

基于SQL/NoSQL的空间目标光学特性 数据混合存储策略*

卢万杰[†] 徐 青 蓝朝桢 吕 亮 施群山

(中国人民解放军战略支援部队信息工程大学地理空间信息学院 郑州 450052)

摘要 为解决海量空间目标光学特性数据入库与检索效率低、非结构化数据存储与管理困难、数据服务扩展性差和无法满足多用户获取实时请求数据的问题,提出一种空间目标光学特性数据混合存储策略.首先,利用光学特性数据中光源与探测角度的规律性,对数据进行处理并构建检索策略;其次,结合空间目标光学特性数据应用领域在数据存储、处理和访问方面的需求特点,构建空间目标光学特性数据混合存储策略,并设计了数据检索体系结构;最后,选取空间目标模拟光学特性数据入库和检索两个场景进行验证.实验结果表明,存储数值数据情况下,混合存储策略数据入库效率较传统关系型数据库提高超过17倍;存储数值数据和图像文件情况下,提高超过34倍;在多用户请求下混合存储策略数据检索效率较传统存储策略有较大提升.提出的混合存储策略能够有效地满足空间目标光学特性相关的仿真、测试与实验对数据的实时请求需求.

关键词 人造卫星, 光学特性, 光度, 数据分析, 数据存储, 数据库

中图分类号: P141; **文献标识码:** A

1 引言

空间目标光学特性数据对国防科技和武器装备建设起着至关重要的作用,是军事信息中的重要和基础资源,能够用于空间目标的状态估计、辅助识别和提前发现^[1-3]等空间态势感知领域,为赢得战争的主动权提供宝贵的预警时间.空间目标的光学特性数据包括不同光源和探测角度组成的相位角序列下的光度数据和光学图像数据等,可以通过光源俯仰角和方位角、探测俯仰角和方位角4个角度的不同序列组合来获取全角度下的光度数据、光学图像文件等信息.为了满足空间目标特性的测试、仿真与实验对数据存储和检索的需求,通常采用数据库对空间目标特性数据进行管理^[4-5].因此,构建满足要求的空间目标光学特性数据库逐渐成为研究的热点.

空间目标的特性数据包括结构化数据和非结构化数据.由于关系型数据库系统(Relational Database Management System, RDBMS)能够高效地处理复杂的数据查询^[6],因此当前对空间目标结构化特性数据的存储方案主要是使用成熟

2019-07-08收到原稿, 2019-08-17收到修改稿

*国家自然科学基金项目(41701463)、河南省科技公关项目(172102210020)资助

[†]lwj285149763@163.com

的RDBMS, 如Oracle^[7-8]、微软的结构化查询语言服务器(Structured Query Language (SQL) Server)^[9]等, 并通过数据库提供的应用程序编程接口(Application Programming Interface, API)查询目标在不同参数下的特性数据. 但是, RDBMS不擅长处理非结构化数据, 因此对于视频、图像、文本等非结构化的特性数据, 通常进行序列化处理后存储于本地文件系统, 导致非结构化特性数据管理不便, 检索困难, 不易维护与RDBMS中结构化数据的关联关系. 针对这种情况, 文献[6]采用RDBMS存储海量的结构化数据和非结构化数据的元数据, 而使用Hadoop分布式文件系统(Hadoop Distributed File System, HDFS)和Apache Hadoop数据库(Hadoop Database, HBase)分别存储不同大小的目标特性非结构化数据文件. 该方案在一定程度上缓解了非结构化数据的存储与管理问题, 但是RDBMS无法有效地维护海量空间目标特性数据和大尺度数据, 尤其是千万级别的数据2维表, 并且RDBMS服务集群的扩展需要更多的数据服务器, 导致系统过于复杂和昂贵^[10-11]. 因此, 使用RDBMS存储海量特性数据时, 面临着数据入库效率低、数据库扩展性差、无法满足云计算和大数据计算环境下多用户实时数据请求等问题^[12].

非关系型数据库(Not Only SQL, NoSQL)的出现则解决了这种难题. NoSQL数据库能够很好地满足大事务量、海量数据低延迟访问和高服务可用性的需求^[13], 并且具有灵活的数据模型, 非常适合存储各种非结构化的数据. NoSQL数据库可以在不影响现有数据结构的情况下进行数据结构的更新. 基于NoSQL的优良特性, 其已经在互联网领域取得了非常成功的应用^[14-15]. 但是, NoSQL数据库不具备复杂的数据查询能力, 其查询方式较为有限.

结合上述分析, 针对现有研究的不足, 为了构建能够提供稳定高效的空間目标光学特性数据服务, 本文尝试对空間目标特性数据中结构化的光度数据和非结构化的图像等数据进行简化和建立索引关系, 并在RDBMS和主流NoSQL数据库特性的基础上构建空間目标光学特性数据混合存储策略; 同时, 通过构建空間目标光学特性数据存储原型系统对模拟光学特性数据的存储与检索效率进行试验验证.

2 空間目标特性数据混合存储策略

针对空間目标光学特性数据存储面临的问题, 充分利用RDBMS和NoSQL数据库的优势, 本文构建了混合数据存储策略, 并以全角度下空間目标模拟光学特性数据为例进行说明.

2.1 特性数据简化与检索策略

通过设置不同的光源俯仰角和方位角、探测俯仰角和方位角, 探测器可以获取光源照射下空間目标的全角度模拟光学特性数据. 其中, 光源俯仰角、光源方位角、探测俯仰角和探测方位角的定义如图1所示, 空間目标本体坐标系以空間目标质心 O 为原点, Z 轴为空間目标质心与地心连线并且背离地心为正, X 轴为空間目标轨道面内与 Z 轴正交的向量并且指向卫星运动方向为正, Y 轴为 X 轴与 Z 轴叉乘获取. 方位角定义为从 X 轴正方向起, 依顺时针方向到空間目标指向光源/探测器的矢量在 XOY 面上的夹角, 范围为 $0^\circ-360^\circ$; 俯仰角定义为空間目标指向光源/探测器矢量与 XOY 面夹角, 在 Z 轴指向一侧为正, 否则为负, 范围为 $-90^\circ-90^\circ$.

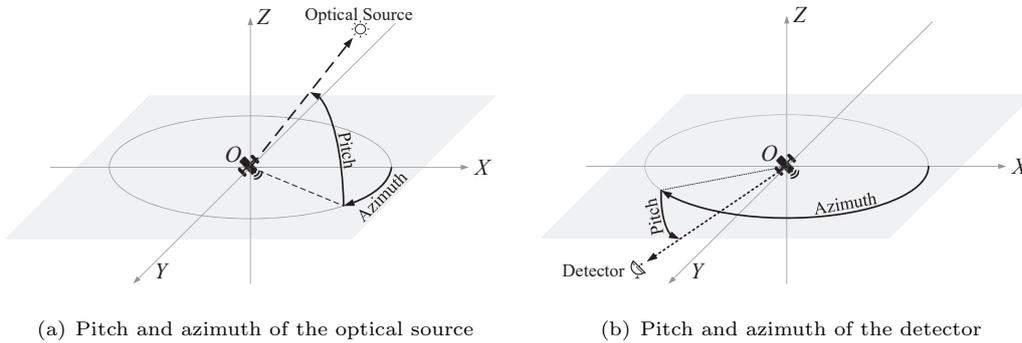


图 1 光源俯仰角和方位角、探测俯仰角和方位角的定义

Fig. 1 Definition of the pitch and azimuth of the optical source and the detector

光源俯仰角和方位角、探测俯仰角和方位角4个角度下的全角度空间目标模拟光学特性数据结构如表1所示. 其中, 4个角度按照探测方位角、探测俯仰角、光源方位角和光源俯仰角的顺序, 以 2° 为间隔, 循环取值. 每组角度数据下均存在空间目标光度值、可见光图像等特性数据. 从表1可以看出, 光源俯仰角和方位角、探测俯仰角和方位角4个角度存在大量重复数据. 当角度间隔为 2° 时, 单个目标的全角度光学特性数据总量超过2.6亿条, 如此庞大的数据集在RDBMS数据库中进行入库和检索会导致系统性能瓶颈, 甚至崩溃. 即使采用分库分表策略进行存储, 数据的重复和冗余也导致耗费空间较多.

针对以上问题, 对空间目标模拟光学特性数据进行简化. 根据4个角度的规律性变化, 构建4个角度之间对应关系的树形结构, 如图2所示. 表1中4个角度之间的关系转变为图2中光源俯仰角与光源方位角、光源方位角与探测俯仰角、探测俯仰角与探测方位角的一对多关系. 简化前角度数据总量 V_{pre} 和简化后角度数据总量 V_{post} 计算公式为:

$$V_{pre} = 90 \times 180 \times 90 \times 180 \times 4 = 1049760000, \quad (1)$$

$$V_{post} = 90 + 90 \times 180 + 90 \times 180 \times 90 + 90 \times 180 \times 90 \times 180 = 263914290. \quad (2)$$

从(1)–(2)式计算可得, 简化的光学特性数据中角度的数据量为简化前的25.14%, 有效地降低了光学特性数据的存储空间. 根据4个角度与特性数据的对应关系构建数据集, 重新构建后的光学特性数据在进行数据检索时, 可以按照光源俯仰角、光源方位角、探测俯仰角、探测方位角的顺序进行数据的检索.

2.2 数据混合存储策略

为实现空间目标模拟光学特性数据的高效存储, 并为相关的测试与实验提供有效的数据服务, 数据混合存储策略主要包括了传统RDBMS、NoSQL数据库和缓存数据库, 具体方案如下:

(1)针对空间目标光学特性数据的元数据, 采用RDBMS中的MySQL数据库进行管理^[16], 并在每条元数据中增加与光学特性数据对应的字段完成数据关系的映射;

表 1 全角度空间目标模拟光学特性数据结构

Table 1 Data structure of the simulated full angle optical property of space object

Optical Source		Detector		Magnitude	Optical Image	...
Pitch/°	Azimuth/°	Pitch/°	Azimuth/°			
-90	0	-90	0	value001	image001	...
-90	0	-90	2	value002	image002	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮
-90	0	-90	360	value180	image180	...
-90	0	-88	0	value181	image181	...
-90	0	-88	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
-90	0	-88	360
⋮	⋮	⋮	⋮	⋮	⋮	⋮
-90	0	90	360
-90	2	-90	0
-90	2	-90	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
-90	2	-90	360
⋮	⋮	⋮	⋮	⋮	⋮	⋮
90	360	90	358	value $m - 1$	image $m - 1$...
90	360	90	360	value m	image m	...

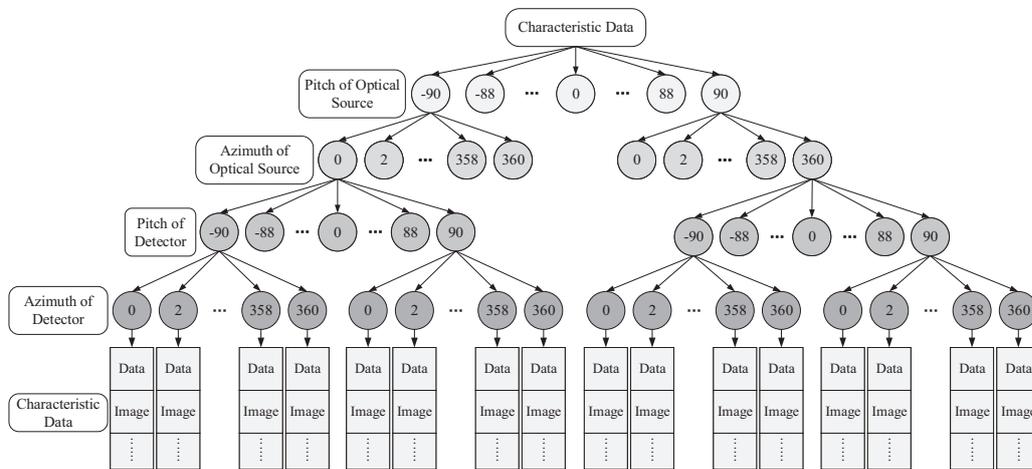


图 2 数据简化与检索策略

Fig. 2 Data simplification and retrieval strategy

(2)简化的光学特性数据结构与JavaScript对象简谱(JavaScript Object Notation, JSON)形式的数据结构非常相似, 因此可采用以2进制格式的JavaScript对象简谱

(Binary JSON, BSON)形式存储非结构化数据的NoSQL数据库—MongoDB进行简化光学特性数据的存储. MongoDB是一种基于文档存储的NoSQL数据库^[17], 在诸多方面的性能均优于RDBMS^[18-20], 适合解决海量大数据访问和存储效率问题. 在MongoDB中, 每个空间目标单独构成一个数据库集合(collection). 为了便于数据存储与检索, 本文使用光源俯仰角对光学特性数据集进行分割, 获取的每份数据组成单独的文档(document)存储于对应目标的集合中, 每个文档中数据结构如图3所示;

```
{
  "OpticalSourcePitch": -90,
  "OpticalSourceAzimuths": [
    {
      "OpticalSourceAzimuth": 0
      "DetectorPitches": [
        {
          "DetectorPitch": -90,
          "DetectorAzimuths": [
            {"DetectorAzimuth": 0, "Magnitude": value 01, "Image": image 01,.....},
            {"DetectorAzimuth": 2, "Magnitude": value 02, "Image": image 02,.....},
            .....
          ]
        },{
          "DetectorPitch": -88,
          "DetectorAzimuths": [
            {"DetectorAzimuth": 0, "Magnitude": value 181, "Image": image 181,.....},
            {"DetectorAzimuth": 2, "Magnitude": value 182, "Image": image 181,.....},
            .....
          ]
        },
        .....
      ]
    },{
      "OpticalSourceAzimuth": 2,
      "DetectorPitches": [
        .....
      ]
    },
    .....
  ]
}
```

图3 Document中的数据存储空间格式

Fig. 3 Data storage format in the document

(3)在对空间目标光学特性进行测试与实验时, 需要对数据库中的数据进行频繁访问, 并希望能够及时获取查询结果. 通过构建缓存数据服务, 存储被频繁请求的数据, 在提供更快的数据获取速度的同时, 也缓解了多用户访问下对数据服务中RDBMS和NoSQL数据库的压力. 远程字典服务(Remote Dictionary Server, Redis)作为典型的内存数据库和缓存数据库, 具有优异的性能^[21], 因此本文选用Redis作为缓存数据库.

空间目标模拟光学特性混合存储策略中, 不同数据之间的映射关系如图4所示. MongoDB中每个目标数据均存储为一个collection, 目标的每条文档均以BSON的形式存储于集合中. MySQL数据库中除目标光学特性元数据外, 通过“Collection Name”字段与MongoDB中对应的集合进行数据的关系映射. Redis中的每组数据为数据请求获取数据的缓存.

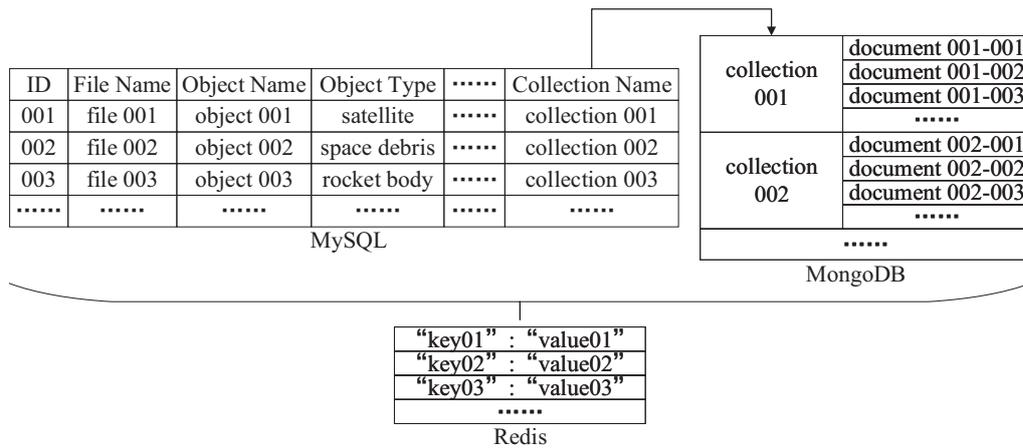


图 4 元数据与特性数据关联关系

Fig. 4 Relationship between the metadata and property data

2.3 数据混合检索结构

混合存储策略下访问特性数据流程如图5所示. 图5中, 数据检索与集成服务提供对数据服务进行统一访问的接口, 用户无需关心存储策略内部逻辑. 外部请求查询数据的流程如下:

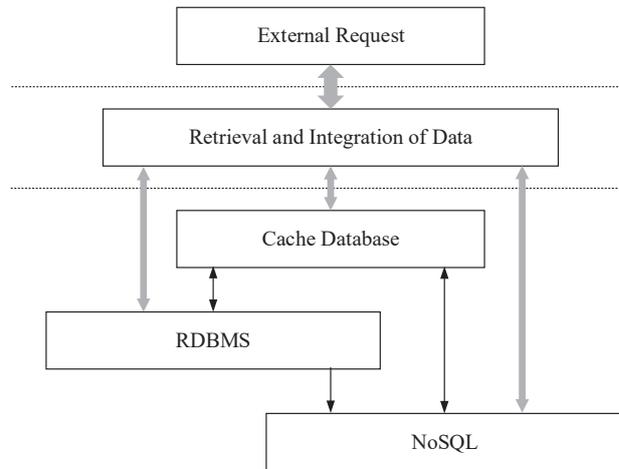


图 5 光学特性数据混合检索结构

Fig. 5 Hybrid retrieval structure of the optical property data

步骤1: 数据检索与集成服务首先通过从缓存数据库中查询是否存在已经缓存的数据, 如果存在, 执行步骤4; 否则, 执行步骤2;

步骤2: 数据检索与集成服务从RDBMS数据库中查询空间目标光学特性数据. 查询结果在返回数据检索与集成服务的同时, 在缓存数据库中进行缓存, 为随后对相同数据的重复使用提供高效的数据服务. 数据检索与集成服务通过对外部请求和从RDBMS中返回数据的分析, 判断是否需从NoSQL数据库中查询非结构化数据. 如果需要, 则执

行步骤3; 否则, 执行步骤4;

步骤3: 数据检索与集成服务访问MongoDB数据库, 获取非结构化的目标光学特性数据. 查询结果在返回数据检索与集成服务的同时, 在缓存数据库中进行缓存, 为随后对相同数据的重复使用提供高效的数据服务. 当数据检索与集成服务获取查询数据后, 执行步骤4;

步骤4: 数据检索与集成服务整合获取的数据, 为外部请求提供格式化后的数据结果.

3 混合存储策略实践

本文选取空间目标模拟光学特性数据入库和检索两个场景对混合存储策略进行测试和验证. 实验测试数据选择光源俯仰角、光源方位角、探测俯仰角和探测方位角的采样间隔为 2° 下的Worldview 1卫星的模拟光学特性数据. 测试数据包括180个光度数据文件和每组角度下对应的光学图像文件. 其中, 光度数据文件数据量为5.58 GB, 包括约2.6亿条光度数据; 光学图像文件数据量为1245.6 GB, 约2.6亿张图像. 实验硬件为一台4核8线程、内存32 G的普通计算机, 操作系统为Ubuntu 16.04长期支持版本, 使用MySQL 8.0.15、MongoDB 4.0.6和Redis 4.0.13数据库构建混合存储策略服务. 为了更好地测试对非结构化数据的存储, 实验将数据分为两种: 第1种为仅数值数据, 即空间目标光学特性光度数据; 第2种为数值数据和图像文件, 即空间目标光学特性光度数据和图像文件.

3.1 数据入库实验与分析

实验通过设置不同的单次入库数据量分别对比了两种数据的入库效率. 采用MySQL数据库时, 测试光源俯仰角、光源方位角、探测俯仰角和探测方位角4个角度的特性数据全量导入; 采用混合存储策略时, 测试光源俯仰角、光源方位角、探测俯仰角和探测方位角4个角度的简化特性数据导入, 对比结果如图6和图7所示. 对比结果表明, 在数据为数值数据的情况下, 混合存储策略的入库效率比MySQL数据库提高超过17倍, 在数据为数值数据和图像文件情况下提高超过34倍. 对比结果说明混合存储策略在面对海量数据入库时具有优异的性能, 尤其是在针对图像等非结构化数据时, 具有比MySQL更好的性能.

3.2 数据检索实验与分析

空间目标模拟光学特性数据存储于数据库中的目的是为特性测试、分析和仿真等应用场景提供服务, 因此需要满足用户对数据的实时访问需求, 尤其是多用户请求下, 能够尽快返回用户请求的数据. 实验使用压力测试软件Apache JMeter设置不同并发访问度来模拟多用户请求, 测试从MySQL数据库和混合存储体系结构中请求数据的响应速度, 对比结果如图8和图9所示. 对比结果表明, 随着并发访问量的上升, MySQL数据库的响应时间线性上升, 本文存储体系结构下混合数据存储策略的响应时间缓慢增长, 并总是低于MySQL数据库的响应时间, 尤其是用户请求数量越大, 本文存储体系结构下混合数据存储策略的优势越明显, 证明了混合数据存储策略在提供数据服务方面的高效性和实时性.

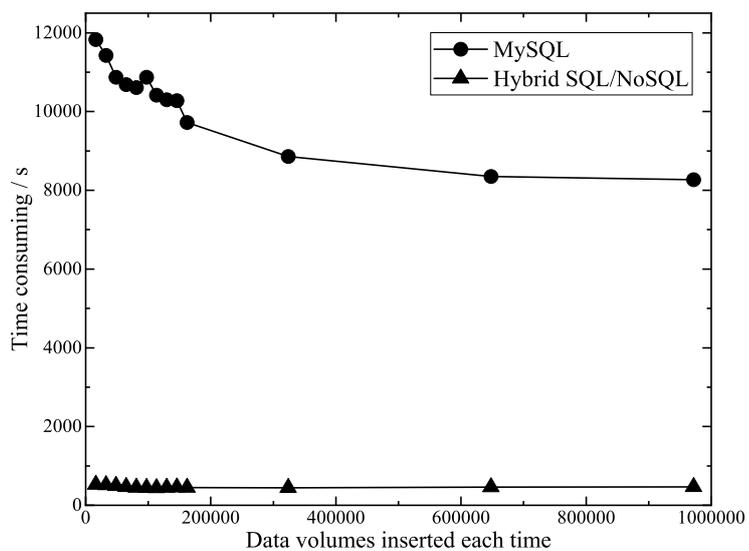


图 6 数值数据入库效率对比

Fig. 6 Comparison of data input efficiency in case of numerical data only

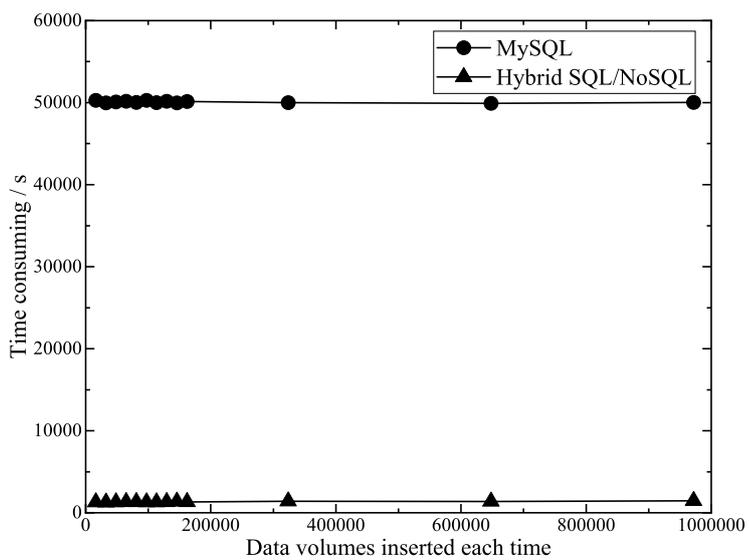


图 7 数值数据和图像文件入库效率对比

Fig. 7 Comparison of data input efficiency in case of numerical data and image files

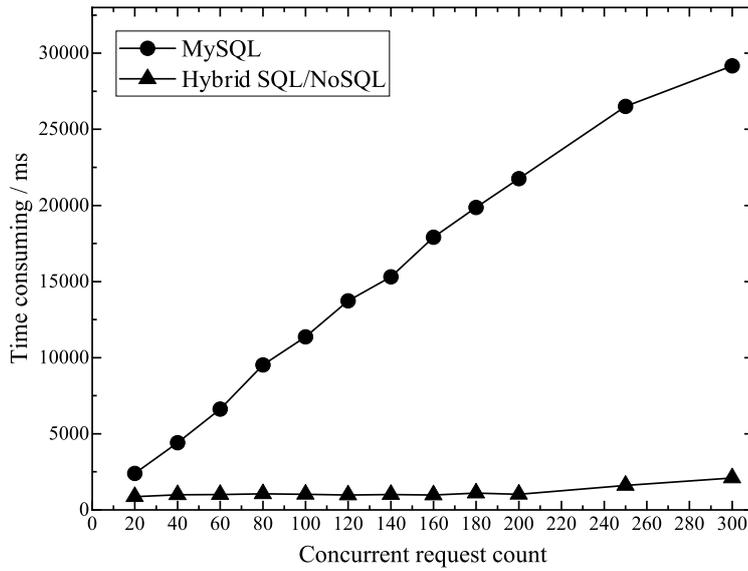


图8 数值数据多用户实时请求响应效率

Fig.8 Comparison of the response efficiency of multi-user real-time requests in case of numerical data only

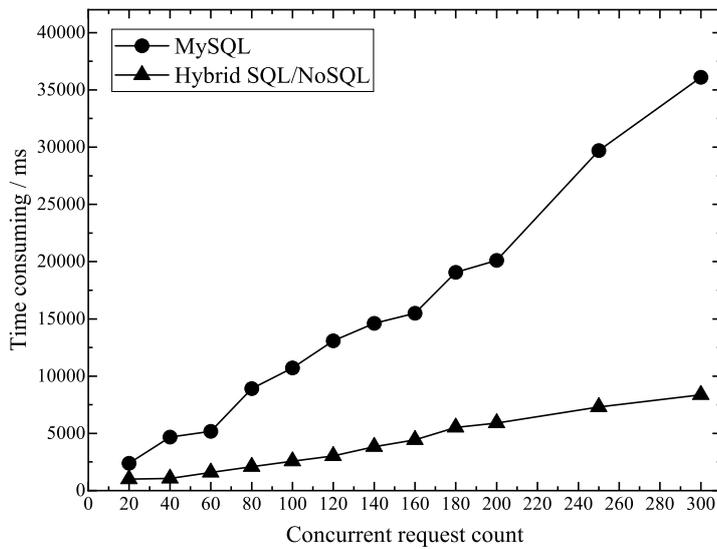


图9 数值数据和图像文件多用户实时请求响应效率

Fig.9 Comparison of the response efficiency of multi-user real-time requests in case of numerical data and image files

4 结论

为解决空间目标光学特性数据存储面临的挑战,提出了空间目标光学特性数据混合存储策略.通过对空间目标光学特性数据中的光源俯仰角、光源方位角、探测俯仰角和探测方位角进行简化,设计了简化后数据的检索策略;在分析关系型数据库和非结构化数据库优缺点的基础上,构建空间目标光学特性数据混合存储策略和空间目标光学特性数据存储原型系统.使用空间目标模拟光学特性数据对传统关系型数据库和混合存储策略下数据的入库和检索效率进行测试.结果表明,仅存储数值数据时,混合存储策略数据入库效率较传统关系型数据库提高超过17倍;存储数值数据和图像文件时,入库效率提高超过34倍.多用户实时请求下混合存储策略数据响应效率较传统关系型数据库有较大提升,尤其是在访问量较大时,优势更加明显,能够迅速响应用户的请求并返回数据.文章提出的空间目标光学特性数据混合存储策略,能够为目标光学特性的管理、测试与仿真实验提供高效和及时的数据服务.

参 考 文 献

- [1] 张逸平, 赵长印, 张晓祥, 等. 天文学报, 2014, 55: 322
- [2] Zhang Y P, Zhao C Y, Zhang X X, et al. ChA&A, 2015, 39: 100
- [3] Xu C, Li Z, Zhang F. SPIE, 2018, 10815: 1081511
- [4] 李冬宁. 空间目标光度测量方法研究. 长春: 中国科学院研究生院(长春光学精密机械与物理研究所), 2015: 1-10
- [5] 徐骏. 激光半主动制导仿真系统中自然环境数据库的建立与应用. 西安: 西安电子科技大学, 2017: 23-57
- [6] Wang Y Q, Jia Y S, Xie X D. 2016 International Conference on Identification, Information and Knowledge in the Internet of Things. Beijing: IEEE, 2016
- [7] 阳召成, 聂镭, 黎湘. 计算机应用研究, 2009, 26: 3843
- [8] Jia Y S, Yang S, Zhong H, et al. Procedia Computer Science, 2019, 147: 542
- [9] 贾雪梅. 导弹卫星目标特性数据库建立与雷达红外目标RCS特性提取. 兰州: 兰州大学, 2008: 4-20
- [10] 李绍俊, 杨海军, 黄耀欢, 等. 武汉大学学报(信息科学版), 2017, 42: 163
- [11] Bjeladinovic S. EntIS, 2018, 12: 1202
- [12] Chang M L E, Chua H N. SQL and NoSQL Database Comparison: From Performance Perspective in Supporting Semi-structured Data. 2018 Future of Information and Communication Conference, Singapore, April 5-6, 2018
- [13] Nance C, Lossner T, Iype R, et al. Proceedings of the 2013 Southern Association for Information Systems Conference. Savannah: AIS Electronic Library, 2013: 111
- [14] Jain V, Upadhyay A. IJCA, 2017, 167: 16
- [15] Pokorny J. International Journal of Web Information Systems, 2013, 9: 69
- [16] Ajayi W. International Journal of Computer Trends and Technology, 2018, 62: 14
- [17] Györödi C, Györödi R, Olah I A, et al. International Journal of Advanced Computer Science and Applications, 2016, 7: 73
- [18] Lawrence R. The 2014 International Conference on Computational Science and Computational Intelligence. Las Vegas: IEEE, 2014
- [19] Kim H J, Ko E J, Jeon Y H, et al. The Journal of Supercomputing, 2018, 74: 1
- [20] Abdullah A, Zhuge Q F. Research Journal of Applied Sciences, Engineering and Technology, 2015, 11: 434
- [21] Spal G K, Kaur J. IJCA, 2018, 180: 26

Hybrid SQL/NoSQL Storage Strategy for Space Object Optical Property Data

LU Wan-jie XU Qing LAN Chao-zhen LYU Liang SHI Qun-shan

(Institute of Geospatial Information, PLA Strategic Support Force Information Engineering University, Zhengzhou 450052)

ABSTRACT In order to solve the problems of low efficiency of data input and retrieval, difficult storage and management of the unstructured data, poor scalability and weak ability to respond to multi-users real-time requests that the massive data of space object optical property faced, a hybrid storage strategy is proposed. Firstly, the data are dealt with and the retrieval strategy is constructed by using the combinatorial sequence regularity of the pitch and azimuth of the optical source and the detector. Secondly, combined with the requirements of data storage, processing and retrieval in the applications of space object optical property data, the efficient hybrid storage strategy is proposed, and storage architecture is designed. Finally, scenarios of simulated data input and retrieval are selected for the experimental verification. The experimental results show that the data input efficiency of the strategy proposed is more than 17 times higher than that of the traditional relational database in the case of storing only numerical data, and more than 34 times separately in the case of storing numerical data and image files. The data retrieval efficiency under multi-user requests is much higher than that of the traditional relational database. The results indicate that the hybrid storage strategy proposed can well meet the real-time data requirements of simulation, testing, and experiment related to the space object optical properties.

Key words artificial satellite, optical property, luminosity, data analysis, data storage, database