Research on cloud computing and services framework of marine environmental information management

SHI Suixiang$^{1,2,*}$, LIU Yang$^3$, WEI Hongyu$^{1,2}$, QIAO Baiyou$^4$, WANG Guoren$^4$, XU Lingyu$^3$

$^1$ Key Laboratory of Digital Ocean, State Oceanic Administration, Tianjin 300171, China
$^2$ National Marine Data and Information Service, State Oceanic Administration, Tianjin 300171, China
$^3$ School of Computer Engineering and Science, Shanghai University, Shanghai 200072, China
$^4$ College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

Received 13 November 2012; accepted 22 May 2013

©The Chinese Society of Oceanography and Springer-Verlag Berlin Heidelberg 2013

Abstract

Based on the features of marine environmental data and processing requirements, a cloud computing architecture of marine environment information is proposed, which provides a new cloud technology framework for the integration and sharing of marine environmental information resources. A physical layer, software platform layer and an application layer are illustrated systematically, at the same time, a corresponding solutions for many difficult technical problems such as parallel query processing of multi-dimensional, spatiotemporal information, data slice storage, software service flow customization, analysis, reorganization and so on. A prototype system is developed and many different data-size experiments and a comparative analysis are done based on it. The experiment results show that the cloud platform based on this framework can achieve high performance and scalability when dealing with large-scale marine data.

Key words: cloud computing, marine information, software services, virtualization


1 Introduction

Marine environmental information has the features of multi-source, diverse, multi-state and huge amount of information, and cloud computing has many advantages in dealing with this type of massive data and providing business services which can be summarized as efficiency, security, reliability, low cost and flexible organizing and service. Therefore, in recent years, developed countries led by the United States, consider cloud computing as one of the strategic key technologies. The Berkeley Laboratory studies have shown that both the cloud service providers and the cloud service users can benefit from the framework of cloud computing. In the aspect of cloud computing architecture, many new frameworks have been proposed to solve different problems. This paper proposed the architecture which included leveraged layer based on market resource allocation strategies, negotiating the allocation of a resources based service layer and network-based global market cloud platform layer. The framework proposed by Wu Weihua is to solve the e-commerce development model which is divided into three layers, the infrastructure layer, the application layer. This paper and other papers proposed a cloud bank model, including the level of provider, cloud bank and cloud resources consumers, and developed an accelerated multimedia synchronous communication cloud computing framework, including the control layer, the virtual instance layer and the cloud storage layer.

In order to satisfy the requirements of sharing information resources, model resources of the marine environment, fully integrate existing computing resources and storage resources, we plan to use three levels to realize the cloud computing and cloud services framework applying research of the marine environment, including the physical layer, the software platform layer and the application layer. In addition, the development of the prototype system and the experimental data test are conducted to prove its advancement and reliability.

2 Physical layer

The physical layer is the basis of the prototype demonstration system and the main function is to provide computing resources, storage resources and data resources for the software platform layer. The physical layer mainly consists of hardware resources and virtual resources.

2.1 Hardware resources

Hardware resources include servers, network equipment, storage devices and other hardware facilities, all hardware resources must have high scalability and form the infrastructure of the cloud computing data center. Users can use these resources and pay a service fee based on usage according to their needs. The hardware resources are based on the two physical nodes that are the National Oceanic Information Center and the Information Center of the South China Sea. The researches of marine environmental information cloud computing system involve physical resources constitution and its virtualization technology, management techniques about physical resources and virtual resources, including virtualization of computing resources, storage resources, and information reso-
urces, as well as the deployment of resources, monitoring and dynamic migration of virtual machine. In the National Marine Information Centre gives the main node providing 20 servers, a high-performance computing clusters and a SAN storage array. The South China Sea Information Centre gives an auxiliary node providing three servers and a SAN storage array. In addition, it is also equipped with a high-performance server and client terminal hardware resources. The computing resources, storage resources, as well as high-performance management servers and clients are connected with a digital marine network. They constitute the physical infrastructure of cloud computing system for marine environmental information, and the hardware resources logical structure is shown in Fig. 1.

Fig.1. Hardware resources logical structure of marine environment information cloud computing system.

2.2 Virtual resources

On the basis of physical infrastructure, we will realize the virtualization of physical resources through inducing the virtualization platform layer and transform the physical resources into unified virtual resources. The virtualization platform layer consists of a virtualization hypervisor and a virtualization management software. The virtualization hypervisor can transform the physical processor, memory and network resources into virtual processor, memory and network resources, and package the marine environment data and marine model resources into virtual processor, memory and network resources, and package the marine environment data and marine model resources into unified virtual marine environment information resources. Thus it forms virtual resource level consisting of many different virtual resource pools, and provides cloud computing platform transparent access to these virtual resources. The virtualization platform layer integrates the physical resources which come from different nodes with different structures and forms large resource pools for the platform layer's usage. The virtualization management software is an important aspect of virtualization technology, the main function includes resources deployment, resources monitoring, virtual machine load management, dynamic optimization, backup management, application programming interface, and so on. The architecture of resources virtualization platform is shown in Fig. 2.

Fig.2. Architecture of resource virtualization.

The main researches focus on basic services of cloud computing platform, including the system level service, tolerant service, cloud services, resources and user management. The system level service research is based on the distributed file system and the MapReduce parallel computing framework. The tolerant service mainly studies the backup mechanism of data and application service, and the mechanism of failure detection and recovery. The cloud service mainly studies the diverse cloud services for the terminal application. The resources and user management studies the resources registration, querying, monitoring, user access control and so on.

Hadoop is an open source cloud computing platform which provides a cloud computing framework for distributed parallel processing on large clusters consisting of cheap hardware equipment and provides a set of stable and reliable transparent access application interface. Hadoop implements the MapReduce framework which can send applications to corresponding data node for executing or repeated executing. Moreover, Hadoop provides a distributed file system to store data at data nodes and a high throughput data read and write mechanism. What is more, because of the use of a kind of parallel programming model for large-scale data sets a distributed file system architecture developed by MapReduce and Apache Foundation–Hadoop, the platform can achieve high efficiency, credibility, openness, and fault tolerance.

3 Query and storage service

3.1 Data slice

Based on the feature of marine data, marine cloud computing system will use a relational database to store relational data while directly using the Hadoop distributed file system (HDFS) to store non-relational data, such as XML data, text data, graph data and image data. Because of using a normalized design for data stored in the relational database, it may need to execute multiple connection operations to complete a query, thus makes the query efficiency very low. So we will use a denormalized design method to improve the query efficiency of relational data, this needs to reconstruct the existing application's data structure to reduce the connection operations. But there is a problem that data pattern of cloud system is different from existing application system, and it can bring confusion to application programmers. To solve this problem, we will use view technology to define the virtual tables with the structure as same as that of the existing application system.