Design and optimization of distributed cloud storage in marine IoT

Rongming Tian
Chongqing Chemical Industry Vocational College, Chongqing 401228, China
tian_rm@sina.com

Abstract. The traditional cloud storage technology has the defect of little consolidation of actual storage files, due to lack of storage strategy. This paper designs and optimizes the distributed cloud storage technology in the marine IoT. Firstly, we use distributed computing to calculate the node data in the Internet of Things. Then we build a cloud storage framework based on the characteristics of the Internet of Things nodes. Then establish a direct mapping between files and storage strategies in the Internet of Things. Finally, according to the calculated node performance data, a cloud storage strategy is formulated to complete the research on the distributed cloud storage technology in the marine Internet of Things. The experimental results prove that, compared with the traditional cloud storage technology, the distributed cloud storage technology in the marine Internet of Things can finally merge more files when storing files, reducing the storage amount occupied by actual files. It is suitable for practical use.

Keywords: Internet of things, distributed, cloud storage, storage strategy.

1. Introduction
The Internet of Things refers to the collection of any objects or processes that need to be monitored, connected, and interacted in real time through various devices and technologies such as various information sensors, radio frequency identification technology, global positioning technology, etc. [1]. Collect all kinds of required information such as sound, light, location, and access through various possible networks to realize the ubiquitous connection of objects and objects, objects and people, and realize the intelligent perception, identification and management of objects and processes. The actual Internet of Things is an information carrier based on the Internet, traditional telecommunications networks, etc., which allows all common physical objects that can be independently addressed to form an interconnected network. Cloud storage is a model of online storage. Data is stored on multiple virtual servers hosted by a third party, which is convenient for users to use storage resources to store files. Distributed computation can refine a problem that requires a very large amount of computing power to solve into many small parts, and then distribute the small parts to multiple computers for processing, and finally combine the results of multiple computers as the final result. Study the distributed cloud storage technology in the marine Internet of Things, which can process marine files in a distributed manner and reduce the memory occupied by ship files.
2. Key technologies of distributed cloud storage in marine IoT

2.1. Distributed node computing

Distributed node calculation is mainly aimed at the hardware parameters in the marine Internet of Things. The calculation parameters mainly include machine word length, memory size and storage capacity, etc., so when using distributed node calculation, the calculation formula of the storage performance of task-type data is set for:

\[ P(i) = \beta_1 F_i + \beta_2 W_i + \beta_3 M_i + \beta_4 S_i + \beta_5 V_i \]  \hspace{1cm} (1)

Among them, \( F \) is the CPU main frequency, \( W \) is the word length, \( M \) is the memory of the virtual machine, \( S \) is the external storage capacity, \( V \) is the disk read and write speed, \( \beta_i \sim \beta_5 \) is the weight coefficient of each parameter. Standardize the storage performance of each node and define the storage performance of relative nodes. \( P(i)_{\min} \) is the minimum storage performance of all nodes in the cluster. At this time, the ratio of the storage performance of other nodes to it is the relative storage performance of the node \( P(i)_2 \):

\[ P(i)_2 = \frac{P(i)}{P(i)_{\min}} \]  \hspace{1cm} (2)

When the job in the Internet of Things is actually running, the storage file records the task start time, which is the time required for each node in the Internet of Things to run the task. The calculation formula is:

\[ T_i = \frac{T}{B}, 1 \leq i \leq n \]  \hspace{1cm} (3)

In the formula (3), \( B \) represents the number of data blocks processed by node \( i \), and \( T \) represents the time of data blocks processed by node \( i \). The final calculated node performance result of formula (3) is relatively uncertain, so when distributed computing nodes are combined with the above three formula, the calculation is:

\[ P(N_j) = \frac{E_{\text{total}} t_i}{m * S} \]  \hspace{1cm} (4)

In the formula (4), \( N_j \) represents the node to be calculated, \( m \) represents the number of tasks completed by the IoT node \( N_j \) in time \( t \), \( t_i \) represents the task \( i \) execution time, \( S \) is a constant representing the size of the data block transmitted in the IoT. After determining the node data in the distributed computing IoT, build a marine cloud storage framework, formulate a cloud storage strategy, and complete the distributed cloud storage technology research in the marine IoT [2].

2.2. Build a marine cloud storage framework

The Marine Internet of Things need to support data management and background maintenance. When building a cloud storage framework, design a four-layer framework. The basic layer is the storage layer, use SAN, FC and iSCSI and other equipment to connect the ship WAN for managing and storing ship transmission data files [3]. In the basic management layer, the upper layer interface is used to connect the storage devices of the storage layer to realize the management function of cloud storage. The application interface layer mainly connects the storage layer and the basic management layer to the cloud computing network. In order to ensure the real-time nature of the transmission of IoT sensor data to the cloud computing network, and reduce the processing time of sensor data, the calculation is:

\[ T_{\text{total}} = T_e + T_T + T_H \]  \hspace{1cm} (5)

Among them, \( T_{\text{total}} \) represents the total elapsed time, \( T_e \) represents the sensor data acquisition time, \( T_T \) represents the sensor data transmission time, and \( T_H \) represents the sensor data processing time. After shortening the processing time of the sensor data, the transmitted data is finally transmitted to the final
2.3. Develop storage strategy

When formulating storage strategies, first establish a direct mapping between files and storage strategies in the Internet of Things, and form six storage strategies based on the node performance calculated above, as shown in the following table:

| No. | Strategy          | Date block (n copy)          |
|-----|------------------|-----------------------------|
| 1   | LAZY PERSIST     | RAM DISK:1, DISK: n-1       |
| 2   | ALL SSD          | SSD: n                      |
| 3   | ONE SSD          | SSD: 1, DISK: n-1           |
| 4   | HOT (default)    | DISK: n                     |
| 5   | WARM             | DISK: 1, ARCHIVE: n-1       |
| 6   | COLD             | ARCHIVE: n                  |

According to the storage strategy shown in the table above, complete the research of the distributed cloud storage technology in the marine Internet of Things [4].

3. Experiment

3.1. Build an experimental platform environment

The experimental environment uses a fully distributed model, and 5 Ubuntu 12.04 virtual machines are built on the cloud platform for the experiment. Configured as 2 CPUs (2VCPU), 4GB memory and 50GB memory. The five virtual machines are: algo2, algo5, algo6, algo8, and algo10, where the algo2 is as the NameNode, master, and JobTracker of the cluster, and the remaining four virtual machines are used as DataNode, slave, and TaskTracker. Java version is version 1.6.0-37. The allocation of virtual machines is shown in the following table:

| No. | System | IP address  | Node type             |
|-----|--------|-------------|-----------------------|
| algo2 | Hadoop-A | 192.168.22.19 | NameNode master JobTracker |
| algo5 | Hadoop-B | 192.168.22.20 | DataNode slave TaskTracker |
| algo6 | Hadoop-C | 192.168.22.24 | DataNode slave TaskTracker |
| algo8 | Hadoop-D | 192.168.22.23 | DataNode slave TaskTracker |
| algo10 | Hadoop-E | 192.168.22.25 | DataNode slave TaskTracker |

Randomly select 10,000 marine transmission files, and set the file size between 2MB-5MB. The default block size is 64MB, the number of copies of each block is 3, starting from 1000 number of files, incrementing 1000 files each time, using three storage technologies to store the 10,000 transfer files prepared by the experiment, comparing the three types of storage The final combined amount of the mode file.
3.2. Analysis of experimental results

The marine storage files start from 1,000. The three storage technologies will merge and store the files to be stored according to the tightness of the coordinates. The files of the same cluster will be merged into one large file by the storage technology, reducing the number of file storage, thereby reducing the number of files in the memory. As the amount of data decreases, the amount of memory used decreases. The experimental results of the three storage technologies are shown below:

![Fig.1 Traditional storage technology 1 experimental results](image1)

![Fig.2 Traditional storage technology 2 experimental results](image2)

![Fig.3 Optimized Distributer storage technology experimental results](image3)

Fig.1-3 are the Experimental results of distributed cloud storage technology in marine Internet of things. It can be seen from the above three experimental result graphs that the three storage methods will gradually increase the memory usage and processing time as the number of stored files increases. Observe that the number of files corresponding to the widest part in the grid graph is the final number of merged files. When traditional storage technology 1 processes the same cluster of files, the maximum file merge amount is 4000, the traditional storage technology 2 processes the maximum file merge amount is 6000, and the distributed cloud storage technology in the marine Internet of Things file merge The quantity is 8000. The distributed cloud storage technology in the Marine Internet of Things has more file merges, reducing the excessive memory occupied by the marine files, which is more suitable for actual use.
4. Conclusion

With the continuous application of science and technology, all fields have gradually developed into a "connection" mode using the network as a medium, using sensing and computing capabilities to connect all areas of the ship to the Internet, ship storage technology faces massive data generation. The distributed cloud storage technology in the marine Internet of Things. Studying this technology can increase the file merge ability and reduce the memory occupied by related marine files.

Reference

[1] QIN Weirong. Construction of informationized Internet of things under cloud computing and big data [J]. Laser Journal, 2018, 39(05): 120-123.

[2] Tang Hong. Storage System Design of Laser Scanning Point Cloud Database [J]. Bulletin of Science and Technology, 2018, 34(09): 127-131.

[3] Chen Qing, Yang Ming, Xiao Yun, Wu Jingsheng, Teng Zhou, Wei Shuang, Hu Yongliang, Pang Jun. APPLICATION OF CLOUD DATA STORAGE TECHNOLOGY IN METEOROLOGICAL DATA STORAGE [J]. Computer Applications and Software, 2018, 35(08):124-127+158.

[4] YI Hua, HAN Xiao, WANG Kai-lun, NIU Guang-li, HUANG Yue-wen. Application of Internet of Things Technology to Safety Monitoring Automation System of Large Hydropower Station [J]. Journal of Yangtze River Scientific Research Institute, 2019, 36(06): 166-170.