

Study on the Organization and Storage of Remote Sensing Data Oriented to Visualization Management

Haijun Yu¹

Key Laboratory of Space Utilization, Chinese Academy of Sciences; Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences

Beijing, 100094, China

University of Chinese Academy of Sciences

Beijing, 100049, China

E-mail: hjyu@csu.ac.cn

Shengyang Li²

Key Laboratory of Space Utilization, Chinese Academy of Sciences; Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences

Beijing, 100094, China

E-mail: shyli@csu.ac.cn

Tao Zhang³

Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences

Beijing, 100094, China

E-mail: zt@csu.ac.cn

Aiming at the management problems of growing multi-source and heterogeneous remote sensing data, this paper studies and realizes the organization and storage method of mass remote sensing data on the basis of analysing the data characteristics. The metadata storage technology is used to realize the unified storage and fast retrieval of mass and multi-resolution remote sensing data information. Through the pre-calculation of the spatial attributes of remote sensing data, the management convenience can be improved. The multi-resolution pyramid model and its storage technology are used to improve the efficiency of remote sensing data reading and visual display, and optimize the operability and visual experience of data management. Finally, the method is applied to the remote sensing data visualization management system based on three-dimensional virtual globe, which to realize the efficient management of multi-load and multi-spectrum remote sensing data as taken by Tiangong-1 aircraft and Tiangong-2 aircraft, and thus satisfy the actual task requirements of the users.

ISCC2017

16-17 December 2017

Guangzhou, China

¹Speaker

²Corresponding Author

³This study is supported by the Prospective Project of Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences (Y7030741WY)

1.Introduction

Remote sensing data as the product of human space application experiment, has become an important data source of earth system science and other disciplines research. Its application can further promote the development of society, economy and other aspects. With the rapid development of ground observation technology and the long-term on-orbit operation of all kinds of flight platforms, the remote sensing data as acquired are gradually showing a series of features such as multi-source, multi-scale, high resolution and global coverage, and the amount of data is also exploded, which makes the orderly and efficient management of high-resolution remote sensing data to becomes one of the important research contents in the field of spatial information science [1].

Modern information technology has been applied to assist the effective management of massive remote sensing data from the research to the application. Many experts have made significant achievements, and developed a large great number of remote sensing data management systems. Here are a few of the more famous: including Earth Observing System Data and Information System of NASA [2], Google Earth of Google [3], ArcGlobe of ESRI [4], Ground European Network for Earth Science Interoperations-Digital Repositories of ESA [5], the online geographic information public service platform-map world of National Administration of Surveying Mapping and Geoinformation of China [6], Satellite data archiving system of National Satellite Meteorological Center of China [7], GeoGlobe of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing of WuHan University [8]. Lu Xuefeng [9] summarized the current mass remote sensing data storage and management technology both at home and abroad. From the aspects of remote sensing data organization and management methods, they compared and analyzed the status and characteristics of the various types of remote sensing data management technology.

By comparing and analysing the existing systems, it is known that the majority are based on two-dimensional graphical interface, and without fully utilizing the spatial information of remote sensing data. They also have some shortcomings, such as simple management method, weak operability and, lack of expressive, etc. With the development of three-dimensional visualization and NoSQL database technology, a more intuitive and easy-to-use method is provided for the efficient organization and storage of mass remote sensing data. Google Earth and World Wind have been in the forefront of exploration and practice, and with good results having been achieved [3,10]. For the actual demand for mass and heterogeneous remote sensing data management, this paper studies the organization and storage method of remote sensing data based on meta-database and distributed NoSQL database, and explains some key technologies as it involved. Finally, the method is applied to the unified management of manned space remote sensing data. In the course of the practical application, the validity of the remote sensing data organization and the storage method have been verified, indicating sound stability and expansibility.

2. Approach Overview

In order to meet the actual needs of remote sensing data management, we use the meta-database and distributed NoSQL database to build remote sensing data organization and storage service, and provide the interface through Web service [11,12]. This approach can not only guarantees the data security, but also improves the efficiency of data management. The meta-database is mainly used to store the metadata information of the remote sensing data and its

spatial attributes. The distributed NoSQL database is mainly used to store the pyramid tile model of the remote sensing data. The data organization and storage technology flow is shown in Fig. 1.

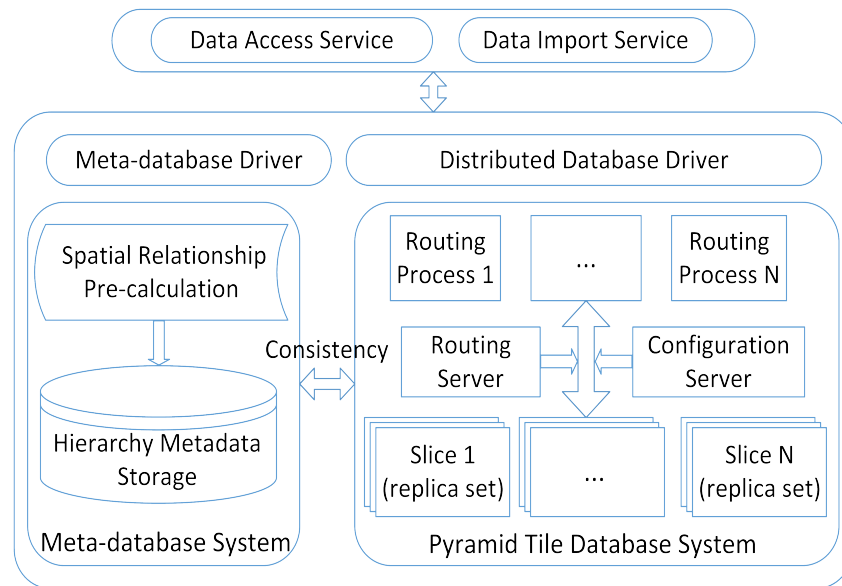


Figure 1: The Technology Flow of Data Organization and Storage

3.Key Technology

3.1Meta-database System Based on Spatial Relationship

The metadata is considered as one of the core technologies for data management as to a data that describes the contents, quality, representation, spatial reference, management style and other features. It is especially suitable for the centralized and unified management of mass and heterogeneous data, which coincides with the management objective of remote sensing data. The objects described by the remote sensing data have spatial distribution characteristics, which express the spatial relationship between objects [13]. The use of its spatial characteristics in the management of remote sensing data can significantly improve the efficiency of query and spatial analysis. In combination with the data characteristics of remote sensing data, this paper designs a remote sensing data metadata system based on their spatial relationship.

3.1.1Meta-database System

The data obtained by different loads have many similarities in terms of the attribute, but there are also differences. The technology flow of the remote sensing data meta-database system designed in this paper is shown in Fig. 2.

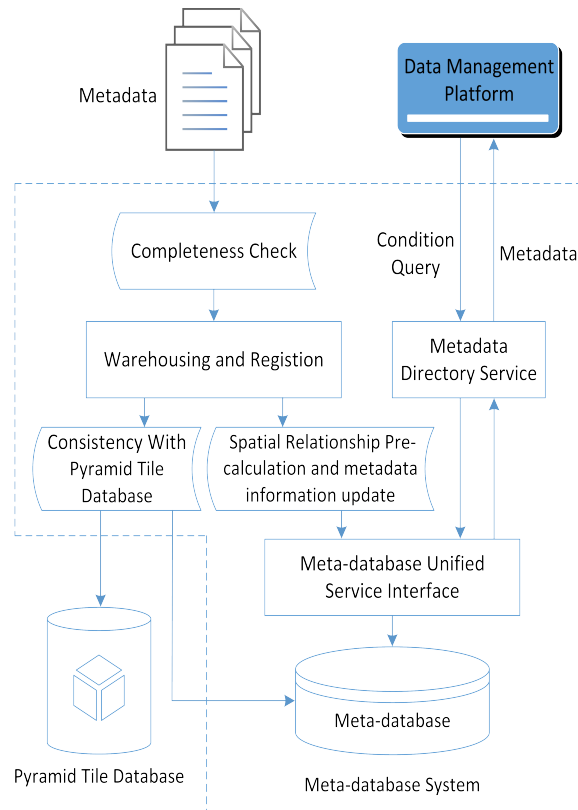


Figure 2: The Technology Flow of Meta-database System based on Spatial Relationship

Through the metadata normalization and warehousing services, we unified and standardize all kinds of heterogeneous metadata information [14], thus to ensure the semantic consistency of data attribute information. In order to fully utilize the spatial attributes of remote sensing data, the necessary processing of metadata information is done by vector calculation. Finally, through the meta-database unified service interface, a unified service is provided for the metadata storage and query.

3.1.2 Structure of Meta-database Table

In the face of the ever-increasing amount of massive and heterogeneous remote sensing data, the design of meta-database tables directly affects the efficiency of data management, especially the user experience of data access [15]. Massive and heterogeneous remote sensing data have a wide range of types varieties, and its metadata attributes are also different. If we create one database table for each subdivision type data, the database is bound to become bloated, and it is not conducive for data query and retrieval. In this paper, we use the method of total-sub-sub table model. First, the load total table is set up according to the load types (such as Hyperspectral imager and Microwave altimeter). We write the common attribute information of the load type into the corresponding load total table. Then, in the same load total table, we establish the spectrum sub-table and write the common attribute information of the spectrum type into the corresponding spectrum sub-table according to the different spectrum types (such as Hyperspectral full spectrum and Hyperspectral thermal infrared spectrum). Finally, in the same spectrum sub-table, according to different data levels (such as L2 and L3) to establish the level sub-table, and write the metadata information of this level into the corresponding level sub-table. We use the keyword index to link the load total-table, spectrum sub-table and level sub-table, to achieve step-by-step access. By using this method, the fusion of heterogeneous attribute information is realized, and to ultimately achieve the purpose of efficient query and retrieval. The structure of meta-database table is shown in Fig. 3.

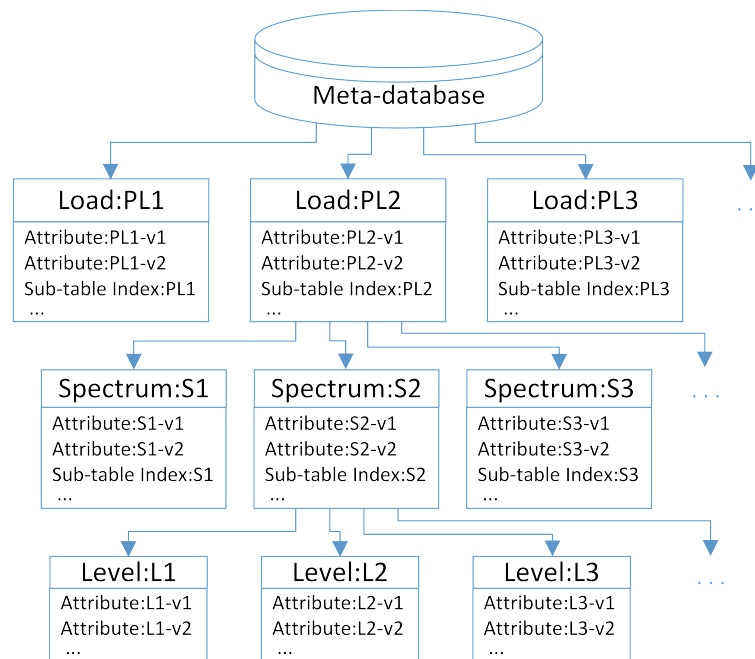


Figure 3: The Structure of Meta-database Table

3.1.3 Pre-calculation Based on Spatial Relationship

The objects described by the remote sensing data have spatial distribution characteristics, such as spatial coordinates and topological information, to express the spatial relationship between objects. In the management of remote sensing data, full utilization of its spatial characteristics can greatly enhance the convenience of data management [16]. Query and statistics based on administrative divisions are frequently used by the filtering methods.

In the vector calculation, the administrative divisions are generally converted into irregular polygons, which brings about a problem. If you want to know all the administrative divisions covered by one data, you need to map the spatial attributes of the data with a large number of administrative divisions, and intersect one by one, which will take much time. If this process is carried out when the user queries the data, it will seriously affect the user's experience. This paper designs the pre-calculation method of remote sensing data coverage administrative divisions. That is, when the metadata is stored, the administrative divisions covered by each data are directly calculated and updated into the metadata information. In this way, each data needs to be calculated once, and subsequent concurrent access requirements can be met fulfilled.

3.2 Multi-resolution Pyramid Model and Its Storage

Practice shows that intuitive visual information is more conducive to people to improve the work efficiency. For the remote sensing data management, the meta-database system is far from meeting the need for the demand of management. In this paper, through the remote sensing data band extraction and conversion, we generate a visual image for each data. At the same time, in order to improve the display efficiency and display accuracy of the view, this paper uses the multi-resolution pyramid technology to divide the visual image into the same size rectangle according to the spatial position grid, called tile [17]. Then, a number of tile layers are created, which with decreasing resolution. These tile layers are stored in a high-performance NoSQL database for access.

3.2.1 Pyramid Tiles Generation

The multi-resolution pyramid model established in this paper adopts the organization of quadruple tree [18]. Assuming that the initial resolution of the acquired remote sensing data is R_0 , the magnification of each layer of the pyramid model is set to 2, then the data resolution of the k layer is $R_k = R_0 * 2^{-k}$. The use of magnification method to build the pyramid model generates a number of tile layers. From the bottom of the pyramid model to the top, the resolution is getting lower and lower, but the data space is unchanged. The pyramid tiles are organized as shown in Fig. 4.

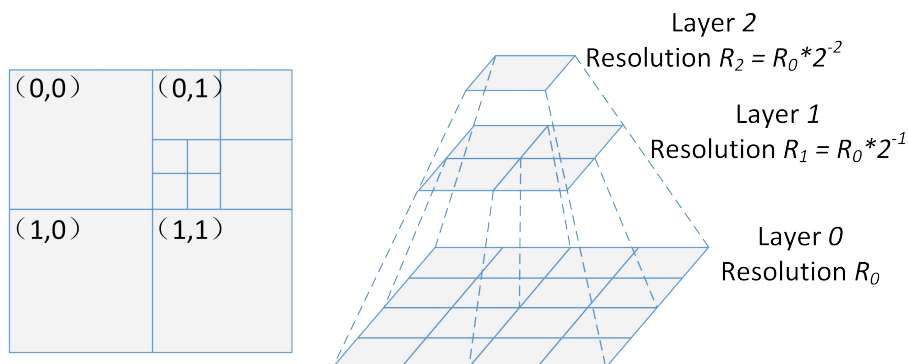


Figure 4: The Organization of The Pyramid Tiles

The pyramid tiles generation process is shown as follows:

- 1) Set the global effective geographic coordinates of the longitude range $[-180^\circ, +180^\circ]$, the effective geographic coordinates of the latitude range $[-90^\circ, +90^\circ]$.
- 2) The multi-resolution pyramid model's layer number starts at 0 and is the highest resolution layer with the most details. Set the layer number is k , the resolution of the $k + 1$ layer is $1/2$ of the k layer.
- 3) Taking into account the transmission rate of the network, the size of each tile data is set to $512px \times 512px$, and the minimum resolution of each block is set to $36^\circ \times 36^\circ$. The data which fails to meet the above rules is filled with a transparent colour.
- 4) The number of tile in each layer consists of the line number and the column number, starting from 0, from left to right and from top to bottom.
- 5) Use the name of the data as the dataset name for each tile. Each tile is solely determined by the dataset name, the layer number, the line number, and the column number.

3.2.2 High Performance Database System for Pyramid Tiles

A remote sensing data may produce hundreds of tiles, and the amount of tiles produced by massive remote sensing data will be countless. How to efficiently and securely store these tiles, and how to quickly and flexibly access the requested tile become a key issue in the visualization management of remote sensing data [19].

MongoDB is a new document-oriented NoSQL database that expresses the "row" of the traditional relational database with a de-structured "Document" model by using BSON (Binary JSON) as the data structure and, using the binary as the storage format. Its data structure is loose, but with high flexibility. It has features the advantages of large storage capacity, fast reading and writing speed, support dynamic query and, support full index, etc. It is easy to expand its storage capacity and maintain the data [20-22]. Therefore, this paper uses MongoDB to store massive pyramid tiles.

Frequent access to mass tiles will take up a lot of computing resources. The increasing amount of data will also exceed the storage capacity of a single server. Data sets that eventually

exceed the memory capacity seriously will affect the I/O performance of the system. The way of increasing CPU and storage resources to extend the database system vertically will be vertically constrained by the operation system and the hardware interface. On the contrary, the horizontal expansion, that is, the fragmentation mechanism, will split and distribute the dataset on multiple servers. Each fragment is a separate database, and all the fragments will be composed to a logically unified database server through the routing process. This approach may reduce the computational pressure on a single server. And the entire fragmented cluster has greatly increased the system's storage capacity and throughput.

In this paper, we design a 12-byte identifier based on time, hardware and random number as the fragment primary key, and the hash function is used as the fragment function. This method can make the data to be distributed randomly and evenly across the slices, as shown in Fig. 5.

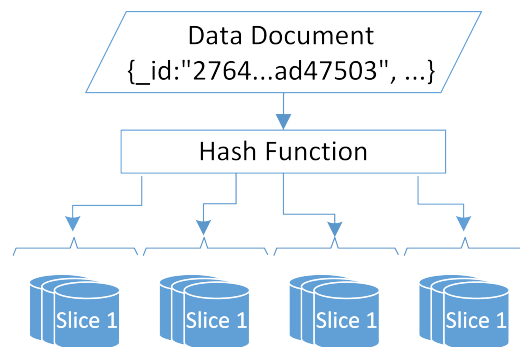


Figure 5: Slice Based on Hash Function

4. Case Studies with Real World Remote Sensing Data

We have verified the effectiveness and usefulness of our method with real world the remote data in fact as taken by Tiangong-1 aircraft and Tiangong-2 aircraft, and designed a visualization management system for the remote sensing data of manned space.

The data centrally organized and stored in the system includes: data from four different aspects spectrums of the hyperspectral imager of Tiangong-1 aircraft, data from three different spectrums of the wide band imager of Tiangong-2 aircraft, data from four different observation modes of the microwave altimeter of Tiangong-2 aircraft. In these data, the highest spatial resolution is 5 meters. The system provides a variety of ways of data retrieval methods, such as according to the specified latitude and longitude, mouse delineation and, administrative divisions, etc. The interface and the result of remote sensing data query are as shown in Fig. 6. The data retrieval results can be displayed in the three-dimensional virtual globe environment at real time and, can be enlarged to view, as shown in Fig. 7. The statistics of administrative divisions covered by the data are as shown in Fig. 8. The system also integrates some of the other vector data, such as 30m resolution elevation data of global land, the name of world countries and regions, the name of China administrative regions and the boundary, etc. These data provide an intuitive and convenient means for understanding the contents and spatial characteristics of remote sensing data. The actual operation of the system realizes the faster data retrieval of massive and heterogeneous remote sensing data, as well as the real-time visual rendering. The response time of each data storage is less than 60s. The query response time of remote sensing data is less than 3s. The frame rate of the three-dimensional visualization rendering is greater than 15fps.



Figure 6: The Interface and the Result of Data Query **Figure 7:** Enlarged Data to View



Figure 8: The Statistics Result

5. Conclusion

This paper analyses the management of massive and heterogeneous remote sensing data, studies the organization and storage method of remote sensing data based on meta-database and distributed NoSQL database, and finally realizes the efficient organization, storage and retrieval of remote sensing data in three-dimensional environment. The practical application shows that this method can realize the organization and storage of massive, multi-source and heterogeneous remote sensing data, and provide fast and safe remote sensing data management. It can well satisfy the actual task requirements.

References

- [1] J. Yang, G. Li, *The China GEO data center: Bringing order to open earth-observation data* [J], IEEE Geoscience and Remote Sensing Magazine, Vol. 5, no. 2, pp. 77-85, 2017.
- [2] M. Esfandiari, H. Ramapriyan, J. Behnke, E. Sofinowski, *Earth observing system (Eos) data and information System (Eosdis) — evolution update and future* [C], Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, pp. 4005-4008, 2007.
- [3] S. Di Martino, S. Bimonte, M. Bertolotto, F. Ferrucci, V. Leano, *Spatial online analytical processing of geographic data through the Google Earth interface* [M], Geocomputation, Sustainability and Environmental Planning, Springer Berlin Heidelberg, pp. 163-182, 2011.
- [4] *A new member of ArcGIS 9: ArcGlobe* [M], ArcGIS World, 2004.
- [5] V. Beruti, M. E. Forcada, M. Albani, E. Conway, D. Giaretta, *ESA plans—a pathfinder for long term data preservation* [C], Proceedings of the 7th International Conference on Preservation of Digital Objects, Vienna, Austria, pp. 53-59, 2010.
- [6] National Administration of Surveying, *Technical guidelines for national geographic information public service platform* [M], Beijing: National Administration of Surveying, 2009.
- [7] S. Jia, J. Yang, J. Shi, *Study and application of parallel-scheduling algorithm in new-generation meteorological satellite data-processing system* [J], Meteorological Science and Technology, Vol. 38, no. 1, pp. 96-101, 2010.

- [8]J. Gong, J. Chen, L. Xiang, Y. Wang, *GeoGlobe: Geo-spatial information sharing platform as open virtual earth* [J], *ActaGeodaetica et CartographicaSinica*, Vol. 39, no. 6, pp. 551-553, 2010.
- [9]X. Lv, C. Cheng, J. Gong, L. Guan, *Review of data storage and management technologies for massive remote sensing data* [J], *Sci China Tech Sci*, Vol. 41, no. 12, pp. 1561-1573, 2011. (in Chinese)
- [10]L. Boschetti, D. Roy, C. Justice, *Using NASA's World Wind virtual globe for interactive internet visualization of the global MODIS burned area product* [J], *International Journal of Remote Sensing*, Vol. 29, no. 11, pp. 3067-3072, 2008.
- [11]Y. Zhu, R. Liu, M. Feng, J. Song, *Research on earth system scientific data sharing platform based on soa* [J], *Journal of Geo-Information Science*, Vol. 11, no. 1, pp. 1-9, 2009.
- [12]F. Wang, *Design and implementation of ground data storage and management system for earth resources satellite* [J], *Spacecraft Engineering*, Vol. 18, no. 3, pp. 66-71, 2009.(in Chinese)
- [13]D. Wicaksono, A. S. Prihatmanto, T. Mardiono, *Design and implementation of GIS data server development for 3D simulation in SAR operation* [C], *System Engineering and Technology (ICSET)*, 2012 International Conference on, pp. 1-6, 2012.
- [14]S. Li, H. Yu, J. Han, B. Hei, *Design and implementation of efficient visualization management system for massive remote sensing images based on three-dimensional globe* [J], *Remote Sensing Technology and Application*, Vol. 31, no.1, pp. 170-176, 2016. (in Chinese)
- [15]F. Wang, X. Wang, W. Cui, X. Xiao, Y. Zhou, J. Li, *Distributed retrieval for massive remote sensing image metadata on spark* [C], *Geoscience and Remote Sensing Symposium (IGARSS)*, 2016 IEEE International, pp. 5909-5912, 2016.
- [16]Y. Zhong, J. Han, T. Zhang, Z. Li, J. Fang, G. Chen, *Towards parallel spatial query processing for big spatial data* [C], *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW)*, 2012 IEEE 26th International, pp. 2085-2094, 2012.
- [17]J. T. Sample, E. Ioup, *Tile-based geospatial information systems: principles and practices* [M], Springer, 2010.
- [18]J. Kang, *Technologies of storage and efficient management on cloud computing for high resolution remote sensing image* [D], Zhejiang University, 2011. (in Chinese).
- [19]J. Liu, X. Fan, J. Tan, *Design and implementation of remote sensing image management system based on high-performance storage* [C], 2012 5th International Congress on Image and Signal Processing (CISP), pp. 1038-1042, 2012.
- [20]K. Chodorow. *MongoDB: the definitive guide* [M], O'Reilly Media, Inc., 2013.
- [21]H. Liang, *Application and research on shardingtechnology in MongoDB database* [J], *Computer Technology and Development*, Vol. 24, no. 7, pp. 60-67, 2014.
- [22]K. B. Sundhara Kumar, Srividya, S. Mohanavalli, *A performance comparison of document oriented NoSQL databases* [C], *Computer, Communication and Signal Processing (ICCCSP)*, 2017 International Conference, pp. 1-6, 2017.