

doi:10.3799/dqkx.2015.128

基于 NoSQL 的高分高光谱遥感影像 存储模型设计与实现

王晓蕊¹, 杨强根², 陈凤敏³, 马维峰³, 唐湘丹³, 谭兴³, 张时忠⁴

1. 中国地质大学资源学院, 湖北武汉 430074

2. 中国地质大学网络与教育技术中心, 湖北武汉 430074

3. 武汉地大信息工程股份有限公司, 湖北武汉 430074

4. 中国地质大学计算机学院, 湖北武汉 430074

摘要: 如何高效存储、管理呈几何增长的高分辨率、高光谱遥感影像数据, 实现遥感数据的快速处理、检索和可视化是急需解决的问题。应用非关系型数据库技术, 设计了由遥感元数据库、遥感影像数据库以及影像金字塔 3 个部分组成的海量遥感影像存储模型; 建立了由硬件支撑层、数据层、数据服务层及应用层组成的遥感影像存储中间件。通过实验分析, 验证了基于非关系型数据库的遥感影像数据存储模型及中间件对影像数据的读写、提取性能优于传统的关系数据库。研究成果可满足高分高光谱遥感探测与评价模型对海量影像高效存储、管理的需求, 具有重要的实用价值。

关键词: 遥感影像; 非关系型数据库; 元数据; 金字塔。

中图分类号: P237

文章编号: 1000-2383(2015)08-1420-07

收稿日期: 2015-03-25

Storage Model Design and Implementation of High Resolution and Hyperspectral Remote Sensing Image Based on NoSQL

Wang Xiaorui¹, Yang Qianggen², Chen Fengmin³, Ma Weifeng³,
Tang Xiangdan³, Tan Xing³, Zhang Shizhong⁴

1. Faculty of Earth Resources, China University of Geosciences, Wuhan 430074, China

2. Network and Educational Technologies Center, China University of Geosciences, Wuhan 430074, China

3. Wuhan InfoEarth Information Co. Ltd., Wuhan 430074, China

4. College of Computer Science, China University of Geosciences, Wuhan 430074, China

Abstract: It is an issue to be addressed as how to efficiently store and manage massive data that in high resolution and high spectral remote sensing images, and achieve rapid processing, retrieval and visualization of remote sensing data is the problem to be solved. Using non-relational database technology, we designed massive image storage model consists of remote sensing metadata database, image database and image pyramid; set up the remote sensing image storage middleware that consists of the hardware support layer, data layer, data service layer, and application layer. Experiment verifies the remote sensing image data storage model and middleware that based on non-relational database is superior to relational database for video data read and write. Research results prove have great practical value for efficiently storage and management of massive remote sensing image data.

Key words: remote sensing image; not only structured query language(NoSQL); metadata; pyramid.

基金项目: 国家高技术研究发展计划(863 计划)项目(Nos.2012AA12A308, 2008AA121100); 国家自然科学基金项目(Nos.2013023123, 41402293)。

作者简介: 王晓蕊(1979-), 女, 博士, 讲师, 从事为矿床学、地球化学、地学数据库和地学遥感研究。E-mail: wangxiaorui@cug.edu.cn

引用格式: 王晓蕊, 杨强根, 陈凤敏, 等, 2015. 基于 NoSQL 的高分高光谱遥感影像存储模型设计与实现. 地球科学——中国地质大学学报, 40(8): 1420-1426.

随着遥感、电子和计算机技术的不断进步,各部门对遥感观测数据的需求逐渐加大,不同级别和类型的遥感数据不断增加,传统的数据库技术和数据存储管理方法已经不能满足存储和管理海量影像数据的性能需求.海量遥感影像需要快速更新、高效提取的瓶颈已严重制约了遥感技术和空间数据库及地理信息系统(geographic information system, GIS)的集成应用.如何更好的存储和管理海量遥感影像数据,并实现对遥感数据的快速处理、检索及可视化,从而在海量的信息中提取所需信息,提高影像利用率,已经成为了当前空间数据库和地理信息系统研究中迫切需要解决的问题(樊小泊和陈红,2006;李小文,2006;杨永崇和贡建明,2007;王华斌等,2008;刘伟等,2009;牛得学等,2011;周敬利和周正达,2012;赖积保等,2013).

非关系数据库(not only structured query language, NoSQL)是近年来随着互联网技术发展出来的一种新型数据管理技术,具有高性能、易扩展、高可用、灵活存储、大数据等优点,可以高效存取数据,并对数据并发访问(范凯,2010).本文面向高分高光谱遥感影像存储和管理,应用非关系数据库(NoSQL)技术研究并建立基于非关系数据库和传统关系数据库结合的遥感影像数据存储模型和中间件,实现影像大数据集的快速处理、入库、更新和提取读写效率.

1 基于 NoSQL 的遥感影像存储模型设计

由于非关系数据库(NoSQL)技术建模方式和传统的关系数据库建模方式有所差异,本文首先需要基于 NoSQL 技术设计遥感影像数据模型.通过分析遥感数据的数据结构,对海量遥感数据采用统一的空间参考,合理的栅格数据分块分层(王旭东,2012),建立有效的多级影像金字塔和高效的空间索引技术,以实现海量遥感影像快速、高效地存取.

1.1 遥感影像模型

遥感影像包括遥感影像数据和遥感元数据.本文将遥感影像数据模型设计为 3 个部分:元数据(吕雪峰等,2011)、影像数据和影像金字塔(图 1).

(1)遥感元数据部分作为数据检索查询的重要依据,对数据建立索引机制,根据查询条件来高效的查询所需要的数据.由于遥感数据的元数据是规则化的表格数据,可基于关系数据库进行建模,并应用关系数据库实现数据的各类索引、查询、筛选和分析.

(2)影像数据部分通过设计合理的分块方案,将影像数据进行分块,并将分块的数据进行编号,然后按顺序存入非关系数据库.由于 NoSQL 在单个数据对象的读写效率方面优于传统的关系数据库,因此从理论上保证了影像数据的高效读写.

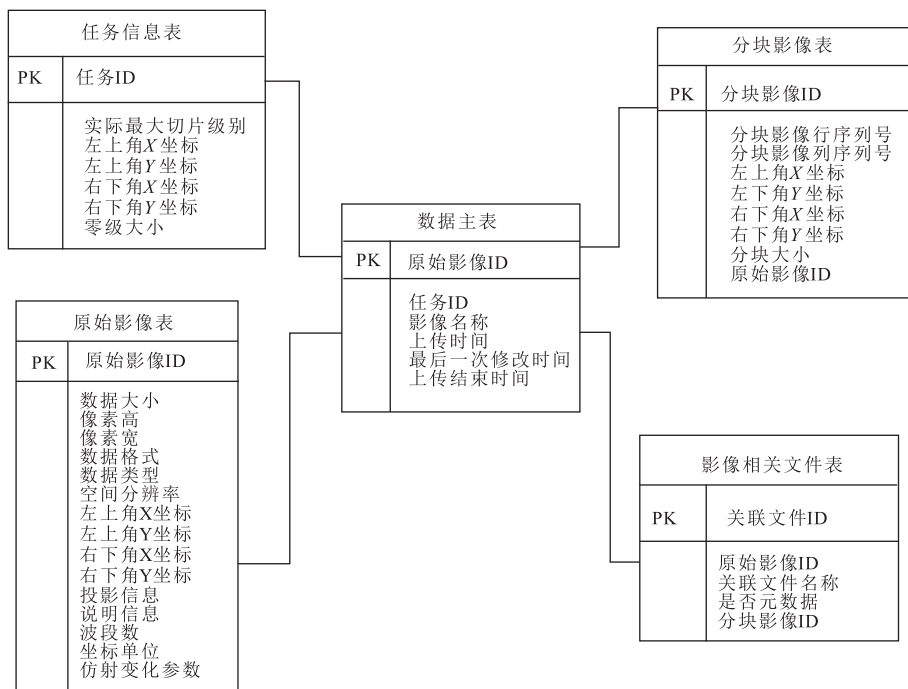


图 1 影像元数据的概念模型

Fig.1 Conceptual model of image metadata

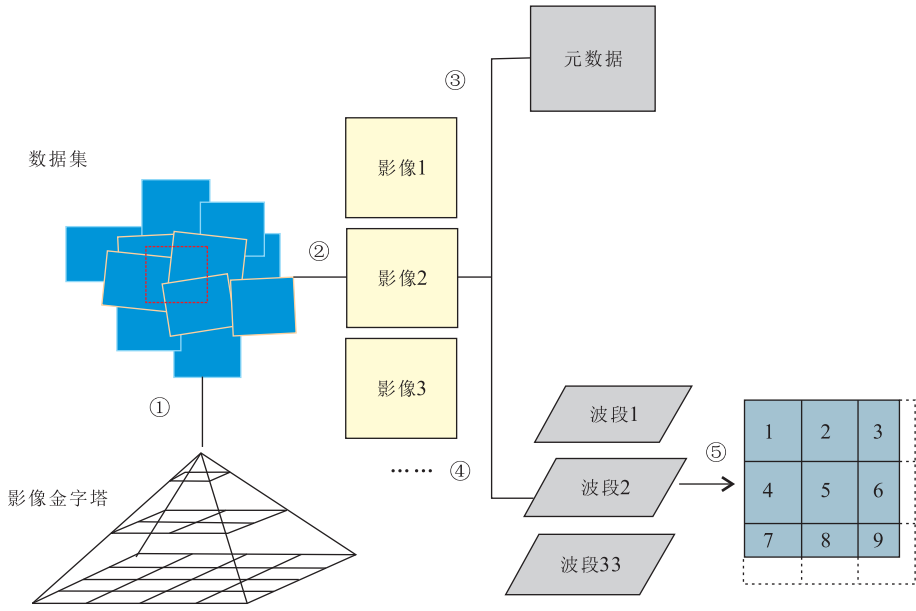


图 2 遥感影像数据模型

Fig.2 Remote sensing image data model

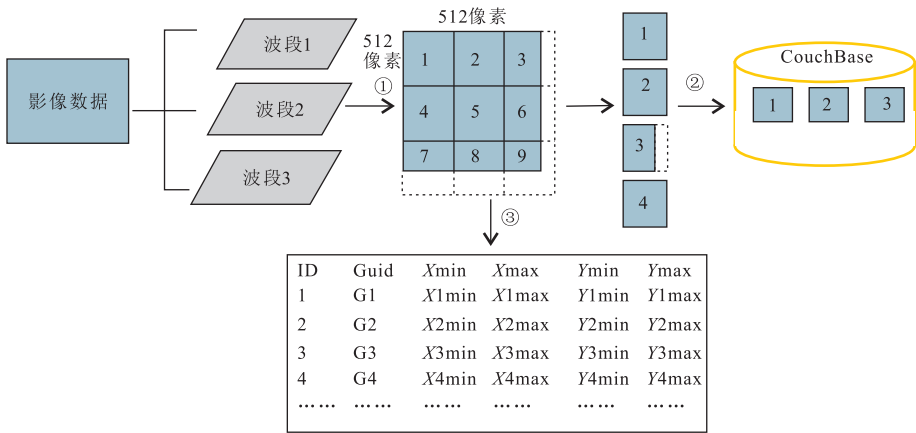


图 3 影像数据分块管理及其存储模型

Fig.3 Image data block management and storage model

①影像分块;②影像切片存储;③切片元数据存储

(3)影像金字塔通过对影像建立不同的分辨率进行预处理和切片,将海量遥感影像按照不同的分辨率进行切分,从而可实现应用层对影像数据的快速浏览和可视化。

1.2 遥感影像存储模型

根据遥感数据的特征以及数据模型设计,数据模型的 3 个部分,需要设计不同的存储模型,分别存储在关系数据库和非关系数据库中,数据的存储模型也分为 3 个部分:元数据部分、影像数据部分和影像金字塔部分。

(1)遥感元数据.影像元数据主要是检索、访问数据库,由于影像元数据的数据量比较小,数据之间

有联系,所以选择关系数据库存储影像元数据.参照相关标准,设计影像元数据的概念模型(图 1)。

(2)影像数据.基于 NoSQL 的遥感影像数据存储模型如图 2 所示.其核心思想是对遥感影像数据进行分块处理,按照规则的分块方案,把遥感影像切分为规则的数据块,然后对数据块进行编码,以编码为 key,以数据块内容为 Value,存入非关系型数据库.影像数据存储模型分为 4 块:影像数据分块、影像数据入库、影像数据出库以及影像数据的空间索引。

(3)影像金字塔.金字塔技术目前已经非常成熟,Google map、Bing map、百度地图和高德地图等

都采用了金字塔技术(谭庆全等,2008)(图 3)。

2 基于 NoSQL 的遥感影像存储中间件

2.1 系统架构

通过设计影像系统存储中间件原型系统,对本文的遥感影像数据存储模型进行验证。

系统的设计要求高内聚、低耦合,从而可以使系统具有良好的可扩展性、可复用性以及可维护性。系统由于属于 IO 密集型系统,并且受网络宽带限制,故在形式上采用客户端和服务端(Client/Server,C/S)架构。大部分工作将在服务端完成,客户端仅作为输入输出端和提供操作指令,可以多个客户端并行使用。系统属于数据业务型系统,从业务逻辑上自底向上划分为软硬件支撑层、数据层、数据服务层、应用层和系统架构如图 4 所示。

其中,影像数据存入 CouchBase 数据库,关系数据和元数据存入 SQLite 数据库。

2.2 SQLite 数据库

SQLite 是一款遵守数据库事务正确执行 4 个基本要素(即原子性、一致性、隔离性和持久性——简称 ACID)的轻型、开源数据库,SQLite 引擎可以嵌入到程序中成为其一部分,主要的通信协议直接通过 API 调用,因此选择 SQLite 做为关系数据和元数据的存储引擎。实际应用中也可以选择其他关系数据库引擎。

2.3 影像入库和出库实现

图 5 展示了影像数据入库流程:用户发出影像数据入库请求,从客户端输入入库的影像数据,客户端可以有多个,针对多个客户端的情况,服务器端会建立线程池,每个上传入库和下载出库任务都会启动独立的线程完成,根据任务的需求做出并发的读写控

制。客户端在收到用户入库请求时,会发送命令到消息队列,服务器端在空闲时读取上传文件命令,会将文件路径传回到客户端,客户端再将文件上传至服务器,生成临时文件。服务器端待文件上传完成后,自动将影像切片,存入服务器磁盘上,切片的过程使用工作线程完成,同时将影像数据切分成边长为 512 像素的正方形图像块,存入 CouchBase 数据库,待所有数据块存储完成时,再将关系数据和元数据存入 SQLite。数据存储完成时,服务器端发送操作完成的消息至客户端。

影像数据出库流程图是影像数据从数据库下载至本地路径的过程。用户向客户端输入需要出库的文件信息,客户端发送命令到消息队列,服务器端将在空闲时读取到文件出库命令,服务器端和客户端建立连接,服务器收到请求后从 SQLite 中查询请求任务的相关信息,包括元数据信息和分块数据信息,根据分块信息从 CouchBase 数据库中取出数据块拼装成临时数据文件,拼装完成后通知客户端下载至本地路径以供用户使用。

3 实验分析

3.1 实验环境

根据以上设计,笔者完成了遥感影像数据模型的原型创建,应用不同大小的影像数据集对网络环境下的遥感影像数据读取效率进行了测试。通过测试传统的关系数据库和非关系数据库在相同的条件下,存取影像数据的效率,验证基于非关系数据库进行海量影像数据处理的可行性。本文搭建的实验环境如下:

(1)硬件环境。数据库服务器端采用 IntelCore(TM)i3-3220 CPU @3.30 GHZ,内存为 4.00 GB,硬盘为 500 GB;客户端采用 IntelCore(TM)2 Duo CPU P9600 @2.66 GHZ,内存为 4.00 GB,硬盘为 300 GB。

(2)软件环境。客户端系统采用 Windows 7,数据库服务器使用 Linux 操作系统;关系数据库使用 Oracle11g数据库;非关系数据库使用 Couchbase2.0 数据库。

本文应用 3 幅分别为 0.6 m 的高分辨率遥感影像数据和 3 幅高光谱遥感影像数据来进行实验,数据集大小分别为 87.5 MB、351.0 MB、879.0 MB 和 1 167.0 MB(1.14 GB),1 310.0 MB(1.28 GB),2 048.0 MB(2.00 GB)。关系数据库选用 Oracle 数据库,非关系数据库选择 CouchBase 数据库,将 6 幅

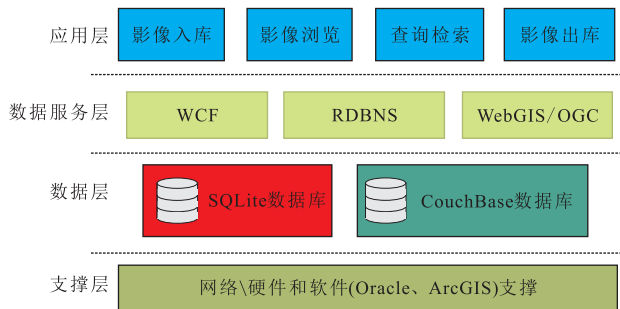


图 4 基于 NoSQL 的遥感影像存储中间件系统总体架构
Fig. 4 The system architecture of storage middleware based on NoSQL for remote sensing image

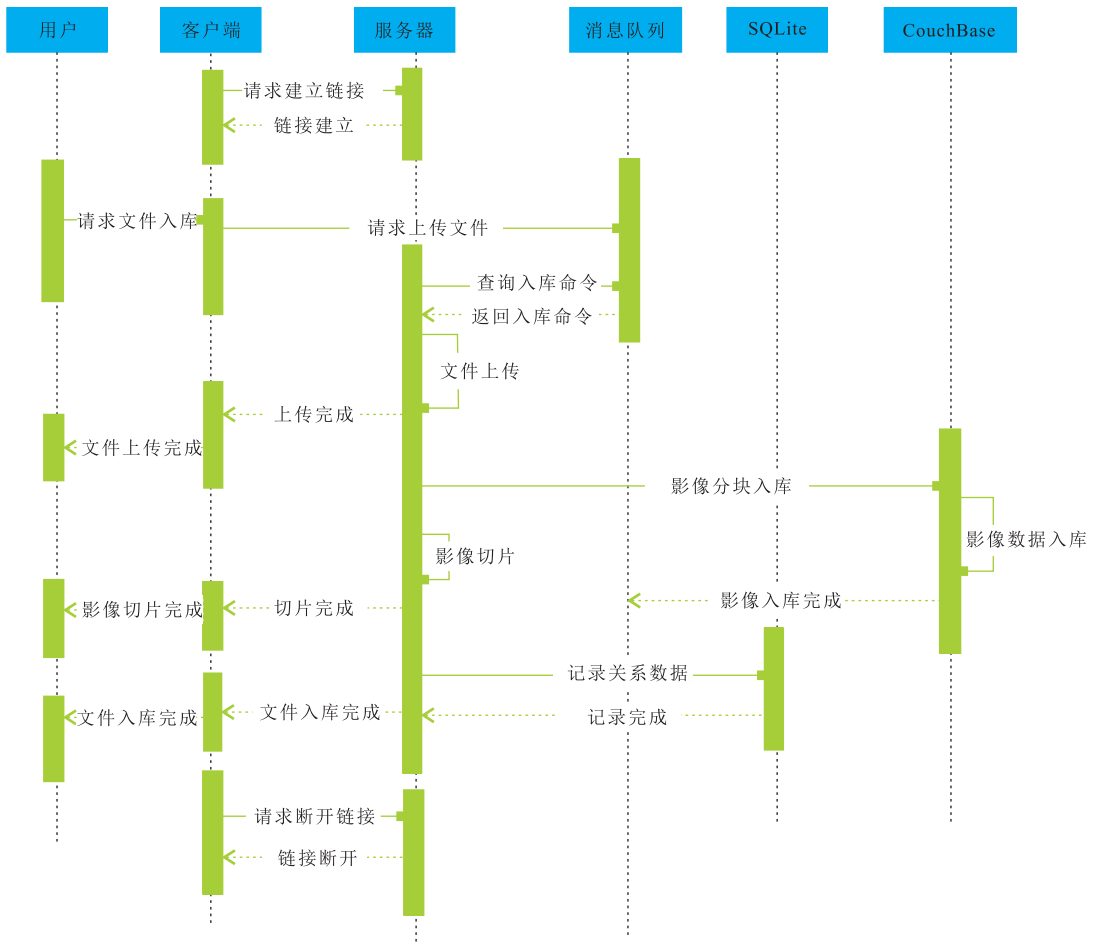


图 5 影像数据入库流程
Fig.5 The flow chart of image data warehousing

表 1 不同大小影像数据集入库/出库平均时间

Table 1 Average time of different size image data collection of inbound/outbound

数据集 (MB)	Couchbase	Oracle
87.5	入库时间: 5.695	入库时间: 17.254
	出库时间: 5.747	出库时间: 14.704
351.0	入库时间: 21.258	入库时间: 61.686
	出库时间: 21.846	出库时间: 65.031
879.0	入库时间: 55.289	入库时间: 187.422
	出库时间: 55.151	出库时间: 186.133
1 167.0	入库时间: 67.600	入库时间: 247.507
	出库时间: 66.910	出库时间: 217.216
1 310.0	入库时间: 80.624	入库时间: 273.341
	出库时间: 83.906	出库时间: 340.847
2 048.0	入库时间: 124.913	入库时间: 529.176
	出库时间: 164.234	出库时间: 614.513

影像数据集分别存入 Oracle 和 CouchBase 数据库, 然后再将数据读出. 由于网络传输的原因, 每次相同的数据入库和出库的时间可能不一样, 本文采取每组数据反复测试 3 次, 时间取平均值的方法进行测试.

3.2 实验结果与分析

表 1 列出了不同大小的数据集分别在 CouchBase 数据库和 Oracle 数据库中入库和出库的时间测试数据. 根据测试数据分别做出二者出库时间和入库时间的对比 (图 6), 直观显示了影像数据存入关系数据库 Oracle 和 CouchBase 数据库效率的差别.

在图 6a 中, 红色代表 Oracle 入库所用时间, 蓝色代表 CouchBase 入库所用时间. 图 6 中可以看出, 本文设计的原型系统入库时间均优于 Oracle 数据库, 且随着数据集的增大, Oracle 数据入库所用时间和 CouchBase 数据入库所用时间之间的差距越来越大. 对于 87.5 MB 的小数据集, 本文设计的原型系统的入库时间是 Oracle 数据库的 33%; 对于 2 GB 的大数据集, 本文设计的原型系统的入库时间是 Oracle 数据库的 23%. 图 6b 中影像数据从数据库读出的效率也是 CouchBase 快于关系数据库 Oracle. 同样, 随着数据集的增大, Oracle 数据出库

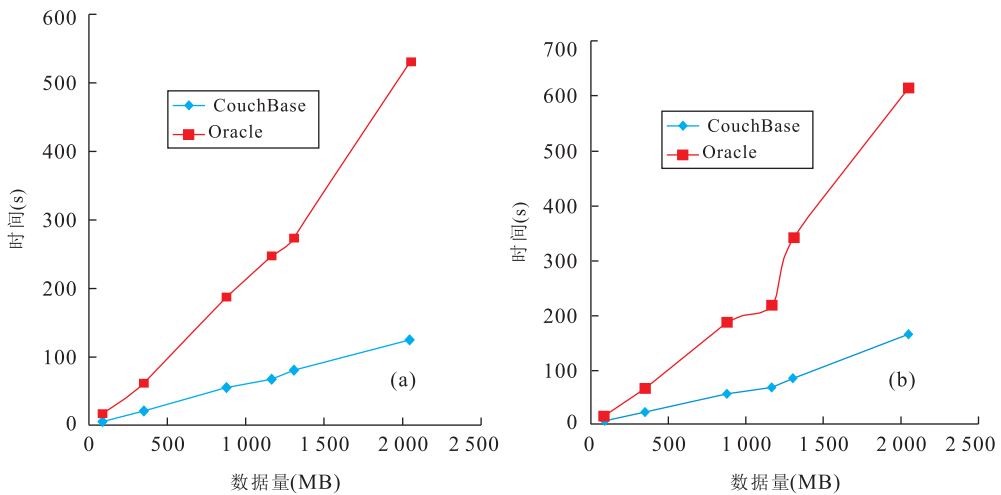


图 6 影像数据入库效率(a)和影像数据出库效率(b)

Fig.6 Efficiency of image data storage (a) and efficiency of image data outbound (b)

所用时间和 CouchBase 数据出库所用时间之间的差距越来越大。

通过实验测试,可以得出以下结论:

(1)本文设计的基于非关系型数据库的原型系统的影像数据入库、出库效率优于基于关系数据库的 Oracle 数据库。

(2)影像数据的数据量不大的时候,关系数据库和非关系数据库对其存取效率相差不大,但是随着数据量的增大,二者的存取数据的时间差也增大。数据量越大,Oracle 所用的时间就越多,曲线的斜率越高于 Couchbase 数据库。

(3)针对高光谱遥感影像,数据模型采用常规的分波段存储方式,波段数目不影响实验结果。

另外,如果随着影像数据量的增大,Oracle 数据库会由于服务器端存储空间及内存不够而不能存储成功,服务器的配置就成了制约系统规模扩展的主要因素,而 Couchbase 数据库由于服务器端采用集群架构,可在任意时间进行水平扩展,增加服务器节点,服务器端自动平衡每个节点上的容量,因此不会出现存储空间不足的问题,所以存储空间影响数据采用非关系数据库是比较理想的方案。

4 结论

本文设计了基于 NoSQL 的遥感影像数据存储模型,以此开发了遥感影像存储中间件原型系统。通过实验测试,笔者认为基于非关系数据库设计的影像数据存储模型及中间件对影像数据读写和提取性能优于传统关系数据库。该研究成果可为基于遥感

技术的矿产资源与能源遥感探测模型研究提供技术支持,满足遥感探测与评价模型对海量影像存储和管理效率的需求,具有重要现实意义和实用价值;同时对我国遥感技术以及基础 GIS 平台软件研究,特别是分布式环境下海量影像的存储管理研究也具有重要理论意义。

References

- Fan, K., 2010. An Overview of NoSQL Database. *Programmer*, (6): 76–78 (in Chinese).
- Fan, X. B., Chen, H., 2006. Key Technology Research of High-volume Image Database Management System. *Computer Engineering and Applications*, (30): 10–12, 37 (in Chinese with English abstract).
- Lai, J. B., Luo, X. L., Yu, T., et al., 2013. Remote Sensing Data Organization Model Based on Cloud Computing. *Computer Science*, 40 (7): 80–83, 115 (in Chinese with English abstract).
- Li, X. W., 2006. Review of the Project of Quantitative Remote Sensing of Major Factors for Spatial-Temporal Heterogeneity on the Land Surface. *Advances in Earth Science*, 21 (8): 771–780 (in Chinese with English abstract).
- Liu, W., Liu, L., Chen, L., et al., 2009. Research on Mass Remote Sensing Image Data Storage Technology. *Computer Engineering*, 35 (5): 236–239 (in Chinese with English abstract).
- Lü, X. F., Cheng, C. Q., Gong, J. Y., et al., 2011. Review of Data Storage and Management Technologies for Massive Remote Sensing Data. *Science China Technological Sciences*, 54 (12): 3220–3232. doi:10.1007/s11431-011-4549-z

- Niu, D. X., Cui, M. M., Huang, C., et al., 2011. Study of the Image Data Storage Technology Based on Oracle Spatial. *Journal of Anhui Agri Sci*, 39(7): 4254—4255, 4261 (in Chinese with English abstract).
- Tan, Q. Q., Bi, J. T., Chi, T. H., et al., 2008. A Flexible and Efficient Algorithm for Constructing Remote Sensing Image Pyramid. *Computer Systems Applications*, (4): 124—127 (in Chinese with English abstract).
- Wang, H. B., Tang, X. M., Li, Q. X., 2008. Research and Implementation of the Massive Remote Sensing Image Storage and Management Technology. *Science of Surveying and Mapping*, 33(6): 156—157 (in Chinese with English abstract).
- Wang, X. D., 2012. Distributed File System Management Technology Research Based on The Massive Remote Sensing Image Data (Dissertation). Lanzhou Jiaotong University, Lanzhou, 19—30 (in Chinese with English abstract).
- Yang, Y. C., Yun, J. M., 2007. Study and Implementation of Image Data Kept on File Management System Based on GIS. *Bulletin of Surveying and Mapping*, (7): 30—34 (in Chinese with English abstract).
- Zhou, J. L., Zhou, Z. D., 2012. Improved Data Distribution Strategy for Cloud Storage System. *Journal of Computer Applications*, 32(2): 309—312 (in Chinese with English abstract).

附中文参考文献

- 范凯, 2010. NoSQL 数据库综述. 程序员, (6): 76—78.
- 樊小泊, 陈红, 2006. 海量影像数据库管理系统关键技术研究. 计算机工程与应用, (30): 10—12, 37.
- 赖积保, 罗晓丽, 于涛, 等, 2013. 一种支持云计算的遥感影像数据组织模型研究. 计算机科学, 40(7): 80—83, 115.
- 李小文, 2006. 地球表面时空多要素的定量遥感项目综述. 地球科学进展, 21(8): 771—780.
- 刘伟, 刘露, 陈萃, 等, 2009. 海量遥感影像数据存储技术研究. 计算机工程, 35(5): 236—239.
- 吕雪峰, 程承旗, 龚健雅, 等, 2011. 海量遥感数据存储管理技术综述. 中国科学: 技术科学, 41(12): 1561—1573.
- 牛得学, 崔苗苗, 黄超, 2011. 基于 Oracle Spatial 的影像数据存储技术研究. 安徽农业科学, 39(7): 4254—4255, 4261.
- 谭庆全, 毕建涛, 池天河, 2008. 一种灵活高效的遥感影像金字塔构建算法. 计算机系统应用, (4): 124—127.
- 王华斌, 唐新明, 李黔湘, 2008. 海量遥感影像数据存储管理技术研究与实现. 测绘科学, 33(6): 156—157.
- 王旭东, 2012. 面向海量遥感影像数据的分布式文件系统管理技术研究(硕士学位论文). 兰州: 兰州交通大学, 19—30.
- 杨永崇, 负建明, 2007. 基于 GIS 的存档影像资料管理系统的研制. 测绘通报, (7): 30—34.
- 周敬利, 周正达, 2012. 改进的云存储系统数据分布策略. 计算机应用, 32(2): 309—312.