Research on Detection and Analysis of Marine Ecological Environment Based on Big Data

Chen Chen*, Yao Zhang, Xiaoyong Shi
National Marine Hazard Mitigation Service, Beijing 100194, China

*Corresponding Author: chenchen@nmhms.org.cn

Abstract. Facing the integration of multi-source and heterogeneous massive data, traditional integration methods and technologies cannot be realized until the emergence of big data technology. The application of big data technology to realize the integration of multi-source heterogeneous marine environment monitoring data is conducive to the sharing of marine environment monitoring data, avoids the emergence of information islands, and provides the required data for data analysis and mining. Therefore, the thesis is based on the research of the marine environment pollution information intelligent image monitoring technology as the core, and analyses the marine environment pollution information intelligent image monitoring technology, including intelligent digital remote sensing technology, reasonable use of water quality transmitters, comparative analysis of big data, etc. Based on this, a more in-depth analysis of the testing of information intelligent image monitoring technology is carried out.

1. Introduction
Massive marine environmental monitoring data (hereinafter referred to as: monitoring data), coupled with its multi-source, polymorphism, and multi-temporal characteristics, makes the monitoring data a big data. These scattered structured, semi-structured, and unstructured monitoring data contain a lot of information and knowledge. However, the current marine monitoring system integration is generally only system integration, and there are few data integrations, even if there are only With traditional relational data-oriented integration schemes, this approach cannot meet the requirements of multi-source heterogeneous data integration and sharing, which easily leads to the lack of authoritativeness and scientific in the analysis results, and ultimately makes the managers' decisions unreliable. Grid technology has played an important role in achieving a deeper and larger range of resource sharing and eliminating data islands, but it has not put much effort on how to share and use the resources organized into one, and there are shortcomings in ease of use [1]. Cloud computing and big data technology use distributed computing mode to overcome the shortcomings of network computing and provide technical guarantee for massive multi-source heterogeneous data collection, storage, integration, sharing, analysis and decision-making.

2. Marine environment observation data sharing needs
(1) With the increase in the national emphasis on marine scientific research, the National Public Welfare Project and the "863" Program have supported the creation of many marine engineering
projects, which have explored various fields such as marine biology, chemistry, and physics. Generated a lot of data. However, due to the block management mode of China's ocean observation platform, the observation data has distributed and regional characteristics, and many data can only be used in a certain field or a certain subject, and the value of the data cannot be fully tapped. Considering the distributed characteristics of data resources, it is necessary to introduce the concept of big data, establish a data sharing mechanism, realize the reuse of data, and tap the potential value of data.

(2) The different types of data are obtained due to the different methods and expressions of marine environmental observation data. According to the data structure, it is divided into structured data, that is, data stored in the database, which can be logically expressed using a two-dimensional table structure; semi-structured data, that is, the format is more standardized, and each item of data can be parsed in a certain way; no Structured unstructured data refers to non-plain text data. There is no standard format, and the corresponding values cannot be directly parsed, such as images and videos. Based on the heterogeneous characteristics of marine environmental observation data, it is necessary to learn from the data storage concept of big data to integrate heterogeneous data and re-plan and deploy its storage.

(3) Under the rapid development of sensor technology, ocean observation technology, and communication technology, the frequency of acquisition of marine environmental observation data is constantly increasing. How to quickly complete data update, analysis, and provide corresponding applications and products is to realize data Share important issues that need to be resolved. In data storage and calculation, the model of distributed storage + computing platform of big data can be adopted to meet the rapid flow of observation data [2].

3. Database design of marine ecological environment pollution

HBase stores data in the form of a table, and the data is stored in a table in the form of key-value pairs of Key-Value. Key is a row key is an identifier of a row of information. Value is specific data information. According to GBT18208.4-2011, the formula for calculating the direct loss of earthquake marine ecological environment damage is:

\[
L = \sum_{i=1}^{m} \sum_{s=1}^{n} \sum_{j=1}^{5} \left( S_{is} \times R_{isj} \times D_{isj} \times P_{is} \right)
\] (1)

In the formula: \(S_{is}\) represents the total area of \(s\) type marine ecological environment in the \(i\) evaluation sub-region, the unit is square meters (m²); \(R_{isj}\) represents the destruction ratio of \(j\) level destruction of the \(s\)-type marine ecological environment in the \(i\) evaluation sub-region; \(D_{isj}\) represents the \(i\)-evaluation sub-region Loss ratio of \(j\) level destruction of \(Fs\)-type marine ecological environment; \(P_{is}\) represents the replacement unit price of \(s\) type marine ecological environment in the sub-area of \(i\) assessment, the unit is yuan per square meter.

3.1. Cloud computing model design

MapReduce is a programming model. Programs written using the MapReduce model can run on thousands of computers at the same time. As can be seen from the following MapReduce working principle diagram, it mainly contains two core function abstract classes Mapper and Reducer. Large files will be divided into multiple data blocks and then processed at different nodes. The Map module of each node converts the input data into Key-Value pairs and outputs them. These key-value pairs are then converted by Obtained by the Reduce module, the Reduce module will merge Value values with the same Key value and finally get the result. Figure 1 describes the working principle of the MapReduce programming model. Through the above steps, the computing task originally on a
computer is distributed to multiple computing nodes to cooperate together, which greatly improves the speed of computing. This is the advantage of cloud computing different from traditional computing.

Figure 1. MapReduce working principle diagram

3.2. Program flow design
From the calculation formula, the direct loss of earthquake house damage can be easily related to the MapReduce calculation model. In the formula, it is necessary to calculate the earthquake damage loss of a single marine ecological environment, and then add up one by one to get the total. Related calculation data is pre-stored in the HBase database table. Here, the calculation process can be divided into two MapReduce tasks. The first MapReduce task is used to calculate the individual seismic damage losses of all marine ecological environments, and the second MapReduce task is calculated.

4. Selection of integrated framework, model and data for marine environment monitoring data

4.1. Integration of related frameworks and models
The marine environment monitoring system mainly includes marine ecological environment supervision and management system, aviation remote sensing monitoring system, satellite remote sensing monitoring system, underwater unmanned automatic monitoring station monitoring system, ecological buoy monitoring system, marine environment routine monitoring business system, marine conventional hydrometeorological observation Business systems, etc. Due to the different spatial and temporal distribution, monitoring frequency, data format, spatial observation scale, and data complexity of these data, the integration of marine environmental monitoring data becomes difficult to achieve. This is also the reason why most of the current ocean monitoring system integration is the integration of front-end systems, and there is very little integration on comprehensive ocean observation data [3].

Ontology semantic framework can solve the differences in concepts and terminology in the field of marine monitoring, so that different users can communicate and communicate smoothly and maintain semantic equivalence, while allowing different tools, software and application systems to achieve mutual Operation is helpful for the construction of integrated systems. The MMI ORR framework is shown in Figure 2. The DIF of MMI ORR and IOOS can be used as a reference framework for the integration of marine environmental monitoring data.
4.2. Framework design of data sharing platform

According to the requirements of ocean environment observation data sharing, combined with the concept and technology of big data booming in recent years, this study proposes a framework of ocean environment observation data sharing platform based on the concept of big data, as shown in Figure 3.

Figure 3. The framework of the marine environment observation data sharing platform

The framework is mainly composed of data sources, data operations, data storage, data access, data analysis, application services, and related data sharing management systems and security mechanisms. Each layer is relatively independent, and the coupling between layers is relatively small. The data source at the bottom of the framework mainly includes basic data that needs to be shared, various historical data, real-time data and delayed data obtained by various projects and regions, and information products obtained and produced, etc., as the basis for the operation of the entire platform. Data operation is to extract, sort / filter and transform heterogeneous multi-source data to realize data re-integration and deployment, and complete the standardized storage of data.
5. National Marine Innovation Index Forecast Technology Ideas

On the basis of an in-depth analysis of the development status of China's marine innovation, determine the marine innovation indicators to be predicted, collect and organize marine innovation data, construct Big Data distributed data storage system + MapReduce technology big data architecture, the prediction method of marine innovation indicators and its Application for research (Figure 4).

![Figure 4. Evaluation route of marine ecological environment detection technical indicators](image)

The technical route can be divided into four stages: 1. to conduct research on the status and problems of China's marine innovation development, define the national marine innovation system, and determine the marine innovation indicators to be predicted; 2. Collect the data required for marine innovation indicators, improve and perfect the existing data 3. Organize and pre-process the data to establish Bigtable distributed data storage system; 4. Comprehensively apply suitable algorithms to map and redefine marine innovation data to form a perfect marine innovation index prediction method.

System [4].

The core of the technical route is in the research part of the marine innovation index prediction method based on the idea of big data, specifically divided into two aspects: 1. Data collection and collation and Bigtable database construction. Collect the data required for marine innovation indicators, sort and pre-process the data, establish a Bigtable distributed data storage system, and store and manage the marine innovation data; 2. Research on the index prediction method based on the Mapreduce computing framework. Aiming at various types of unstructured, non-linear and non-causal data, based on the MapReduce calculation framework, a systematic marine innovation index prediction method is formed, which includes Map processing and Reduce processing. The maximum height of the decision tree is set to limit the growth of the tree or the minimum number of records that each node must contain is used to improve the decision tree algorithm. The improved decision tree algorithm is used to map the marine innovation data. According to the different types of data, BP
neural network and different types of clustering algorithms are used comprehensively to process the marine innovation data, which lays the foundation for the prediction of marine innovation indicators.

The realization of the best method for forecasting marine innovation indicators under Big Table + MapReduce is a difficult research. The specific performance is as follows: How to use the Bigtable distributed data storage system to effectively manage massive multi-source marine innovation data based on the idea of big data? How to use the appropriate algorithm to map and process marine innovation data under the MapReduce computing framework, to realize and find the optimal algorithm? If it can be achieved, it will provide comprehensive and accurate forecast information for the formulation of the marine innovation master plan and the formulation of marine science and technology policies, and realize the quantitative forecast of the development trend of China's marine innovation and the future marine innovation war, effective layout with a little emphasis [5].

6. Conclusion
According to the current status of marine environmental pollution monitoring technology, based on the background of scientific development, the intelligent image monitoring technology used is mainly composed of intelligent digital remote sensing technology and water quality sensor technology, and the pollution data for the marine environment is collected. You can also use data comparison and analysis to calculate the data, and use intelligent image monitoring technology to monitor the pollution in the marine environment, so as to pass the test and simulation experiments of intelligent image monitoring technology such as monitoring resolution test and monitoring timeliness test. To expand the scope of application of information intelligent image monitoring technology.

References
[1] Qu, C. F. Song, J. M., & Li, N. The effects of jellyfish decomposition on marine ecological environment. Acta Ecologica Sinica, 35(18) (2015) 6224-6232.
[2] Zong, L. Du, W., & Bai, Y. An enhanced transmission control protocol scheme on the marine ecological environment sensor monitoring. Sensor Letters, 13(2) (2015) 152-155.
[3] Ashu Khosla. Palaeoenvironmental, palaeoecological and palaeobiogeographical implications of mixed fresh water and brackish marine assemblages from the cretaceous-palaeogene deccan intertrappean beds at jhilmili, chhindwara district, central india. revista mexicana de ciencias geológicas, 32(2) (2015) 344-357.
[4] Lee, Y. M. Cho, K. H., Hwang, K., Kim, E. H., & Lee, H. K. Succession of bacterial community structure during the early stage of biofilm development in the antarctic marine environment. Korean Journal of Microbiology, 52(1) (2016) 49-58.
[5] Luo, Y. Ni, W., Jiao, L., Huang, H., & Ye, Y. Water quality comparison and ecological environment assessment near sea area of fujian province. Journal of Computational & Theoretical Nanoscience, 13(4) (2016) 2306-2314.