

Research of Cloud Computing in Application of University Information

Bin Li

School of Computer Engineering, Shenzhen Polytechnic College, China

libin@szpt.edu.cn

Keywords: Cloud Computing, Virtualization, Scalability

Abstract. There is a huge contradiction between the cost and the application in the context of accelerating University information, this paper proposes a way to resolve by making full use of the large number of idle computing resources to build cloud computing platform.

Background

As information technology continues to progress at home and abroad for college construction investment continued to increase, rising standards of university information, especially in recent years, the rise of the digital campus to achieve from the environment (including equipment, classrooms, etc.) and resources (such as books, handouts, courseware, etc.) to the application (including teaching, learning, management, service, office, etc.) of all digital technology, traditional campus foundation to build a digital space, in order to expand the reality campus of temporal and spatial dimensions, enhance the operating efficiency of the traditional campus expansion the traditional campus business functions, and ultimately the educational process of comprehensive information.

Development in information technology to enhance the college management efficiency, but also produce a variety of data, and how to effectively analyze mining these data provide strong support for the decision analysis of the various departments of the college bound college next information technology the trend of development. These data analysis often has the following characteristics:

(1) a large amount of data, are usually TB level, and even Class PB, and cross analysis, multi data source;

(2) various types of data, both structured data, there are a lot of unstructured data, such as text, voice and image, and a large number of used full-text search;

(3) data timeliness requirements, and some of the data has a higher timeliness requirements.

With the increase of the amount of data and analysis needs diversification, colleges and universities often need to purchase more and more high-end computer in order to meet the increasingly high computing needs. At the same time, colleges and universities every year, a large number of low-end computer to be phased out, to be deposited in the warehouse, the ability to use idle computing power? The emergence of cloud, provides the possibility for this idea.

Concepts and characteristics of cloud computing

Cloud computing is grid computing, distributed computing, network storage, virtualization, load balancing and other traditional computer and network technology development and integration of the product. Aimed at through the network to a plurality of relatively low-cost computing entity integrated into a system having a powerful computing capability.

In the traditional manner, the construction of a data processing center, the need to purchase a minicomputer, to respond to the needs of a large number of data analysis calculated. In this way the main drawback is that: (1) a one-time investment is huge and difficult to adapt to the needs of future business development; (2) system is often the reference the maximum computational requirements to configure, regardless of the treatment is a small problem or a big problem, occupied by the computing and storage resources, a waste of resources; computing resources (3) colleges and universities do not lack a lot of low-end equipment is often in the idle state, and can not play a role.

The idea of using cloud computing, we will be a large number of low-end computer configured as a compute node build virtualized computing cloud, when the need to solve the problems of computing power, task scheduling layer is responsible for the large tasks into many small parts, then assigned to these parts of many low-performance computers to process, the final results of these calculations together, to achieve the purpose to solve the problem quickly. And when the computing tasks lighter, you can dynamically reduce the number of computing nodes to reduce resource consumption.

Cloud computing has the following characteristics:

(1) Ultra-large-scale: a considerable scale, already has more than 100 million servers Google cloud computing, Amazon, IBM, Microsoft, Yahoo and other "cloud" with hundreds of thousands of servers. Enterprise private clouds generally have hundreds of thousands of servers. "Cloud" gives users unprecedented computing power.

(2) Virtualization: support user access to application services at any location, using a variety of terminals. The requested resource from the "cloud", rather than a fixed physical entity. Application running somewhere in the "cloud", but in fact do not need to know, do not worry about the specific location of the application running. Only need a laptop or a mobile phone, can be achieved through the network services we need, even this task supercomputing.

(3) High reliability: the "cloud" data more copies of fault-tolerant, high reliability services same constitutive interchangeable, and other measures to protect the compute nodes to use cloud computing reliable than using the local computer.

(4) Versatility: Cloud computing is not targeted at a specific application, in an ever-changing applications can be constructed in support of the "cloud", with a "cloud" can support different applications running simultaneously.

(5) High scalability: the size of the "cloud" dynamically scalable, to meet the needs of the application and user scale growth.

(6) on-demand services: the "cloud" is a huge resource pool, you demand to buy; cloud like running water, electricity, gas, billing.

System Design

The system uses the Eucalyptus build infrastructure clouds, virtualization technology, low-performance computers configured to facilitate the management of compute nodes, the formation of the system infrastructure cloud. Eucalyptus is an open source software infrastructure, to achieve flexible computing clusters or workstation group, practical cloud computing. Logically, Eucalyptus cloud consists of four components: storage controller (walrus), cloud controller (CLC or "cloud"), the cluster controller (cc), and the node controller (nc). When the need for more computing resources interface add in Eucalyptus, new resources join; contrary, in the Eucalyptus delete interface, release resources in order to achieve the scalability of the system and scalability.

On this basis hadoop in distributed parallel framework to build a distributed computing environment. System through a task scheduler module, the big task is decomposed into multiple sub-tasks, intelligent scheduling run between compute nodes. Has a certain degree of autonomy to the various sub-tasks, that is, each sub-task is not simply a passive acceptance of the scheduler allocate computing resources, but will take the initiative to monitor where computing node status, when the compute node failures or resource constraints, subtasks consultation with the scheduler, continue transferred to other nodes, thus effectively avoiding the compute node failures result in task execution failed to ensure that the system has a certain degree of fault tolerance.

The entire system uses a five-tier structure, as shown in the following figure 1.

(1) Presentation layer through a variety of terminal access system, and want to check on the investigation, including desktop computers, PDAs, smart phones, laptop computers and other smart devices anytime, anywhere.

(2) Scheduling layer works as the center console, depending on the load, real-time scheduling, intelligent distribution service request to each compute node.

(3) Business logic layer encapsulates the model of a variety of business topics, and is responsible for the management and create a business model.

(4) Computing layer is responsible for the creation and management of a variety of mathematical models, to provide operational support services for all other layers, such as probabilistic algorithms mathematical model for trend trend analysis, ant colony algorithm for analysis.

(5) Data buffer layer structure based on pure main memory on the media avoid low-speed devices such as hard disk, and the establishment of hash buckets, data pools, dynamically Manager component, server cluster, unified management and use of memory, and to establish the appropriate buffer and index strategies to boost system performance.

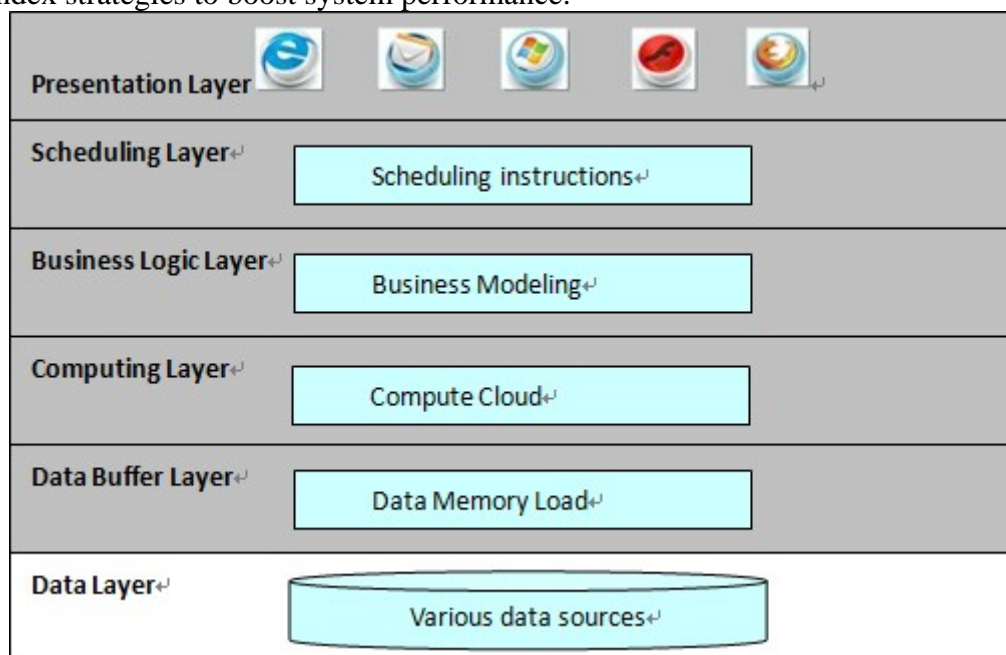


Fig. 1 Five-tier Structure

Summary

With the accelerating pace of the college of information technology, the contradiction between the huge cost and application has become increasingly apparent. Build a cloud computing platform, a large number of low-performance computing resources together to provide powerful, scalable computing power will college teaching cost savings and improve the quality of teaching good breakthrough.

References

- [1] Amazon Inc. Amazon Elastic Compute Cloud (Amazon EC2) Case Studies. <http://aws.amazon.com/ec2/>, 2010
- [2] Buyya, Rajkumar; Chee Shin Yeo, Srikumar Venugopal. Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities [J]. Department of Computer Science and Software Engineering, The University of Melbourne, Australia: 9 [2008-07-31].
- [3] Gruman, Galen. What cloud computing really means [J]. InfoWorld. 2008-04-07 [2009-06-02].
- [4] Huan Liu, Dan Orban, "GridBatch: Cloud Computing for Large - Scale Data - Intensive Batch Applications, " ccgrid , pp.295-305, 2008 Eighth IEEE International Symposium on Cluster Computing and the Grid (CCGRID), 2008.
- [5] Brian Hayes, Cloud computing, Communications of the ACM, Volume 51, Issue 7, July 2008.