

doi: 10. 6046/gtzyyg. 2017. 04. 32
引用格式: 随欣欣,王彦佐,晋佩东,等. 基于 RasterCatalog 的海量遥感数据存储及快速浏览技术研究[J]. 国土资源遥感, 2017,29(4):214–218. (Sui X X,Wang Y Z,Jin P D,et al. Research on mass remote sensing image storage and rapid browse based on RasterCatalog[J]. Remote Sensing for Land and Resource,2017,29(4):214–218.)

基于 RasterCatalog 的海量遥感数据存储 及快速浏览技术研究

随欣欣¹, 王彦佐¹, 晋佩东¹, 魏英娟¹, 王文凯², 马骏欢²

(1. 中国国土资源航空物探遥感中心,北京 100083; 2. 北京吉威数源信息技术有限公司,北京 100043)

摘要: 高分一号卫星自 2013 年 4 月 26 日发射至今,已获取了大量的数据,数据量已达到了 TB 级,而且还在持续增长。基于高分数据的有效管理、检索和展示的需求,对海量数据存储模型和影响图像浏览速度的因素进行分析。通过实验验证,采用压缩质量为 50% 的 JPEG 压缩和双线性插值法构建影像金字塔,以 64×64 切片大小进行数据存储,并通过构建 GDB RasterCatalog Layer,能够实现海量遥感数据的存储与快速浏览展示,满足高分系列卫星在行业应用中数据管理与显示需求。
关键词: 高分数据; RasterCatalog; 快速浏览
中图法分类号: TP 79 **文献标志码:** A **文章编号:** 1001–070X(2017)04–0214–05

0 引言

随着遥感技术的不断发展,遥感数据的获取途径也在不断的增多,并且随着空间分辨率、波段数的提高,遥感影像的数据量级别也在不断增长。中国国土资源航空物探遥感中心(下文简称航遥中心)2016 年 5 月就已存储各类遥感影像数据达 176 万余景,数据量达 533.9 TB,其中高分一号卫星影像数据 81.8 万余景,189.7 TB; 高分二号卫星影像数据 45.5 万余景,220.2 TB,且国产卫星数据仍以每天 2 000 余景,1.0 TB 数据量的增长速度不断增加。航遥中心承担着高分国土资源业务卫星数据的统筹、管理、服务及保障任务,巨大的数据量对海量数据的高效管理和服务提出了挑战,急需应用合理的技术高效管理这些海量高分影像,为应用提供数据支持。近年来,很多专家和学者针对海量数据的存储管理进行了大量的研究与系统建设,如许自舟等^[1]进行了海量影像数据储存与发布的技术研究,表明采用影像压缩、影像金字塔和影像切片等技术能有效提高海量影像数据储存与发布性能;薛涛等^[2]研究了基于空间数据库、文件目录方式和 Web Service 等方式的数据存储和发布。虽然数据存储

管理方式层出不穷,但是应用最广泛的是基于关系型数据库+空间数据引擎与文件系统相结合的方式实现海量数据的管理^[3–5]。而针对遥感影像数据的浏览展示,以往主要通过单一 Oracle 关系数据库 BLOB 字段存储,影像检索及叠加展示效率较低;而后续的研究主要运用金字塔、影像分块等技术^[6],能够实现较好的浏览展示效果。
本文通过研究发现,基于 RasterCatalog 栅格数据组织方式结合影像金字塔、数据压缩技术能够实现海量遥感影像数据的高效存储与快速展示,其效果明显优于关系数据库 BLOB 字段的影像数据存储方式。

1 影像数据存储技术

近年来,随着遥感影像数据量的不断增长,海量遥感影像数据的存储技术也随之迅速发展,国内外学者提出了以关系数据库、空间数据库为主的栅格数据存储技术^[7],其中应用广泛的有 Oracle 关系数据库 BLOB 字段存储和空间数据引擎 ArcSDE 存储 2 种。
Oracle 关系数据库 BLOB 字段存储方式易于管理,具有可伸缩性、安全性及可用性的优点;空间数

收稿日期: 2016–05–05; 修订日期: 2016–06–21
基金项目: 中国国土资源航空物探遥感中心项目“高分基础数据支撑分系统建设”(编号: 201330106)和“资源一号 02C 卫星应用系统(地矿)影像产品数据整合建库及管理”(编号: 201530605)共同资助。
第一作者: 随欣欣(1983–),女,博士,工程师,主要从事遥感信息化研究。Email: sxx5666@163.com。

据引擎支持超大型数据集,可灵活、高性能地搜索提取空间数据,是专门为多用户、分布式环境设计开发的体系架构,拥有逻辑上的无缝、连续的非瓦片式的空间目标数据模型,并支持多用户并发查询的快速响应。以下针对关系数据库 BLOB 字段和空间数据引擎的存储原理进行分析。

1.1 关系数据库 BLOB 字段存储

在 Oracle 数据库中,大对象(large objects,LOB)是用来存储大量的二进制和文本数据的一种数据类型(一个 LOB 字段可存储多达 4 GB 的数据)。Oracle8i 以上支持 3 种类型的内部 LOB:二进制数据(BLOB)、单字节字符数据(CLOB)和多字节字符数据(NCLOB)。其中 BLOB 字段适用于存储大量的二进制数据,如图像、视频和音频等。在通常的遥感影像数据管理系统中,影像数据以 BLOB 字段存储在 Oracle 数据库中。BLOB 字段包含 BLOB 列、BLOB 数据段和 BLOB 索引段 3 部分: BLOB 列由 BLOB 定位符和 BLOB 值组成,BLOB 定位符是一个指向 BLOB 字段实际存放位置的指针,BLOB 值存储 BLOB 数据; BLOB 数据段由多个数据块(chunk)组成,每个 chunk 用来存放 BLOB 数据; BLOB 索引段由多个 entry 组成,主要用于存放 BLOB 索引。

遥感影像由于数据量较大的原因,通常存放在 BLOB 数据段中。若 chunk 未超过 12 块,通过 BLOB 定位符指向存储在 BLOB 字段中的影像,不使用 BLOB 索引段; 否则需要用 BLOB 索引段进行寻址,数据段中每一个 chunk 地址都存放在 BLOB 索引中,通过索引段寻址方式,可以定位到影像在 BLOB 字段中的存储位置,从而读取高分影像进行展示。这就导致 BLOB 字段存储和网络资源占用大,影响遥感影像的检索获取效率,多用户对海量遥感影像的并发访问受限。

1.2 空间数据引擎存储

空间数据引擎(spatial database engine,SDE)是 GIS 中介于应用程序和空间数据库之间的中间件技术,能够在常规数据库管理系统功能之外获得空间数据存储和管理的能力,具有代表性的是 ESRI 的 ArcSDE。RasterCatalog 数据模型是 ArcSDE 中存储、管理栅格数据的一种组织方式,通过 ArcEngine 提供的接口,将多个独立的影像数据文件无缝地显示在一个 RasterCatalog 图层中,并同时存储影像数据的空间位置信息。

RasterCatalog 图层在 Oracle 数据库中存储在若干张相互关联的表中,其中真正存储栅格影像数据的表包括: 附录信息表、分块表、波段表和栅格表。附录信息表存储影像颜色映射、统计信息等; 分块

表真正存储栅格像素值; 波段表存储影像的所有波段信息; 栅格表存储 RasterCatalog 中包含的栅格的属性。对于高分影像数据,SDE 通过建立 Raster_Columns(SRID 外键 < Spatial_References >)元数据表,来建立与导入的业务表(business table)之间的关系。

栅格目录中的每个影像数据都是独立的,利于数据更新和数据库维护,常用于管理分幅数据或者同一地区的多期数据,航遥中心现有影像多数为原始分幅影像,具有数据量大、覆盖面积广和部分地区多次覆盖的特点,使用 RasterCatalog 存储方式能够有效管理影像数据。

2 海量影像数据显示优化技术

2.1 影像数据压缩类型与压缩质量

海量遥感影像数据管理系统在管理海量影像数据时,为了节省数据库中的数据存储空间,在影像数据入库前,对数据进行压缩存储。系统内部在导入栅格影像数据时,可以根据需要选择不同的压缩类型和压缩质量,常见的压缩类型有无压缩、LZ77 和 JPEG 等,其中 LZ77 为无损压缩类型,JPEG 为有损压缩类型,对于 JPEG 有损压缩类型还可以选择不同的压缩质量完成数据压缩。

2.2 影像数据金字塔

对原始图像创建金字塔是改善影像数据显示性能的一种重要方法。影像金字塔的创建过程就是一个重采样的过程,金字塔文件中存放了多种空间分辨率的栅格数据,同一空间分辨率的栅格数据被组织在一个层面(layer)内,而不同空间分辨率的栅格数据具有上下的垂直组织关系: 越靠近顶层,数据的空间分辨率越小,数据量也越小,只能反映原始数据的概貌; 越靠近底层,数据的空间分辨率越大,数据量也越大,更能反映原始详情。

影像金字塔的各个连续图层间通常以 4:1 的比例进行重采样,常用的重采样方法有最邻近法、双线性插值法和三次卷积法。卫星或航空摄影等 tiff 和 img 格式的影像数据多采用双线性插值法或者三次卷积法^[1]进行重采样,以保证影像平滑显示效果。影像数据越大、金字塔级数越多,创建金字塔花费时间就越长,但在浏览展示时也将节省更多时间。

2.3 影像数据切片

随着遥感技术的发展,卫星遥感影像数据单幅数据量也越来越大,影像切片技术通过将一幅较大影像数据切分成很多小的切片进行存储,在加载影像时,根据视图窗口地理位置加载相应的切片数据,

从而减少数据加载时间,提高数据浏览展示的效率。当利用 RasterCatalog 图层进行影像数据存储时,栅格影像数据(包括金字塔)被切割成一块块存储在 SDE_BLK_<N> 表中,每条记录代表了一个切片。切片默认尺寸大小为 128 像素 × 128 像素,通常根据实际影像大小对切片尺寸大小进行调整。

3 实例分析

本文在研究影像数据高效展示的基础上,结合现有的影像数据存储技术,以 Visual Studio 2010 作为系统开发平台,Oracle11g 作为数据库管理平台,ArcSDE10.1 作为中间件消息处理技术,构建了影像

处理支撑数据库管理系统。作为一个海量数据管理系统,系统存储管理了大量国产卫星影像数据(如 GF-1 和 ZY02C 等)以及一些其他辅助数据,共计 176 万余景,数据量约 534 TB。

通过对影像数据存储技术、展示技术的对比研究,关系数据库 BLOB 字段存储方式存储时直接存储的是影像文件,占用较大存储空间,因此影像处理支撑数据库管理系统中采用了空间数据库 RasterCatalog 图层方式存储。系统构建了以数据压缩入库,创建金字塔,创建切片生成地图缓存,客户端浏览展示的工作流程,可实现影像数据的高效展示。影像数据存储与展示流程如图 1 所示。

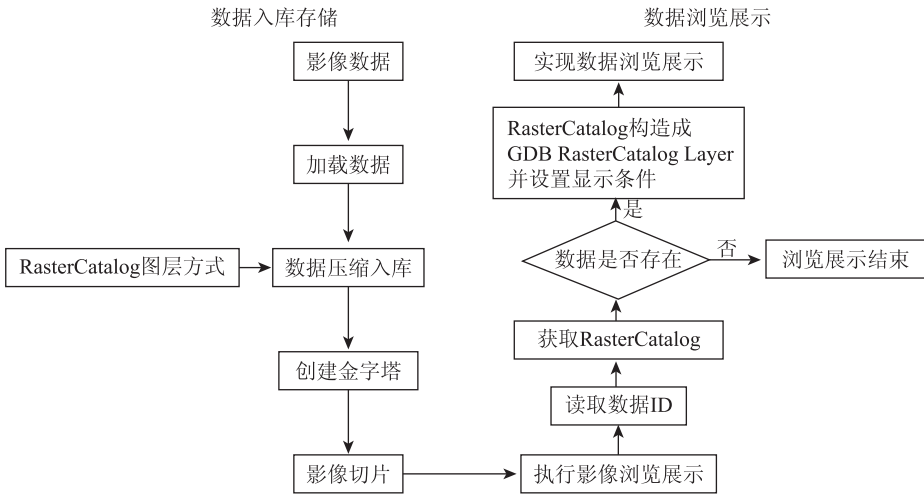


图 1 影像数据存储与展示流程

Fig. 1 Storage and display process of image data

3.1 压缩比对浏览效果的影响

选取一景 4.72 G 大小、tiff 格式无压缩无金字塔的栅格影像数据为数据源,在 ArcCatalog 中进行无压缩、LZ77 和 JPEG 共 3 种压缩类型的压缩测试,得到的测试结果以及视觉效果如表 1 所示。

表 1 压缩类型性能对比

Tab. 1 Performance contrast of compressed format					
压缩类型	压缩质量/%	压缩后数据量	压缩比/%	耗时/s	视觉效果
无压缩	—	4.72 G	100.0	0	纹理清晰
LZ77	—	3.92 G	83.1	968	纹理清晰
	100	2.05 G	43.4	223	纹理清晰
JPEG	75	598 M	12.4	171	纹理较清晰
	50	396 M	8.2	146	纹理明显

对比实验结果得出,LZ77 无损压缩视觉效果很好,但由于是无损压缩,存储空间占用较大。压缩质量为 75% 的 JPEG 在压缩比纹理效果相对较好,在 1:3 000 比例尺下,纹理仍然能保持比较清晰;压缩质量为 50% 时,视觉效果较 75% 略差,但整体效果尚可,在 1:3 000 比例尺下,纹理较明显。

万方数据

通过空间库 RasterCatalog 图层方式入库时,真

正存储栅格像素值的 SDE_BLK_<N> 表存在增长过快的情况,出现 N 倍的存储空间膨胀。考虑到数据存储空间与视觉效果,对比分析无损压缩以及压缩类型为 JPEG,压缩质量分别为 75% 和 50% 这 3 种情况对影像数据入库存储时存储空间膨胀的影响(如表 2 所示)。实验用 RasterCatalog 共 72 条数据,原始数据文件共 11.2 G。存储空间膨胀比率为栅格数据占用存储空间与原始文件大小的比值。

表 2 压缩类型结果对比

Tab. 2 Results of compressed format					
压缩类型	压缩质量/%	BLK 表记录数	BLK 表大小/G	平均单条数据占用空间	存储空间膨胀比率/倍
无损压缩	—	19 810	188.2	2.61 G	16.78
JPEG	75	19 810	14.9	212.38 M	1.33
	50	19 810	0.4	5.69 M	0.04

通过对比得出,当压缩类型为 JPEG、压缩质量为 50% 时存储空间膨胀比率最低,有利于影像数据的存储管理。

通过对压缩类型、压缩质量、视觉效果和存储空

间膨胀比率的综合分析,影像处理支撑数据库管理系统采用压缩类型为 JPEG,压缩质量为 50% 进行压缩入库管理,能够满足大范围影像浏览的需求。

3.2 金字塔构建方式对浏览效果影响

数据入库后,为了实现快速浏览展示,程序内部对已入库的数据建立影像金字塔。对以 JPEG 压缩类型、50% 压缩质量进行数据入库的数据分别采用最邻近法、双线性插值法和三次卷积法 3 种常用的重采样方法进行对比(表 3),设置构建 7 层金字塔。

表 3 重采样方法性能对比

Tab.3 Performance contrast of different resampling methods

重采样方法	完成时间/s	金字塔文件数据量/M	视觉效果
最邻近法	410	465	较差
双线性插值法	433	446	很好
三次卷积法	614	432	很好

对比 3 种重采样方法,就压缩速度、文件数据量和视觉效果结合来看,双线性插值法建立金字塔耗时较短,且能够获得很好的视觉效果,但最终文件数据量略大于三次卷积法。而三次卷积法耗时太长,不建议采用,因此影像处理支撑数据库管理系统选取了双线性插值法建立影像金字塔。

3.3 图像切片大小对浏览速率的影响

空间数据库对影像进行存储时,切片大小默认为 128 像素×128 像素。对于无压缩的数据,128 像素×128 像素切片大小的栅格影像数据切片应该包含 16 K 个像素。就高分影像数据而言,影像位深为 16 Bit,即每个像素占据 2 个字节;因此,128 像素×128 像素切片大小将在数据库中占据 32 KB 存储空间。采用 Oracle 数据库进行入库时,默认创建数据库的数据块大小为 8 K,则 128 像素×128 像素切片占据了 2 个数据块,Oracle 读取这个切片就需要做 4 个 I/O 操作。在数据库中,通过尽量减少 I/O 操作来提高数据库访问效率,因此,针对现有高分数据源以及系统部署环境等实际情况,系统采用 64 像素×64 像素作为最优切片大小进行影像切片。

3.4 影像数据展示速度比较

根据现有技术,主要设计如下 3 种技术路线实现影像数据展示。

方式 1:以关系数据库 BLOB 字段作为影像数据存储方式。在数据库中,通过数据 ID 判断影像数据是否存在。如果存在,则下载影像文件并生成坐标参考文件,读取文件并构造 IRasterCatalog Layer,从而实现影像数据的浏览展示;如果不存在,则不进行展示。

方式 2:以空间数据库作为影像存储方式。在

数据库中,通过数据 ID 获取 RasterCatalog 相关表,判断影像数据是否存在。如果存在,则获取数据库中存储的影像数据构建 Raster Layer,实现影像数据的浏览展示;如果不存在,则不进行展示。

方式 3:以空间数据库作为影像存储方式。在数据库中,通过数据 ID 获取 RasterCatalog 相关表,判断影像数据是否存在。如果存在,则直接通过 RasterCatalog 构建成 GDB RasterCatalog Layer 并设置显示条件,实现影像数据的浏览展示;如果不存在,则不进行展示。

在影像处理支撑数据库管理系统中,针对以上 3 种浏览展示技术进行效率测试,结果如表 4 所示。

表 4 影像数据浏览展示性能分析

Tab.4 Performance analysis of image data's browse and display (s)

数据量/景	方式 1	方式 2	方式 3
10	5 ~ 10	5 ~ 10	<1
50	>20	12 ~ 14	<1
100	>40	24 ~ 37	<2
500	>135	96 ~ 120	<3
1 000	>410	230 ~ 325	<5

通过对比分析,方式 3 的数据浏览展示效率最高,且数据量越大,优势更加明显。因此在影像处理支撑数据库管理系统中最终采用该方式实现影像数据的快速浏览展示。

4 应用分析

在相同的环境下,对采用了方式 3 的影像处理支撑数据库管理系统和采用了方式 1 的其他影像管理系统进行了应用对比分析。通过验证分析,针对超大型数据集空间数据库 RasterCatalog 图层存储方式较 Oracle 数据库中关系数据库 BLOB 字段存储方式更具优势。

将 10 000 景高分一号卫星影像数据分别在影像处理支撑数据库管理系统及其他影像管理系统中进行数据入库管理和浏览展示,统计不同数据量执行 10 次操作的平均耗时,结果如表 5 所示。

表 5 不同系统中数据浏览展示性能分析

Tab.5 Performance analysis of different system's browse and display functions (s)

数据量/景	影像处理支撑数据库管理系统(方式 3)	其他影像管理系统(方式 1)
10	0.79	8.29
50	0.98	15.77
100	1.64	32.70
500	2.31	98.92
1 000	4.20	362.90

对比实验结果可以看出,影像处理支撑数据库管理系统较其他影像管理系统在数据浏览展示效率上存在明显的优势,能够较好地实现海量影像数据的快速浏览展示。

5 结 论

本文通过对关系数据库 BLOB 字段存储和空间数据库 RasterCatalog 存储 2 种存储栅格影像的方式进行了对比分析,选择空间数据库 RasterCatalog 作为航遥中心海量数据的存储方案。通过对影响浏览速度的参数进行实验,定量分析了不同的压缩格式、切片大小和金字塔采样方式等对海量数据浏览性能的影响,实验证明当采用压缩质量为 50% 的 JPEG 压缩和双线性插值法构建影像金字塔,以 64 像素 × 64 像素切片大小进行数据存储时,浏览效果最优。在此基础上完成了航遥中心影像处理支撑数据库管理系统 533 TB 数据的入库管理,并通过与前期开发的其他影像管理系统进行了性能对比,验证了该系统能显著提高高分影像的浏览速度,表明了基于 RasterCatalog 栅格数据模型能够实现海量遥感数据的高效存储与快速展示,满足航遥中心现有影像数据的存储管理需求。

参考文献 (References):

[1] 许自舟,孙淑艳,梁 斌,等. ArcGIS 中海量影像数据存储与发布技术[J]. 海洋环境科学,2014,33(1):99-104.
Xu Z Z, Sun S Y, Liang B, et al. Technique for mass image data

management and publication based on ArcGIS[J]. Marine Environmental Science, 2014, 33(1):99-104.
[2] 薛 涛,刁明光,李建存,等. 资源环境遥感海量空间数据存储、检索和访问方法[J]. 国土资源遥感,2013,25(3):168-173. doi:10.6046/gtzyyg. 2013.03.28.
Xue T, Diao M G, Li J C, et al. Approach to storing, retrieving and accessing mass spatial data in resources and environments remote sensing[J]. Remote Sensing for Land and Resources, 2013, 25(3):168-173. doi:10.6046/gtzyyg. 2013.03.28.
[3] 王昀昀,朱勤东. 基于 ArcSDE 的影像数据入库研究[J]. 测绘通报,2013(1):84-86.
Wang Y Y, Zhu Q D. Research on image data storage based on ArcSDE[J]. Bulletin of Surveying and Mapping, 2013(1):84-86.
[4] 钟永友. 基于 ArcSDE 的物流信息系统空间数据库设计[J]. 计算机与数字工程,2010,38(1):74-77.
Zhong Y Y. Design of logistics information system spatial database based on ArcSDE[J]. Computer and Digital Engineering, 2010, 38(1):74-77.
[5] 胡仪员. 基于 ArcSDE for Oracle 的地震灾害紧急救援队数据库系统建设[D]. 成都:西南交通大学,2008.
Hu Y Y. The Construction of Database for Earthquake Disaster Emergency and Rescue Team Based on ArcSDE for Oracle[D]. Chengdu: Southwest Jiaotong University, 2008.
[6] 王华斌,唐新明,李黔湘. 海量遥感影像数据存储管理技术研究及实现[J]. 测绘科学,2008,33(6):156-157,153.
Wang H B, Tang X M, Li Q X. Research and implementation of the massive remote sensing image storage and management technology [J]. Science of Surveying and Mapping, 2008, 33(6):156-157, 153.
[7] 李宗华,彭明军. 基于关系数据库技术的遥感影像数据建库研究[J]. 武汉大学学报(信息科学版),2005,30(2):166-169.
Li Z H, Peng M J. Remote sensing image database based on RD-BMS[J]. Geomatics and Information Science of Wuhan University, 2005, 30(2):166-169.

Research on mass remote sensing image storage and rapid browse based on RasterCatalog

SUI Xinxin¹, WANG Yanzuo¹, JIN Peidong¹, WEI Yingjuan¹, WANG Wenkai², MA Junhuan²
(1. China Aero Geophysical Survey and Remote Sensing Center for Land and Resources, Beijing 100083, China;
2. Beijing JW Digital Source IT Co. Ltd., Beijing 100043, China)

Abstract: Since the GF-1 satellite was launched on April 26, 2013, a large quantity of data have been obtained, and the data volumes have entered into the multi-terabyte range, and even continue to grow. For effective management, retrieval and display of high-resolution remote sensing images, the authors take the efficient data storage model and faster browsing into consideration. The related experiments show that the storage and fast browsing of mass data can be realized by constructing the image pyramid by method of JPEG compression quality of 50% and method of quadratic interpolation, storing data with slices of 64 × 64, and creating GDB RasterCatalog Layer, which satisfies the need of industry-specific applications.

Keywords: GF image data; RasterCatalog; rapid browse

(责任编辑:李 瑜)