# *Research and Practice of Ocean Data Management Technology with Big Data Philosophy*

**Jinkun Yang[1], Weilu Li[1], Chunhua Han[1], Fangfang Wan[1], Guanghao Wei[1], Dongmei Qi[1]**

[1]*National Marine Data and Information Service Tian, China*

***Keywords:*** Big data, ocean big data, ocean data management, integrative ocean database.

***Abstract:*** Applying big data philosophy and technology to managing marine data resources has become a trend for some time now and in the future. Based on the practice of national marine data resource management for many years and the basic concept of big data, this paper studies the connotation and characteristics of ocean big data, analyzes the opportunities and challenges to be faced by ocean data resource management in the age of big data, and puts forward the classified and hierarchical ocean data resource management system and the concept of integrative ocean database management with "multiple architectures supporting multiple applications" based on the big data philosophy. This paper studies the key technologies such as the construction of ocean data models, the column storage technology of integrative ocean database and the extraction and integration of multi-source ocean data, and carries out the practical application of ocean data management technology based on big data concept and technology, which is of great significance for innovating the ocean data management mode and raising the application value of ocean data

## 1. Introduction

The vast ocean contains rich treasures. The rapid development of information technology drove the rapid accumulation of ocean data. Global ocean programmes have been launched. The Array for Realtime Geostrophic Oceanography (Argo) allows for the first time that information on the physical state of the upper oceans can be systematically measured and collected in almost real time. The international Census of Marine Life (CoML) assesses and interprets the distribution, abundance and diversity of marine life on a global scale [1]. The Ocean Drilling Program (ODP) drills cores and conducts geophysical exploration of different sea areas around the world in an attempt to detect the deep earth and the conversion of matters, investigate lithosphere evolution and seismic formation processes [2]. At the same time, ocean data types is increasing and data volume is exploding, through the implementation of special ocean surveys, polar and ocean scientific investigations, marine remote sensing surveys, as well as ocean economic statistical accounting, integrative sea island management, public opinion analysis at ocean network by internet. The total volume of global ocean data for 2014 is about 25TB, and it is expected to grow to 275PB in 2030 [3]. These ocean data contain rich value and are an important way for human beings to know the ocean, to understand the ocean, and to use the ocean. It is extremely essential to use the advanced information technology to manage ocean data scientifically and effectively so as to fully release the "bonus" of ocean data application.

Ocean data management is a fundamental operational task. Over the years, the concept and technology of ocean data management have been progressing with the rapid development of information technology. Especially with the spurting progress of technologies of information acquisition, cyber-physical system, the Internet, the Internet of Things (IoT), social networks and others, the structured, semi-structured data of text, image, audio, video, etc. emerge in large quantities, and the types, size, storage volume of data are growing rapidly. The world has ushered in the "Big Data Era" [4]. In 2008, Nature published the special issue of "Big Data" [5]. Economist published an article in 2010, "Data, Data Everywhere" [6], and Science launched a special issue in 2011 of “Dealing with Data " [7]. McKinsey Global Institute published the report on "Big Data: The Next Frontier for Innovation, Competition, and Productivity” in 2011 [8]. Twenty experts, including Dr. Divyakant Agrawal of the United States, wrote and published a white paper on" Challenges and Opportunities with Big Data "in 2012 [9]. Starting from 2014, big data has been reported on the work reports of the Chinese government for five times. As a brand-new mode of thoughts, technical means and application methods, big data has reversed the traditional way of data management, which has brought revolutionary changes in data sources, data processing and data thoughts [10, 11].

In the field of scientific research, the application of big data has attracted wide attentions. Data management and application researches based on big data philosophy have been carried out in such scientific fields as remote sensing, geology [12-17], disasters [18], mineral resources, geophysics, and environmental protection [19, 20]. In the ocean field, there are also experts and scholars who have conducted research on the basic concepts and underlying applications of ocean big data centralized on ocean data storage and management [21-28]. Thus, applying big data, cloud computing, virtualization and other modern information technologies to carry out ocean data management, has become the tendency of ocean data management in current and future period.

The first section of this paper addresses the challenges and opportunities for ocean data management in the age of big data. The second section mainly introduces the planning and design of multi-source wide-area, classified and hierarchical ocean data resources based on big data philosophy. The third section addresses the design, research and development of the Ocean Data Management System (integrative ocean database) with "multi-architecture supporting for multiple applications" based on big data technologies. The forth section summarizes the effects of ocean big data management practice and application, and makes suggestions for the future development of ocean big data based on the existing results of research and practice.

## 2. The Age Of Big Data Has Brought New Challenges and Opportunities to Ocean Data Management

### 2.1. The Origin Of Big Data

As a new age of informatization development, big data is the digital portrayal of all things, the core driver of industry and social reformations. With the increasing capacity of human beings for big data acquisition, storage, analysis and application, data has become a basically strategic resource. Now, there is no uniform definition of big data. Michael Cox and David Ellsword, when they presented the term "Big Data" in 1997, stated that data sets are too large for computer memories, local disks and even remote disks to adequately deal with, and visualization of such data is called Big Data [29]. Wikipedia defines big data as “a term refer to data sets that are too large or complex for traditional data-processing application software to adequately deal with” such as capturing data,

data storage, search, sharing, transfer, analysis, visualization, etc. [30]. The Xiangshan Science Conference gives two definitions, technical and non-technical; technically, defined as "big data is a dataset of diverse sources and diverse types, large and complex, potentially valuable, but difficult to process and analyze within a desired time"; non-technically defined as, "big data is a new strategic resource in the age of digitalization, is an essential factor driving innovation, is changing the production- and life-style of human beings" [31].

Although the definition of big data is still not concluded, a consensus on three aspects of big data has been achieved. First, big data involves a huge quantity of data, which cannot be processed, managed, extracted and analyzed by the conventional methods in a reasonable time. Second, big data is a method of using no sample data but all the data for analysis. Third, big data is the massive, high growth-rate and diverse information assets with more excellent decision-making ability, insight capacity and procedure optimization capacity by adopting new processing models. The first emphasizes the processing capacity, the second stresses the application methods, and the third focuses on application value.

Big data has typical "4V" characteristics, namely massive data size (Volume), diverse data types (Variety), fast data flow (Velocity) and huge data value (Value) [32]. With the development of technology, big data currently also presents the characteristics of "5V" and "6V", that is, “Veracity” and/or “Validity” are added to describe it.

#### 2.1.1. The Connotation Of Ocean Data In The Age Of Big Data Is Further Extended, With More Obvious Characteristics

Many scholars have begun to pay attention to and apply the big ocean data. Huang Dongmei et al. summed up the connotation and characteristics of ocean big data, that the ocean big data has the characteristics of multi-source, diversity, massive, hyper- dimension, and sensitivity [33, 27]. The complex of ocean data acquisition, the numerous of ocean data types and the massive of data size make it a typical big data system. [34]. The author thinks that, due to the complex mechanism of ocean itself, the ocean big data has the typical "5H" characteristics (Fig. 1) besides the typical "4V" those of big data: high correlation, high coupling, high diversification, hierarchy, and high regularity. High correlation is embodied that each ocean data has its own clear spatiotemporal attribute, and there are spatiotemporal correlation among several data. High coupling refers to the interaction between the ocean and the land, atmosphere, human activities and so on, and there is a strong coupling between each other. The ocean is changing all the time, a same element in the same place is constantly changing at different times, and multiple elements continue to interact, also produce changes in the environment and phenomena, which refers to the high diversification. Hierarchy refers to the existence of mixed layer, thermocline, acoustic barrier, temperature front, salinity front, vortex, internal wave and other phenomena in the ocean, and all have four-dimensional structural features; reflected in the ocean data, there are continuous variations and intense changes spatially and temporally. In addition, from a macroscopic point of view, marine environment, marine ecology and other systems have monthly, seasonal, yearly, decadal and other periodic variations, and ocean data have a multi-periodically superimposed law, that is, high regularity.

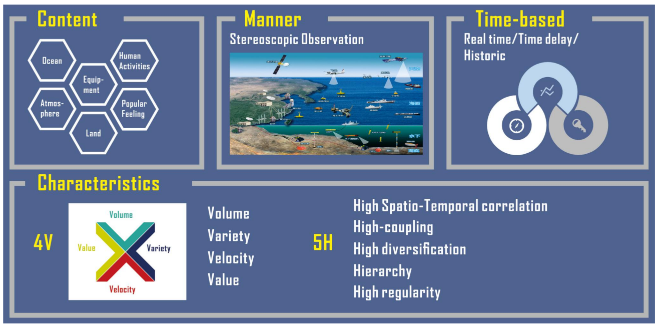


Figure 1: Connotation and characteristic of marine big data.

#### 2.1.2. B. Ocean Data Management Faces New Opportunities And Challenges In The Age Of Big Data

Big data philosophy is a kind of consciousness, a kind of concept. It is an innovative thought. Big data age is the reformation of data, technology and thought caused by data opening [35]. Big Data brings new technologies and solutions to ocean data management and value mining. We are able to obtain the Multi-source, wide-area, multi-time and multi-modal data of ocean development and utilization, equipment, environment, targets and management activities through three-dimensionally spatial, atmospheric, territorial, marine and submarine observation, survey and statistics. How to apply big data philosophy to planning multi-source heterogeneous ocean data resources; how to use data cleaning, conversion, deep fusion, correlation mining, visual analysis and other big data technologies to analyze the potential correlation between various ocean phenomena, ocean phenomena and influencing factors, to reveal the historical laws of the ocean in time and space, and to predict the trend of future variations; how, through the discovered correlations and influencing factors, to further feedback and promote the research and development of physical mechanism of the ocean itself and related scientific issues, and provide quality information services and decision-making support for marine environmental forecasting, marine disaster prevention and mitigation, marine operation and production, economic policy formulation, etc., will all be the main future research and development of ocean big data management and application.

### 2.2. Research On Ocean Data Resource Management With Big Data Philosophy

There is a wide range of operational areas of ocean data sources (in China), including operational observation (monitoring), marine censuses and special ocean investigations, marine scientific expedition, marine environmental protection, marine disaster prevention and mitigation, sea area management, sea island management, marine economy, international cooperation and exchange, etc.. In term of the means of data acquisition, there are marine stations, buoys, shore-based radars, volunteer ships, cross sections, integrative surveys, ocean expeditions, Antarctic and Arctic scientific expeditions, marine satellites and aeronautical remote-sensing. The disciplines of ocean data cover marine hydrology, marine meteorology, marine biology, marine chemistry, marine geology, marine geophysics, ocean acoustics and optics, seafloor topography, marine basic geography and ocean remote sensing, that is, all the ocean-related disciplines. It fully embodies the multi-source, diverse and massive characteristics of ocean data.

Traditional ocean data resources are often confined to an operational field. For example, excellent data management systems have been developed respectively in the fields of ocean observation (monitoring), ocean scientific expedition, sea area and island surveillance and monitoring, marine geographic information [36, 37]. However, there are relative few of research on ocean data resource management system in the national vision of "mount, water, forest, farmland, lake and sea" to carry out. Assembling the practice of national marine data resource management for many years, applying big data philosophy, based on multi-source and heterogeneous ocean data management practice, taking full account of the need for the sharing and application of ocean big data, this paper studies and puts forward a classified and hierarchical national system of ocean data resource management.

### 2.2.1. Classified Ocean Data Management System

Ocean data are classified by different ways from different perspectives. Hou Xueyan et al. divide ocean big data into natural science big data and social sciences big data [38], while Huang Dongmei et al. divide ocean big data into actively generated big data and passively generated ones [33]. These classification methods have their own rationality and advantages, but in the actual work of ocean data management, factors such as data sources, data types, data application requirements need to be comprehensively considered. In particular, different ocean data may differ significantly in their collection methods, which might affect the selection of data and application to a large extent [33]. Through comprehensive reference of the classification methods used in Digital Engineering Data [39], Network Big Data [40], Land Resources [41], marine information classification and coding, marine environment database standards, Integrated Ocean Management Thematic Database Standards (Trial Version), and based on years of national ocean data management practice, ocean data are generally divided into three categories of marine environmental data, basic ocean geographic products and marine thematic information products (Fig. 2). Marine environmental data are then divided into ocean observing data, ocean monitoring data, data obtained via international exchange, polar exploration data, ocean investigation data and special surveys, etc. Basic ocean geographic products are classified as basic ocean geographical products, marine remote sensing information products and seabed topographic products. And marine thematic information products include island management, sea area management, marine economy, ocean policy, marine disaster prevention and mitigation and environmental protection, etc. Under each category, multi-level division is further carried out according to data sources, time, data types, and parameters and so on.

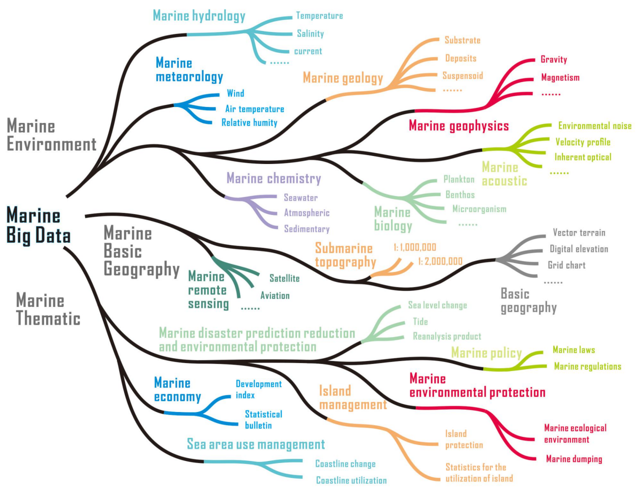


Figure 2: Classified marine data management system.

### 2.2.2. Hierarchical Ocean Data Management System

After the collection of ocean data, data sets with different formats and contents will be formed after different degrees of processing, which have different degrees and values of application. Therefore, the idea of hierarchical management of the raw data layer, basic data layer and integrated data layer is proposed (Fig. 3). The raw data layer consists of received data and raw data. Received data is the most primitive state of data obtained. In the data management of this layer, the original attributes of data are not changed, only sorted by collection time, classified by general types, and then make permanent backup. No other treatment will be done in subsequent processing thereafter. Raw data refers to the data sets formed by decompression, files sorting and other processing of the received data according to data source and discipline, which serve as the data source of data processing and backtracking. The basic data layer refers to all kinds of marine environmental standard data sets, marine geographic and thematic information products and all types of information products obtained by processing of the corresponding raw data sets. The integrated layer refers to data oriented to analysis and application formed by the extraction, transformation, standard processing and quality assessment of various heterogeneous basic data of multi-source. Different levels of data will be managed differently. The classified ocean data management system and hierarchical ocean data management system are combined to form a multi-source and multi-modal ocean big data resource management system, which provides a multi-level data resource basis for subsequent sharing and application of ocean data.

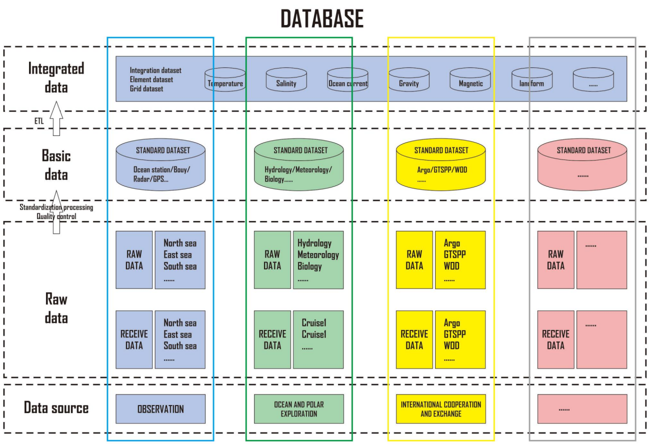


Figure 3: Hierarchical marine data management system.

## 3. Research On Key Technologies Of The Integrative Ocean Database Using Big Data Technology

Scientific research in the age of big data is a process of big science, big demand, big data, big computing and big discovery, and it is of great significance to develop a data management system that supports the whole life cycle of scientific big data [42]. The development of a data management system based on big data technology to support the whole life cycle of ocean data will also become an important technical means of marine data management. Based on the planning and research of data classified and hierarchical ocean big data resources, the key technologies and application practices of Integrative Ocean Database based on big data are explored and implemented.

### 3.1. Overall Design Of The Integrative Ocean Database

At present, the mainstream database technologies based on big data in current market include OldSON, NewSQL and NoSQL, which are applicable to transaction processing applications, data analysis applications and Internet applications, respectively. Among them, the typical representative of the OldSQL applications is transactional database (Oracle), the typical representative of NewSQL applications is large-scale distributed parallel databases (MPP), and the NoSQL applications are typically represented by non-relational databases (Hadoop) [43]. Traditional ocean databases generally use the "an architecture supports multiple applications" model, such as the ocean data management and sharing platform during the "Twelfth Five-Year" period using Oracle database [44]. Along with the rapid growth of ocean data in both types and volumes, this model was no longer able to satisfy the needs of large amount of structured and unstructured ocean data storage, management, analysis, correlative query, real-time processing and construction cost control. To address this prominent contradiction, based on big data technology, the "multiple architectures supporting multiple applications" method, that is, an OldSQL (transactional database) + NewSQL (distributed parallel database) + NoSQL (Hadoop database) mashup model was proposed to build a integrative ocean database [45]. Among them, the OldSQL (transactional database) database is mainly oriented to structured data management, such as temperature and salinity, current, meteorology and other types of marine standard data sets; NewSQL (distributed parallel database) database is mainly aimed at the intelligent mining and analysis of structured ocean data based on big data technologies, such as temperature and salinity element dataset, grid dataset and so on. And NoSQL (Hadoop database) is primarily oriented at the processing management and correlative analysis of unstructured ocean data such as photos and videos [46].

The raw data layer is managed by the file system; the data files of the basic data layer are managed through the data file system, and the database is managed through the transactional database cluster. The integrated data layer realizes the traceability and linkage update of the basic data layer and the raw data layer through database cluster technology. On this basis, combined with data flow management, data content management, portal management and other functions of database supporting management system, the dynamic collection, processing, management, and analysis and sharing service of a variety of marine environment data, marine geographic information and marine thematic information could be realized.

On the whole, the Integrative Ocean Database includes four parts: basic supporting platform, database resource platform, database management platform and service interface. The basic supporting platform consists of compute resource pool, storage resource pool, and network resource pool. The database resource platform is the core of the Integrative Ocean Database. The platform integrates all kinds of data platforms involved in the acquisition, processing, transmission, storage, analysis and application of ocean data, manages and provides data services at the PaaS level. The database management platform mainly realizes the unified management and overall allocation of resources. The database service interface mainly provides service-related interface and general analysis tools for ocean big data, to realize the unified management of the Integrative Ocean Database sharing service (Fig. 4).

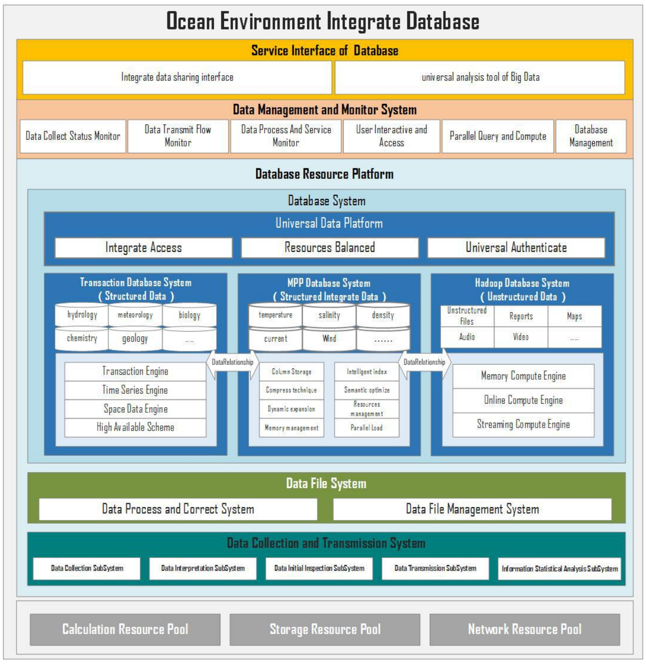


Figure 4: Overall structure of the integrated marine database.

### 3.1.1. Architecture of Ocean Data Model

A reasonable ocean data model can provide important technical support and guarantee for large-scale ocean information decision analysis [47]. Developing a complete set of data object system that can be applied to all fields of ocean operations, highly generalizing the data attribute, data feature and data correlation would, on one hand, provide model guidance for the top-level design of the ocean data management system, and provide a model basis for service-oriented ocean data sharing and exchange on the other hand.

According to the carrier form and processing flow of the raw data layer, the basic data layer and the integrated data layer of the ocean data resource management system, the inventory data object, the metadata object, the raw data object, the basic data object, the product data object and the Spatio-temporal index data object are extracted respectively, and the various ocean data objects are listed according to subjects, elements, topics and other different granularity to build a scalable, highly reusable ocean data domain model [48]. The metadata is summarized and highly abstracted according to the existing metadata industry standards and specifications in the marine field, and metadata models corresponding to the raw data layer, the basic data layer and the integrated data layer are formed. According to the needs of different applications, based on the metadata model constructed, core information is extracted to form the core metadata model based on the built metadata model. Data structure which is generally adapted to various types of ocean data objects are designed and developed by extracting data objects from ocean observation, ocean monitoring, special ocean survey, international exchange and communication, polar and ocean investigation, and various thematic marine fields. By establishing the correlation between ocean data objects, the received data model, original data model, basic data model and application product model are constructed, which further promote the construction of the corresponding database entities based on these models. Among them, the original database entity is constructed according to the original data model, which maps the business logic of the original data collection and management. The basic database entity is built according to the basic data model, mapping the business logic of basic data loading and management. And the application product entity is developed based on application product models, mapping the business logic of application product production and management (Fig. 5).

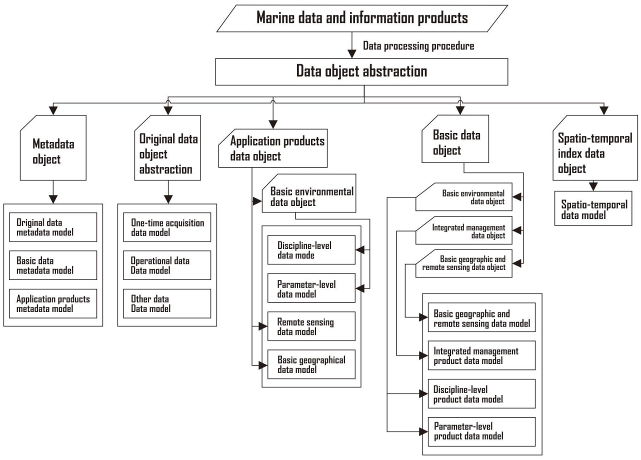


Figure 5: Marine data model construction.

## 3.2. Database Column Storage

The slow reading and analysis speed of massive data in traditional ocean database management has been a key problem in ocean data management for many years. The Integrative Ocean Database adopted the column storage technology, in which the data in the two-dimensional table of the database is stored by column, the invalid column data is not read, which reduces the I/O overhead and greatly improves the data inquiry performance. To further improve the I/O efficiency, data in each column is subdivided into packets, regardless of the size of the single table, the database only needs to operate the relevant packets, so the performance will not decline with the increase of data amount, which greatly improve the data throughput. With the use of column storage technology, it only takes 3 seconds to query the 100 billion records of a single table and its associated tables in a database, while previously, based on transactional databases, billions of record queries took hours, and process crashes were common. At the same time, the data compression ratio can reach up to 20 times more; the space occupied by the data can be reduced to 1/10 of the transactional database, greatly saving the cost of storage devices. In addition, the database design was carried out according to different timeliness and application frequency of various ocean data, the complexity of database table structure, and the degree of correlation between elements. That is, for the more complex structure of the database table, we select the HASH distribution (specified node) mode or random distribution (random node) mode; and for the relatively simple database table, the database is designed by means of replication table, which reduces the consumption caused by data writing and cross-node reading on the premise of ensuring efficient query (Fig. 6). For example, a datum contains the elements of time, temperature, salinity, and air pressure. Time is strongly correlated with other elements, and temperature and salinity are usually collected and used together. Therefore, time, temperature and salinity column group and time and pressure column group are formed.

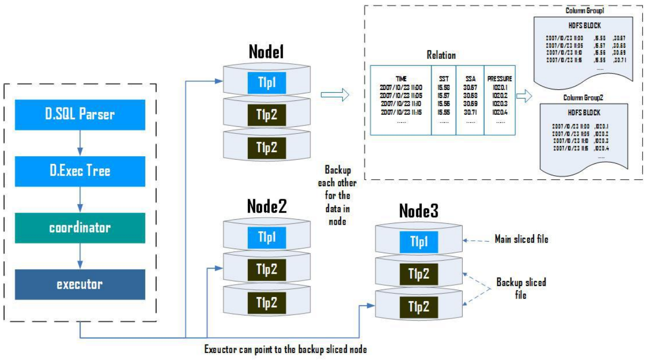


Figure 6: Column storage technology of the integrated marine database.

### 3.2.1. Multi-source Ocean Database Extraction and Integration

High-quality big data integration is an important aspect of big data availability [49]. A prominent feature of ocean data is that data come from multiple business domains. According to the development needs, a number of thematic databases of different ocean fields were constructed at different stages for storage and management of data in fields like marine economy, sea area management, island management, marine ecological protection, marine policy, marine rights and interests, marine forecast and disaster reduction, environmental protection, etc. In retrospect, these database constructed separately using different standards at various levels are no doubt against the application of multi-source ocean big data analysis. So it is urgent to carry out the integrating and dispatching of multi-source ocean database in order to develop collaboratively and efficiently comprehensive marine information support capability.

In the light of the existing marine thematic databases of basic geography and remote sensing, ocean economy, island management and protection, marine ecological protection, marine rights and interests, ocean forecast and disaster reduction, synchronized mirror databases were established to serve as the basic database for further database integration. Then based on the mirror databases of the above themes, the core database table and data content were screened; the data standard, spatial benchmark, unified data dictionary and other processing and repository standardization transformation were carried out to establish various types of marine thematic integrated database, and to extract data from the integrative ocean database using ETL technology by referring to the attributes like data type, discipline, parameter and others.

In order to reduce the performance consumption caused by data writing, data extraction and update are carried out at different times with different rates. Normally, from 6:00 to 21:00 it is real-time extraction according to the data collection frequency; from 21:00 to 6:00 the next day, data is extracted at an interval of one hour. The extraction is carried out by means of full table refresh, timestamp increment, log increment and timestamp comparison. In order to ensure the stability and efficiency of data extraction, Kettle is selected as ETL tool, which could run on Windows, Linux and UNIX, supporting graphical GUI design interface, and then flow in the form of workflows. This method is proved to have a relatively stable performance in either simple or complex data extraction, data cleaning, data conversion, data filtering and so on.

In the comprehensive utilization of multi-source ocean data, duplicates caused by simple combination would inevitably affect the research results. Similar data should be sorted by observation time, receiving time, observation content, data quantity according to the characteristics of data source, observation time, instrument type, data quantity and so on. Based on data sorting results, file rearrangement is carried out for similar real-time and delay-mode data received at different times to avoid duplication caused by multiple compilation and transmission. For the data set itself, the appropriate parameters are selected for the combined conditional duplicates elimination of the data set, and the combination parameters are adjusted for duplicates elimination step-by-step.

### 3.3. Application Practice And Future Development Of Ocean Big Data Management Technology

#### 3.3.1. Application Practice

The ocean data resource management system and the Integrative Ocean Database based on big data concept and technology have been established and run on trial at National Marine Data and Information Service (Fig. 7) for management of multi-source, multi-type data collected from domestic operational ocean observation, monitoring, marine special survey, ocean investigation, polar expedition, international cooperation and exchange, marine geographic information products and marine thematic data products, with the total data volume of hundreds of TB, with the time range from 1662 to date. The problem of database structure redundancy is solved fundamentally. The establishment of industry knowledge model for marine specific knowledge, the optimization of database table structure, and the use of parallel technology and virtualization technology in analytic database make the query of massive data a qualitative leap in timeliness and stability compared with the past databases.

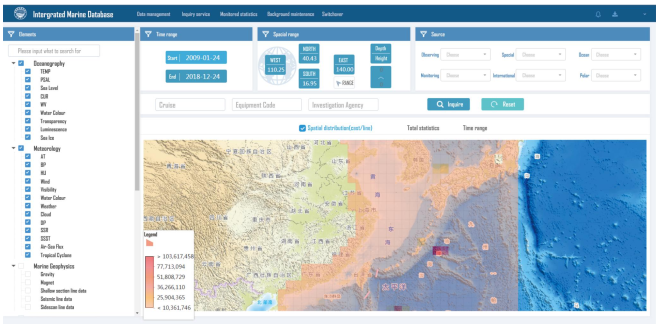

Figure 7: Application of the marine big data management system.

#### 3.3.2. Development Prospect

Using big data concept and technology for ocean data resource management and sharing service will be a general trend of ocean data management in the future. Based on years' practical experience on national marine data management, the author leads his team carrying out the design and application of the multi-source, wide-area ocean data resource management system and the Integrative Ocean Database based on big data technologies, which was only the first step towards ocean data management under big data philosophy. In the next step, further research is also needed in various aspects of ocean big data management standards, ocean big data resource management planning, ocean big data security management, ocean big data governance, ocean big data mining and analysis. By integrating data and information collected in a variety of fields of the space-atmosphere-land-ocean, including environment, basic geography, ocean management, ocean economy, marine equipment, information and documentation, standards and specifications, and policies and regulations, knowledge mining would be carried out for ocean management decision-making, marine environmental awareness raising, marine ecological civilization construction and so on, comprehensively promote the application value of ocean data and information resources.

## 4. Conclusions

Under the thinking pattern of big data, the research on intension and characteristics of ocean data resources, the establishment of classified and hierarchical ocean data management system, the application of composite data management model as columnar storage format, distributed, parallel or virtualization, and overall design of the integrative ocean database, all primarily contribute to solve the problems of mixed management and application, and structure redundancy of database tables for traditional database. After verification, the query aging to retrieve mass data and operational stability of database are promoted greatly, and the management level for ocean data resources is effectively improved.


## Acknowledgments

The writing of this paper benefits from long-term practice and accumulation of operational ocean data management. The author also would like to show his heartfelt gratitude to Dr. Shi Suixiang, Senior Researcher of the National Marine Data and information Service, for his valuable guidance, and to the research team, who devoted themselves to all the related work.


## References


*[1] S. Sun, and X. Sun, Census of Marine Life, Advances in earth science, 2007, 22(10), 1081–1086.*

*[2]Z. Liu, and S. Tuo, Sientific Ocean Drilling Programs: Review and Prospect. Chinese Jourunal of nature. 2007, 29(3), 141–151.*

*[3]H. Wang, N. Liu, R. Pang, and X. Sun, Global ocean forecasting and scientific big data, Chin Sci Bull, 2015, 60, 479–484.*

*[4] H. Guo, L. Wang, F. Chen, and D. Liang, Scientific big data and digital earth, Chin SciBull (Chin ver), 2014 59(12), 1047–1054.*

*[5] S. Nelson, Big data: The Harvard computers, Nature, 2008, 455, 36–37.*

*[6] K. Cukier, Datal, data everywhere: A special report on managing information, Economist, 2010.*

*[7] O.T. Jonathan, G.A. Meehl, S. Bony, and D.R. Eastering, Climate data challenges in the 21st Century, Science, 2011, 331, 700–702.*

*[8] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, H.A. 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute. 1–137.*

*[9] D. Agrawal, P. Bernstein, E. Berntino, S. Davidson, U. Dayal, M. Franklin, ... and J. Widom, Challenges and opportunities with big data: A white paper prepared for the Computing Community Consortium committee of the Computing Research Association, 2012, http://cra.org/ccc/resources/ccc-led-whitepapers/.*

*[10] Z. Yang, J. Tang, P. Zhou, T. Zhang, and X. Jin, Earth science research in U.S Geological Survey under the big data revolution. Geological Bulletin of China, 2013, 32(9), 1337–1343.*

*[11] M. Zhao, and Y. Zhao, Big Data: Data management and data engineering. Beijing: Tsinghua University Press, 2017, 392pp.*

*[12] J. Chen, J. Li, N. Cui, and P. Yu, The construction and application of geological cloud under the big data background. Geological Bulletin of China, 2015, 34(7), 1260–1265.*

*[13] J. Li, J. Chen, and X. Wang, A study of the storage technology of geological big data, Geological Bulletin of China, 2015, 34(8), 1589–1594.*

*[14] J. Miao, W. Shang, Y. Wei, Z. Gao, and Z. Xu, Construction and practice of geological big data management platform based on hybrid architecture, Scientific and technological management of land and resources, 2015, 32(2), 114–119.*

*[15] M. Li, J. Fu, A. Chen, L. Li, and X. Zhou, Preliminary Discussion on the Construction of Geological Data System, Journal of HEBEI Geo University, 2018, 2,16–20.*

*[16] Y. Tan, The study on framework of geological big data construction system, Geological survey of China, 2016, 3(3), 1–4.*

*[17] D. Wang, X. Liu, and L. Liu, Characteristics of big Geodata and its application to study of minerogenetic regularity and minerogenetic series, Mineral Deposits, 2015, 34(6), 1143–1154.*

*[18] H. Fang, Y. Li, X. Yang, C. Yin, and X. Qu, Research on Geo-hazards information management system in China Based on Big Data. Journal of Anhui Agri, Scin, 2014, 42(14), 4478–4482.*

[19] K. Xiao, K. Wang, L. Sun, J. Fan, J. Ding, Mineral resource assessment under the thought of big data, Geological Bulletin of China, 2015, 34(7), 1266–1272.
[20] P. Yu, J. Chen, F. Chai, X. Zheng, M. Yu, and B. Xu, Research on model-driven quantitative prediction and evaluation of mineral resources based on geological big data concept, Geological bulletin of China, 2015, 34(7), 1333–1343.
[21] D. Huang, D. Zhao, L.Wei, Y. Du, and Z. Wang, Managing marine data as big data: uprising challenges and tentative solutions, Computer science, 2016, 43(6), 17–23.
[22] D. Song, K. Cao, J. Zhang, X. Jing, X. Wang, S. Sun, and P. Xie, The conception of marine information construction from the perspective of big data, Ocean Development and Management, 2017, 9, 50–53.
[23] J. Li, Z. Wang, H. Shi, and Q. Zang, Stay on big-data collaborative management system with the emergency events of marine environment. Marine environmental science, 2015, 34(6), 949–953.
[24] J. Li, S. Zhao, and H. Shi, Study on big data sea-mainland collaborative management system with the emergency events of marine environment. Science and technology management research, 2015, 17, 104–108.
[25] C. Yang, Exploring the construction and application of large ocean data platform. China Computer & communication, 2017, 6, 149–151.
[26] C. Sun, Q. Liu, T. Hu, and Z. Guo, Software architecture for oceanographic big data processing, Periodical of ocean university of China, 2015, 45(2), 134–137.
[27] D. Huang, H. Sui, Q. He, D. Zhao, Y. Du, and C. Su, A marine monitoring big data placement strategy in cloud computing environment based on data dependency. Computer engineering & science, 2015, 31(11), 1989–1996.
[28] D. Huang, L. Ji, X. Yuan, and Z. Wang, Marine big data fast platform based on spatio-temporal integration, Marine Environmental Science, 2015, 34(5), 743–748.
[29] M. Cox, and D. Ellsworth, "Application-controlled demand paging for out-of-core visualization," In: Proceedings of Visualization '97. Phoenix AZ: IEEE Computer Society Press, 1997, 235–244.
[30] Wikipedia, Big data, 2015, http://en.Wikipedia.org/wiki/Big data.
[31] The 462nd scientific conference in Xiangshan, 2015, http://xssc.ac.cn/ConfRead.aspx? ItemID=2168.
[32] X. Meng, and X. Ci, Big data management: Concepts, Techniques and challenges, Journal of Computer Research and Development, 2013, 50(1), 146–169.
[33] D. Huang, and G. Zou, Marine big data. Shanghai: Shanghai Scientific & Technical Publishers, 2016, 193pp.
[34] G. Dong, Z. Wang, and Z. Liu, Digital marine system and its security requirements based on big data. Communications Technology, 2015, 48(5), 573–578.
[35] X. Hu, B. Zhang, and D. Li, Overview of big data research and application (Part C), Standard Science, 2013, 11, 29–33.
[36] C. Han, R. Yin, and S. Shi, China's ocean data management system. Marine information, 2012, 2, 1–16.
[37] Z. Xing, G. Xi, R. Zhang, and M. Fan, Design and realization of marine spatial information management and service platform, Science of Surveying and Mapping, 2013, 38(6), 185–187.
[38] X. Hou, Y. Hong, J. Zhang, Y. Zou, X. Shi, L. Ren, X. Cheng, B. Zhang, H. Yu, Z. Guo, and Y. Cui, Marine big data: concept, applications and platform construction, Marine science bulletin, 2017, 36(4), 361–369.
[39] F. Bian, Principles and Methods of Digital Engineering. Beijing: SinoMaps Publishing Press, 2011, 283pp.
[40] Y. Wang, X. Jin, and X. Cheng, Network big data: Present and furture, Chinese journal of computers, 2013, 36(6), 1125–1137.
[41] H. Shi, J. Li, and Y. Feng, Primary study on land and resources data management, Land and Resources in Shandong Province, 2008, 24(3), 45–50.
[42] J. Li, Z. Shen, X. Meng, Scientific Big Data Management: Concepts, Technologies and System, Journal of Computer Research and Development, 2017, 54(2), 235–247.
[43] Z. Lin, Principles and Applications of Big Data Technology, Beijing: Posts & Telecom Press, 2015, 286pp.
[44] C. Han, J. Liang, J. Zhang, R. Yin, Design of ocean data management and sharing platform, JISUANJI YU XIANDAIHUA, 2013, 7,218–221.
[45] S. Wang, H. Wang, X. Tan, X. Zhou, Architecting big data: Challenges, Studies and Forecasts, Chinese Journal of Computers, 2011, 34(10), 1741–1752.
[46] D. Shen, G. Yu, X. Wang, T. Nie, Survey on NoSQL for management of big data. Journal of Software, 2013, 24(8), 1786–1803.
[47] D. Huang, L. Sun, D. Zhao, A composite index strategy for big marine data based on adaptive method of data merging strategy, Journal of University of Science and Technology of China, 2015, 45(10), 813–821.
[48] L. Han, G. Wei, H. Zhang, C. Dong, Study on the Meta object driven data service platform in marine data management, Journal of Ocean Technology, 2017, 36(3), 84–89.
[49] J. Li, X. Liu, An important aspect of big data: data usability. Journal of Computer Research and Development, 2013, 50(6), 1147–1162.