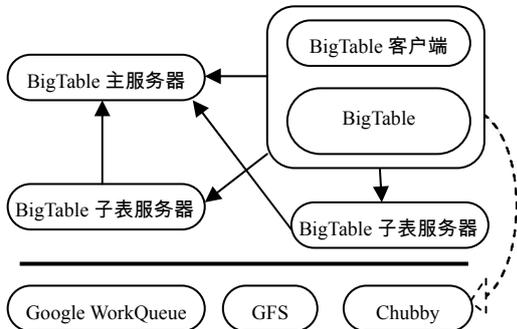


图 5 BigTable 整体结构

主控服务器: 负责整个系统的管理工作、子表的分配、子表服务器的负载均衡、子表服务器失效检测等;

子表服务器: 负责子表的数据存储和管理, 响应客户端程序的读/写请求, 对应所负责的子表以 GFS 的格式存储, 一个子表就是由子表服务器中存储的若干个 SSTable 文件组成的;

客户端: 具体应用的接口程序, 与子表服务器进行通信, 读写子表对应的数据。

BigTable 采用一个三层类似 B+树的结构存储 Tablet 的位置信息, 如图 6 所示. 其中第一级的 Chubby file 中包含 Root Tablet 的位置, Root Tablet 包含所有 METADATA Tablets 的位置信息, 每个 MetaData Tablets 包含许多 User Table 的位置信息.

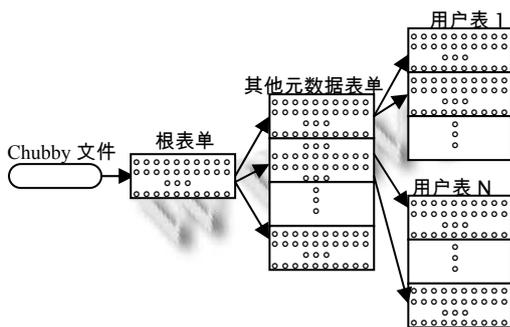


图 6 BigTable 存储结构

对应客户存储的数据, 被分割到对应的表格中, 并且每个表格又被分割为若干子表, 存储在不同的子表服务器中. BigTable 利用 Chubby 系统和元数据表来维护系统管理数据.

Chubby 中某个特殊文件则指出 Root 子表所在的

子表服务器地址;

元数据表: 每一行记载了了整个 BigTable 中某个具体子表存储在那台服务器上管理信息, 而这些元数据表也被分割成若干子表, 并存储在不同的子表服务器中. 第一个元数据子表为 Root 子表, 它记录元数据表自身除 Root 子表外其他子表的位置信息. 通过 Root 子表记录可以找到元数据表中其他子表存储在哪个机器上, 从而得到完整的元数据表, 而在元数据表的每一行, 记录了 BigTable 中应用程序生成的表格的某个子表的管理数据; 每一行以用户表名和子表内存储的最后一个行主键共同构成元数据表内此条记录的行主键; 在记录行的数据里则存储了这个子表对应的子表服务器等其他管理信息.

某个客户端需要查找用户表的某行记录, 先读取 Chubby 系统中的特殊文件, 从中得知 Root 子表的所在位置, 然后根据 Root 子表获取元数据表其他子表的所在位置, 通过和要查询的用户表以及待查询记录的行主键比较, 知道具体那台子表服务器存储着这条记录. 然后客户端将这些信息缓存在本地, 并直接和子表服务器通信来读取对应的数据.

4.4 并行空间分析技术

云 GIS 的数据处理能力必须: 具有面向任务的异步空间数据处理架构, 支持大型集群的并发处理和流程控制, 支持长时间运行、长事务处理, 支持移动终端操作处理大型空间数据库, 具有处理流程的可视化设计和运行状态的实时监控功能, 可以跨平台、跨地域整合空间数据的处理流程, 并能够实时将处理结果进行发布, 等等. 另外, 云 GIS 的空间分析能力必须: 具有统一的空间分析框架, 丰富的、标准化的空间分析模型库, 支持空间分析流程的快速构建和自动化运行, 并能够实时将分析结果进行发布. 为此, 我们必须充分利用云平台的计算能力, 尽可能地使用并行算法, 以提高云 GIS 的空间数据处理和空间分析能力.

Google 的 MapReduce 是一种并行编程模型和分布式计算框架, 它把要执行的问题分解成 Map(映射)和 Reduce(归约)两个阶段来完成, 首先是将每一个任务拆分成若干个 Map 任务并分配到不同的计算机上去执行, 每一个 Map 任务将输入文件的一部分作为自己的输入, 经过计算生成与最终所需的文件格式完全一致但仅包含部分数据的中间文件, 当所有 Map 任务完成后, 系统会生成若干个 Reduce 任务并分配到不同的计

算机上去执行, Reduce 任务的目的是将若干个 Map 任务生成的中间文件为汇总到最后的输出文件中, 最终, 生成所需的目标文件. 整个算法的关键, 就在于增加了一个中间文件生成的流程, 大大提高了灵活性, 使其分布式扩展性得到了保证.

图 7 为 MapReduce 并行处理空间选择查询的数据流程图.

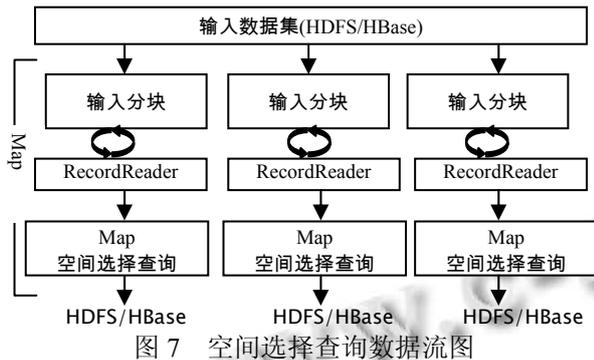


图 7 空间选择查询数据流程图

以 HDFS 中输入文件分块或 HBase 表分块作为 Map 任务划分依据, 使用空间选择查询算子(点查询算子或区域查询算子)作为 Map 函数, k_1 为空间对象在文件中的行号或在 HBase 中的 ID, v_1 为该对象的所有属性值. 各 Map 任务分别对不同的文件块或表分块中的对象执行空间选择查询, Map 任务的计算结果直接放入 HBase 或 HDFS 作为最终结果. 由于不需要对 Reduce 任务的结果进行归约运算, 因此, 空间选择查询处理作业不需要 Reduce 阶段.

下面再以栅格叠加分析来并行空间分析算法的设计.

某一记录的输入格式为 $F\langle MBR, Value \rangle$, 其中, MBR 是这条记录的外包矩形, Value 是此矩形范围内的某个栅格点的属性值的数组. 使用两次 Map 和 Reduce 阶段的算法来完成叠加分析操作.

(1)在第一次 Map 阶段, 按照某一空间划分方法, 根据每条记录的外包矩形将其划分成若干长和宽相同的新记录. 新记录的键是它的外包矩形与属性类别的组合, 若属性值数组中与原记录的坐标重叠则使用原来的值, 新增处则没有值, 暂用一个特殊值来标记.

(2)在第一次 Reduce 阶段, 按照新记录的键进行合并, 在同一个外包矩形内的记录取有值的部分合并.

经过上述 MapReduce 阶段结束后, 中间结果被保存起来, 原来每一个的数据文件被划分成了规则的格

网.

(3)在第二次 Map 阶段, 选择每一个记录的外包矩形作为键, 而将属性类别和值数组作为值.

(4)在第二次 Reduce 阶段, 选择具有相同键的记录按照设定的叠加算法进行处理, 处理后的结果以文件形式保存.

(5)将此文件按照输出要求进行归约, 栅格叠加分析操作就完成了.

4.5 数据和功能互操作技术

云 GIS 必须实现跨操作系统 (Linux/Unix/AIX/Windows)、跨 GIS 平台、数据源(格式)异构、硬件异构、环境异构, 同时, 云 GIS 必须支持单点发布、自动同步、频度统计和自动优化, 支持云内部的数据互操作、私有公有云的互操作和云中心之间的互操作. 为了实现上述功能和操作可从以下几方面去实现:

(1)制定相关的数据政策和标准

根据云 GIS 的六层体系结构, 就要包括实现上述六个层次的标准化, 具体涉及到云计算互操作和集成标准、云计算的服务接口标准和应用程序开发标准、云计算不同层面之间的接口标准、云计算服务目录管理、不同云之间无缝迁移的可移植性标准、云计算商业指标标准、云计算架构治理标准、云计算安全和隐私标准等一系列接口、规范和标准. 目前, 眼下三大国际标准化组织 (ISO、IEC、ITU) 纷纷展开云标准工作, 国内外数十个标准组织也已启动云计算相关标准体系的研究. 这些组织可以分为三类: 一是传统信息技术标准化组织: 包括 ISO/IEC、美国国家标准技术研究所 (NIST)、分布式管理任务组 (DMTF)、网络存储工业协会 (SNIA)、互联网工程任务组 (IETF)、开放移动联盟 (OMA) 等. 二是针对云计算新成立的组织: 包括开放式数据中心联盟 (ODCA)、筋斗云 (Open Cirrus)、云计算安全联盟 (CSA)、Hadoop 社区、开放网格论坛 (OGF)、开放云计算联盟 (OCC)、绿色网格 (TGG) 等. 三是传统电信领域的标准化组织: 包括国际电信联盟 (ITU)、世界移动通信大会 (GSMA)、欧洲电信标准研究所 (ETSI)、中国通信标准化协会 (CCSA) 等.

(2)基于 Web Service 技术构建

云 GIS 是 GIS 与应用模型将以服务模式向用户提供应用, 可通过 Web Service 将云 GIS 各类信息资源包装成统一的形式, 采用 XML/GML 等标准格式来进行

空间信息传输和存储的,实现分布式、异构环境下空间信息标准化、同构化,从而为进一步实现分布式异构网络环境下的 GIS 集成和互操作奠定了基础;构建与 OpenGIS 的空间信息服务体系相对应的包括模型交互、模型管理、 workflow 及任务管理、模型处理、模型通讯、模型系统管理等六大类服务的应用模型服务体系.通过对分布式、异构环境下信息资源的统一管理和调度,云 GIS 真正实现了跨平台、跨系统、跨硬件设施的异构整合. ArcGIS 10.1 for Server 就是一款纯 Web Service 产品.

(3)采用 Web Server 集群技术

采用 Web Server 集群技术,可构建能够整合跨区域、跨平台、跨部门的 GIS 服务器,实现多资源的应用整合,完成企业各类信息的互通互联.在 ESRI 的 ArcGIS 10.1 for Server site 中必须至少有一个集群,当第一次创建 ArcGIS 10.1 for Server site 的时候,会产生一个名为“default”的默认集群.以后创建的 Runnable GIS Server 节点都可以添加到这个集群内,某个 Site 站点也可以创建多个集群.这种集群架构保证了系统的健壮性和扩展性,即使是某一个 GIS Server 节点意外的宕掉,也不会导致整个地图服务的停止运行;同样,当需要增加一个 GIS Server 节点时,以 plug-in 方式插入一个节点为服务提高负载能力.同时,在逻辑上,在多个 GIS 服务器中我们可以根据服务器的性能,或者根据我们应用的不同而进行分组,不同的组用于处理不同的服务,比如说性能比较好的机器用于处理 GP 服务,性能一般的用于处理地图服务,进行弹性的服务部署.

4.6 部署模式

云 GIS 的建设模式与云计算相同,主要有三种的建设模式:共有云 GIS、私有云 GIS 和混合云 GIS.其中,混合云 GIS 是共有云 GIS 和私有云 GIS 之间的权衡模式.

共有云 GIS(Public Cloud GIS),由专业的云 GIS 供应商负责提供各种类型的 GIS 资源服务,用户无需关心云端所有资源的安全、管理、部署和维护,也无需任何前期投入而只要按需获取并使用即可.共有云 GIS 具有前期投入少,周期短;数据更新维护简单;系统维护简单;自动在线升级等优点.图 8 为共有云 GIS 部署模式.目前,发展成熟的共有云 GIS,有 ESRI 提供的共有云 GIS 产品,包括 ArcGIS.COM、ArcGIS

APPs/APIs 以及 ArcGIS In Amazon,国内的有 SuperMap 与微软 Windows Azure 平台开展合作的产品等等.

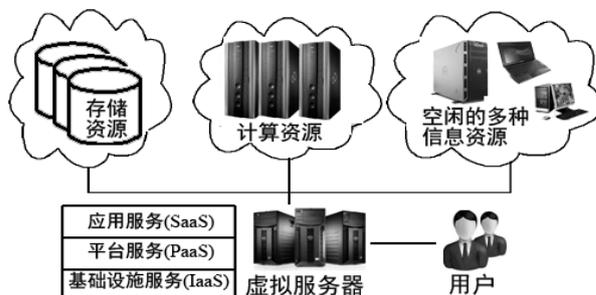


图 8 共有云 GIS 部署模式

私有云 GIS(Private Cloud GIS):私有云是为一个客户单独使用而构建的,因而提供对数据、安全性和服务质量的最有效控制.对于企业应用来说,由于一些其他的原因,例如网络条件的限制、对数据安全的更高要求、有能力建设私有云等等,企业用户同时也需要搭建企业自己的云 GIS 环境,并由企业自己来管理和维护云端的各种资源,即建设企业的私有云 GIS.私有云具有数据安全、网络通达性好、资源可控、易于定制等优点.图 9 为常见的私有云 GIS 结构.当前, ArcGIS Server 可实现“云端发布服务,本地搭建应用”,让用户能够发布具有弹性调整能力的 GIS 服务,并可将其应用于自己的业务应用系统当中,同时,允许用户将开发的应用托管至云端; SuperMap iServer 6R 为用户提供数据、硬件、IaaS 解决方案、GIS 平台软件(PaaS)等支撑服务.



图 9 常见的私有云 GIS 结构

混合云 GIS(Hybrid Cloud GIS):指供自己和用户共同使用的云,它所提供的服务既可以供别人使用,也可以供自己使用.混合云表现为多种云配置的组合,

数个云以某种方式整合在一起。例如有时用户可能需要用一套单独的证书访问多个云,有时数据可能需要在多个云之间流动,或者某个私有云的应用可能需要临时使用共有云的资源。一些单位考虑到安全问题,选择私有云或“混合云”的部署模型,共有云是最通用的云计算,基础设施和应用都由提供云服务的机构维护。当然由于许多的传统提供商和用户目前还不使用共有云计算,因此云服务层只能在私有云环境中复制使用。

5 具体的应用

东莞市政务核心平台基于 GIS 云服务中心现暂时提供政务和公共两方面应用。政务应用主要限于内网、外网和交互之间的网络,而公共服务涉及到个人、企业和团体,需要设施虚拟化、平台虚拟化、平台虚拟化和应用虚拟化。东莞市试点 GIS 云服务中心两个核心内容:应用虚拟化和数据虚拟化。涉及到的部门包括国土、环保、交通等部门。其总体框架如图 10 所示。

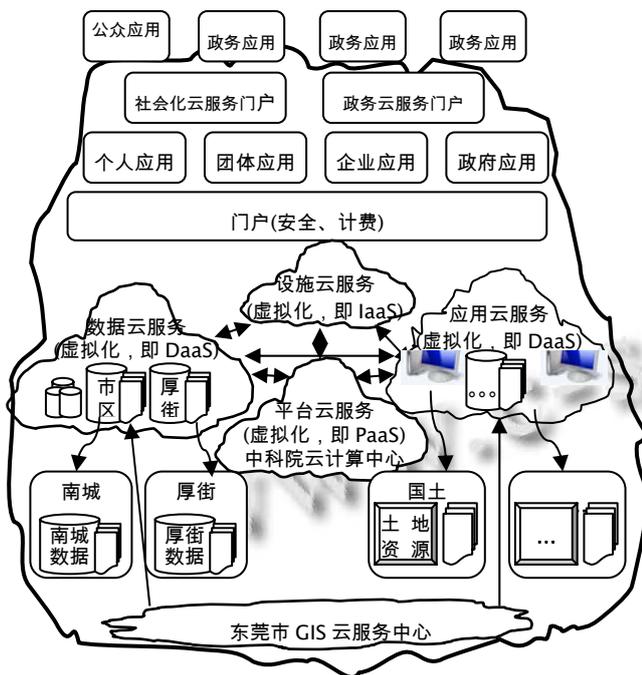


图 10 东莞市政务云核心平台总体框架

6 结语

随着用户量或者使用频率的增加,超负荷并发量

的请求推向后端的 GIS 服务器,此时, GIS 服务处理性能就遇到瓶颈。随着云计算的发展和普及,各种端的应用获得了有力的模式和广阔的前景。云 GIS 因可通过网络服务的方式: 按需提供各种存储和计算能力,从而省去用户大量的运维精力,正在被越来越多的用户所重视和利用,也带给人们良好的用户体验,用户可在云中随时、随地通过任何设备即可获取所需 GIS 资源,并且这种资源是可以计量和灵活扩展的,实现了 GIS 能力从云到端的高效流动。总之,云计算将深刻改变我们应用 GIS 的传统方法和模式, GIS 系统建设和运营模式也都将面临新的机遇和挑战,云 GIS 意味着数据、软件、开发方面的壁垒已被打破,所需 GIS 资源变得唾手可得,信息共享更加方便快捷,这正是许多服务性机构一直在追求的目标。

参考文献

- 1 中国电子学会云计算专家委员会. 云计算技术发展报告. 北京: 科学出版社, 2011.
- 2 Youseff L, Butrico M, Da Silva D. Toward a unified Ontology of cloud computing. Grid Computing Environments Workshop, 2008. GCE'08 12-16 Nov. 2008. 1-10.
- 3 Esri 中国(北京)有限公司. ArcGIS 云计算解决方案——触手可及的云. <http://www.esrichina-bj.cn>
- 4 戴立乾, 陈娜. 浅谈云计算时代下 GIS 的发展. 安徽农业科学, 2009, 37(31): 15556-15557, 15572.
- 5 赵薇, 耿晴. 云计算在 GIS 系统模型中的应用. 地理空间信息, 2010, 8(6): 8-10.
- 6 王尔琪. 云计算与 GIS 技术革新. 新经济导刊, 2011, 10: 83-87.
- 7 王尔琪. 云 GIS 技术路线与发展探析. <http://www.supermap.com.cn/sup/xwtxpage.asp?orderID=416>
- 8 王鹏, 董静宜. 一种云计算架构的实现方法研究. 计算机工程与科学, 2009, 31(A1): 11-13.
- 9 张书彬, 韩冀中, 刘志勇, 等. 基于 MapReduce 实现空间查询的研究. 高技术通讯, 2010, 20(7): 719-726.
- 10 Hofmann P. Cloud computing: The limits of public clouds for business applications. IEEE Internet Computing, November -December 2010, 14(6): 90-93.