

文章编号: 1007-4619(2007)03-0385-05

新型遥感卫星数据目录服务系统的研制

汪超亮^{1, 2, 3}, 李传荣², 贾 斌^{3, 4}

(1. 中国科学院 遥感应用研究所, 北京 100101; 2. 中国科学院 光电研究院, 北京 100080;
3. 中国科学院 研究生院, 北京 100039; 4. 中国科学院 中国遥感卫星地面站, 北京 100086)

摘 要: 随着遥感技术的不断进步, 各种类型的新型传感器不断涌现, 遥感数据源呈现高时间分辨率、高空间分辨率和高光谱分辨率的发展趋势, 数据类型越来越多, 数据量越来越大。对于这些海量遥感数据必须采用有效的存储、管理和检索。首先分析了遥感编目数据的特点, 在对构建高性能、高可靠和可维护的遥感卫星数据目录服务系统的关键技术进行研究的基础上, 实现了新型的目录服务系统, 并描述了该系统实现的功能。

关键词: 编目存档; 编目浏览; 海量数据存储

中图分类号: TP391 **文献标识码:** A

Development of Satellite Data Catalogue and Archive System

WANG Chao-liang^{1, 2, 3}, LI Chuan-rong², JIA Bin^{3, 4}

(1. Institute of Remote Sensing Applications Chinese Academy of Sciences Beijing 100101, China;
2. Academy of Opto-Electronics Chinese Academy of Sciences Beijing 100080, China;
3. Graduate School of Chinese Academy of Sciences Beijing 100039, China;
4. Remote Sensing Satellite Ground Station Chinese Academy of Sciences Beijing 100086, China)

Abstract: With the development of the remote sensing technology, many new types of remote sensing sensor are being developed immediately and used for earth observation. Large numbers of earth images, which are very important to research earth, are obtained. As remote sensing satellite data has a tendency to become high temporal, spatial and spectral resolution, its volume is increasing sharply. It is a great problem on how to efficiently store, manage and search mass of remote sensing satellite data. And satellite data catalogue and archive system, a mass storage and searching web-based system, is just the tool to solve the problem. In order to develop high efficient, high reliable and maintainable archive and catalogue system, the characteristics of satellite data's metadata are analyzed in the paper firstly, which determine the structure of storage organization. Then some critical technology on developing new type of archive and catalogue system are provided, including mass storage, spatial data storage, the implementation of high reliable system, archive and catalogue database's optimization. At last, archive and catalogue system's hardware structure is shown and its main functions are introduced briefly and an example of searching RESOURCESAT-1 satellite data is given.

Key words: catalogue and archive system; catalogue and browse; mass storage

1 引 言

卫星遥感集中了空间、电子、光学、计算机通信和地学等学科的最新技术成果, 为人类研究地球获取了大量珍贵的图像资料。随着遥感技术的不断进

步, 各种类型的新型传感器不断涌现, 遥感数据源呈现高时间分辨率、高空间分辨率和高光谱分辨率的发展趋势, 数据量越来越大, 其在空间探索、全球变化、地表资源环境监测、农作物估产、灾害监测、城市规划等领域的应用越来越广泛, 为国民经济的发展创造着巨大的经济效益。

收稿日期: 2006-09-01; 修订日期: 2006-10-30

基金项目: 中国科学院知识创新工程方向性项目(KGCX3-SW-401)资助。Publishing House. All rights reserved. <http://www.cnki.net>

作者简介: 汪超亮 (1980—), 男, 中国科学院遥感应用研究所博士研究生, 目前主要从事遥感卫星数据预处理、海量数据存储研究。

E-mail: cfwang@aoe.ac.cn

对于海量的遥感卫星数据必须考虑采用有效的手段进行组织、存储和管理,并在此基础上能够充分、有效地实现数据共享和数据发布。遥感卫星数据目录服务系统使用户能够快速、方便、准确地获取感兴趣的遥感数据,实现遥感数据向遥感信息的转化,为遥感应用的实用化、产业化发展提供有效的工具。由于遥感卫星数据量庞大,数据结构复杂,要从上千万景存档图像中快速、准确检索到人们所感兴趣的数据,就对目录服务系统提出了巨大的挑战。同时目录服务系统是卫星数据编目浏览的信息支撑中心,它不仅决定了编目查询的性能,同时也是整个系统不间断运行的前提。所以,通过对目录服务系统的研究,分析当前运行系统的优缺点,综合考虑各种因素,在可获得的支撑环境下设计出更先进、更合理的目录服务系统是当前卫星遥感数据应用发展的一项基础工作与重要研究课题。

遥感数据的编目就是将遥感数据依照一定的规则加以系统的组织与整理,将之编辑成目录,供人们进行数据的检索。为了全面、快速地获取卫星数据的主要信息,首先要对所接收到的卫星数据进行快视处理,按照一定的方式生成以景为单位的元数据和浏览图像,然后传递给目录服务系统组织存储。当人们要检索自己所感兴趣的数据时,即可利用目录服务系统提供的检索功能,根据遥感数据的元数据信息和查询要求组织查询,获取相应的存档数据信息,供后继使用。遥感卫星数据目录服务系统的工作流程如图 1 所示。

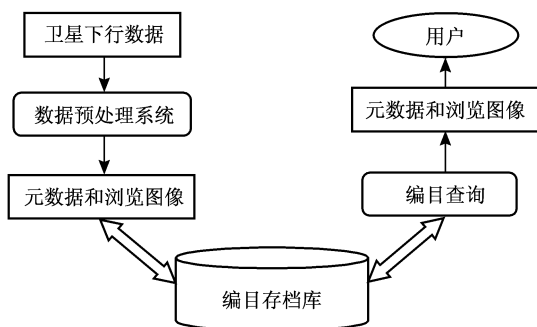


图 1 遥感卫星数据目录服务系统工作流程图

Fig 1 The procedure diagram of remote sensing satellite data catalogue and archive system

2 遥感卫星编目数据的特点

遥感卫星编目数据既有遥感数据本身的描述信息,又包含遥感数据的空间信息和时间信息,是空间

数据和普通属性数据的结合体。因此,它除了具有普通数据的选择性、可靠性、时间性、完备性、详细性和综合性的基本特征外,还具有如下特点:

(1) 空间性。这是编目数据的重要特征,它描述了遥感图像的位置,决定了遥感数据和其他空间数据之间的空间拓扑关系,构成了遥感编目数据空间推理的基础。遥感数据的空间属性可以使用点状和面状要素描述,以表示其地理位置信息,与其他空间数据源建立拓扑关系^[1],实现对遥感数据的空间查询和空间分析。同时,根据卫星上传感器的扫描轨道为固定的或可调整的两类情况,卫星数据的编目数据也具有规则和 irregular 的空间分布特征。

(2) 复杂性。遥感数据描述的是现实世界中的地物和地貌特征,非常复杂,必须经过一定的处理,才能提取人们感兴趣的信息。对于不同的遥感卫星数据,人们所关心的内容也有差别,如光学遥感与微波遥感因为成像原理不同,所形成的遥感图像反映了地物的不同方面,人们关心的元数据信息就有所区别。遥感编目数据的复杂性包括人为的取舍数据,定义遥感数据的元数据。因此,遥感数据的编目数据项也因卫星、数据处理系统和人们关心的内容的不同而有所差别。

(3) 大容量性。遥感编目数据不仅包括普通文本数据,而且包括浏览图像数据,这就使得编目数据量急剧增加,其数据管理量可以达到太字节级。

(4) 多元、多尺度性。遥感数据目录服务系统包括了多传感器、多波段、多分辨率的遥感数据的元数据。不同的传感器有不同的成像原理和成像分辨率,同一传感器又可能有不同的波段和成像模式,因此生成的编目数据有多种数据源和多种尺度。

(5) 永久性。编目数据是对遥感卫星数据性质和质量进行描述的结构化信息,是遥感卫星数据的目录。编目数据的丢失会严重影响用户获取遥感卫星数据信息,所以它们与遥感卫星数据一样属于非常宝贵的资料,需要永久保存。

3 系统研制的关键技术

遥感卫星数据目录服务系统是一个业务化运行系统,是遥感卫星数据分发部门与外部用户的接口,系统对性能、可靠性、可获取性及可维护性的要求非常高,同时系统需要管理海量的遥感卫星数据的编目数据,给目录服务系统的研制带来了很大的挑战。因此,在研制新型业务化运行的目录服务系统过程

中需要解决海量数据存取、空间数据存取、高可靠存档系统的实现、编目存档数据库的性能优化四项关键技术问题,这些关键技术的解决将为建立高性能、高可靠的新型目录服务系统奠定良好的基础。

3.1 ‘统分结合’的海量数据存取

由于目录服务系统不仅需要管理数据记录众多的属性数据,还要管理大量的图像数据,随着系统的不断运行,其数据量急剧膨胀,这对目录服务系统的数据查询性能造成了很大的影响^[2]。如果在系统设计过程中没有认真考虑这种情况,那么随着时间的推移,系统的查询效率会急剧下降。本文根据计算机技术的发展状况和编目数据的特点,采用“统分结合”的方法设计目录服务系统的存储结构,即采用“面向整体存储”和“分时存储”相结合的方案解决目录服务系统海量数据存取问题。

所谓“面向整体存储”的存储方案是指从面向对象的观点出发,将空间数据、属性数据和浏览图像作为一个整体以关系数据库表的形式统一存储,使用者不仅可以像通常那样对表中数据进行查询、合并,还可以进行由空间数据查询属性和由属性数据查询目标的空间分布等操作。这些查询操作在服务器端进行,满足查询条件的数据在服务器端缓冲存放并返回给客户端。这种缓冲存放的方法降低了网络负荷,提高了执行效率。在数据查询时,采用标准的结构化查询语言(SQL)命令来访问和操纵数据库中的数据。空间数据、属性数据以及浏览图像作为一个整体进行存储和管理,可以增强遥感数据的完整性与一致性,减少查询的I/O开销,同时对整个系统的维护也更加方便。

“分时存储”是指按编目数据的接收时间顺序,以统一的逻辑分段存储于不同的物理空间中,即每种编目数据在逻辑上以一个表的形式存在于数据库中,在物理上按遥感数据获取的时间存储于不同的物理空间中,非当前更新的物理空间设置为只读空间,实现以时间分区为基础的存档系统。这种存储方式有如下优点:首先,编目数据的备份和恢复方便。因为在存档系统中绝大部分存储空间处于只读状态,这部分数据只需进行一次永久备份即可。如果属于只读存储空间的数据出现错误,仅需要恢复出错存储空间中的数据,而在恢复过程中数据库其他部分的数据仍然可以被访问,按这种方式对数据库进行恢复所花的时间比以前少得多。这些对海量存储系统是非常重要的。其次,加快了数据的查询速度。编

目数据表进行分区后,逻辑上与未分区的表是相同的,所以分区不会对应用系统开发造成任何影响。但编目数据已经按时间进行了一次划分,当用户以时间作为查询条件时,数据库系统会根据分区时间进行一次过滤,后面的查询处理只在满足条件的分区中进行^[3]。第三,增强了系统的健壮性,减少了数据库出错的可能。只读储存空间只允许用户和管理员做查询操作,不允许修改或删除,因此增加了对数据的保护,进一步维护了编目数据的安全。

3.2 双模式的空间数据存取方案

目前可选的空间数据管理方式有多种,既可以使用数据库厂商提供的空间数据管理模块来实现,也可以使用建立在关系数据库基础上的空间数据操纵引擎来实现。由于编目查询系统存在复杂的空间查询和空间分析,故对空间数据的管理使用专业的空间数据操纵引擎,而属性数据仍然使用当前成熟的关系数据库来管理。在这种双模式的存储方案下,简单的空间查询和分析可以直接利用关系数据库实现,而对于复杂的空间查询和分析,则使用关系数据库和空间数据操纵引擎共同完成。虽然该方案会造成一些数据冗余,但可以提高系统的性能,是一种典型的以空间换取时间的方法。

3.3 经济、高效的集群部署方案

在旧的目录服务系统中,系统结构设计者主要考虑了硬件故障对系统造成的影响,对存储设备采用某种形式的备份,如 RAID 盘柜,但却忽视了系统运行软件出错的可能性。如果系统的运行软件出现错误,系统就会宕机,停止服务。若系统正在更新数据,有造成重要数据丢失的危险。为了确保系统7×24小时可获取能力和系统安全,系统采用集群技术实现高可靠的目录服务系统。

集群系统从功能角度可以分为两种类型:一种是以 Windows 为代表的网络负载均衡集群;另一种是以 Oracle 为代表的真实应用集群服务器。网络负载均衡集群系统由操作系统实现负载的平衡,不需要昂贵的集群硬件支持,但它不具备高可靠的功能,不能实现运行任务的平滑过渡^[4];而真实集群应用服务器不仅可以做到负载均衡,而且支持系统的高可靠性^[5]。当其中一台服务器出现故障后,另一台服务器可以平滑地接过处理任务,实现系统服务的无缝切换。其缺点是需要比较昂贵的集群硬件的支持。对于基于浏览器/服务器结构的目录服务

系统而言,在综合考虑系统的经济性和可靠性的前提下,系统使用网络负载均衡的集群实现 WEB 服务器集群,而采用真实集群应用服务器实现数据库服务器的集群。这样既可以保证编目数据的安全,又可以确保目录服务系统的高可获取能力。

3.4 基于规则和代价相结合的编目存档数据库查询优化方案

编目存档数据库的性能受到多方面因素的影响,除了考虑编目存档数据库的物理设计外,数据库的各种参数设置、SQL语句、操作系统、磁盘 I/O、内存和 CPU 等也会对数据库系统性能产生较大的影响,编目数据库可以从这些方面进行优化。应该指出的是编目数据库可以根据需要,适当地增建索引,以提高系统的查询效率,因为编目数据库的更新操作比较少,索引对数据库产生太多的性能负担。在确定 SQL语句的执行路径时,数据库提供了基于规则的优化和基于代价的优化方法。基于规则的优化是指数据库依赖于一套严格的语法规则,只按照给定的规则确定 SQL语句的执行路径,不管数据表和索引的内容是否发生变化;而数据库使用基于代价的优化时,是根据数据库内容表和索引的统计信息来确定 SQL语句的执行路径^[3]。一般来说,基于代价的优化方法比基于规则的优化方法具有更好的性能,但对于数据记录特别大的表使用基于代价的优化方法会造成检索效率不稳定,有时严重影响系统的检索效率,所以对编目数据库的优化,采用基于规则的优化和基于代价的优化相结合的方式实现。

4 系统实现

当前目录服务系统主要使用浏览器/服务器结构(B/S)向用户提供检索服务,以便用户能够通过网络直接检索遥感卫星数据。目录服务系统由数据库服务器、磁盘阵列、自动磁带库、存储管理软件、Oracle 数据库、微软的 IIS (Internet Information Service)、编目检索应用软件构成,其硬件结构如图 2 所示。两台数据库服务器由光纤通道分别与磁盘阵列、磁带库相连,每台服务器上安装了 Oracle 数据库和真实应用集群软件,实现了负载均衡与单点故障自动切换;两台应用服务器通过千兆以太网分别与两台数据库服务器相连,每台应用服务器安装了 Windows 的网络负载均衡集群软件,用户通过网络使用统一的 IP 地址访问应用服务器,而实际提

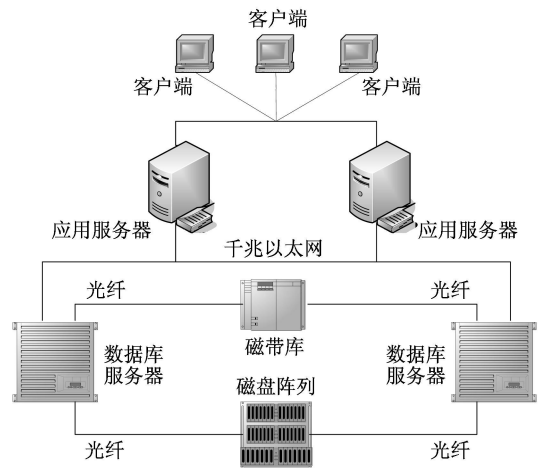


图 2 目录服务系统硬件结构图

Fig 2 The hardware structure figure of catalogue and archive system

供服务的应用服务器则由集群软件根据两台服务器的负载状况确定。如果正在提供服务的应用服务器出现故障,用户只需重新提交请求,系统即会自动切换到另一台服务器,继续为用户提供服务,切换过程对用户而言是透明的。

用户通过浏览器向应用服务器提出数据检索请求后,应用服务器从编目数据库中提取满足条件的数据记录,然后提交给浏览器显示。目录服务系统提供的功能主要包括:用户管理、文本检索、地图检索、高级检索、归档数据统计、订单申请以及检索配置等。下面是目录服务系统中涉及海量数据检索的功能描述。

(1) 文本检索。文本检索是目录服务系统最基本的功能,也是用户使用最频繁的功能。文本检索具有简单文本检索和复杂文本检索两种模式。对于简单文本模式,用户可以设定卫星数据目录、时间方位、空间范围、模式和产品级别参数进行检索,主要供普通用户使用;对于复杂文本模式,用户可以选择卫星平台相关参数、地理分幅、存储参数、轨道号等参数进行检索,此功能主要供专业用户精确定位遥感数据使用。

(2) 地图检索。地图检索为用户提供了直观的空间数据检索环境。在 WEBGIS 技术的支持下,用户可以进行基本的地图操作功能(不同比例尺图层无缝切换,地图放大、缩小、漫游等),可以直观地在地图上选取矩形框、点目标、多边形、各级行政区划等作为空间检索条件,也可以结合文本检索的检索条件实现组合检索。

(3) 高级检索。对于级别较高的用户,系统提

供了更加灵活的高级检索功能。用户可以根据查询需求定制检索条件,克服系统提供的查询选项的限制,但此功能要求用户对系统的数据结构非常熟悉。用户可以手工直接输入 SQL 语句进行检索,也可以根据页面上的选项来构建 SQL 语句。

(4) 归档数据统计。统计功能模块用于统计某一较大时间段或者较大地理空间范围内的数据条目及数据覆盖情况,统计结果可以很直观地在地图上给出数据覆盖情况。

图 3 是从目录服务系统检索到的印度 RESOURCESAT-1 (RS-P6) 卫星数据覆盖图。

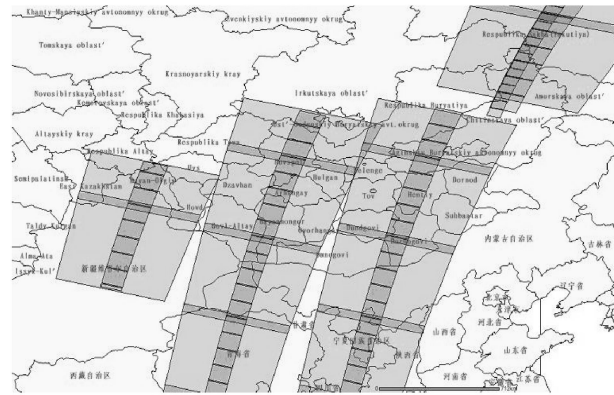


图 3 RS-P6 卫星数据覆盖图

Fig 3 The image of RS-P6 satellite data coverage

5 结 论

遥感卫星数据目录服务系统是实现遥感向产业

化发展转化的不可或缺的应用服务平台,开发高性能、高可靠、可维护的目录服务系统是遥感应用的一项基本工作和重要研究课题。本文从这几个方面出发,根据当前技术的发展及编目数据的特点,设计并实现了新型的遥感卫星数据目录服务系统。目前该系统已经投入业务化的生产运行,取得了较好的效果。因此,遥感卫星数据目录服务系统作为一个专业的地理信息系统,在以后的更新升级过程中,不仅要考虑它和通用 GIS 系统的共性,更要把握其自身的特点,不断引入地理信息系统技术和计算机技术的优秀成果,使遥感卫星数据目录服务系统趋于成熟和完善,以适应遥感飞速发展的需要。

参 考 文 献 (References)

- [1] Wu L. Geographic Information Systems [M]. Beijing: Science Press, 2001. [郭伦. 地理信息系统——原理、方法和应用 [M]. 北京: 科学出版社, 2001.]
- [2] Eos Ltd. Catalogue and Browse Systems Administrator Manual [M]. <http://www.eos.co.uk>; 26-36.
- [3] Oracle. Oracle9i Database Administrator's Guide Release 2(9.2) [M]. <http://www.oracle.com>. 2002.
- [4] Oracle. Establish High Availability Database Server Using Oracle Real Application Clusters [R]. <http://www.oracle.com>. March, 2002. [Oracle 使用 Oracle Real Application Clusters 建立高度可用的数据库服务器 [R]. <http://www.oracle.com>. 2002.]
- [5] Microsoft. Microsoft Windows 2000 Advanced Server Clustering Services [M]. Beijing: Tsinghua University Press, 2002. [Microsoft. Microsoft Windows 2000 Advanced Server 群集服务 [M]. 北京: 清华大学出版社, 2002.]