The Application of Artificial Intelligence Technology in Cloud Computing Environment Resources

Xiangbin Wen and Yuan Zheng*

Information Center, GuangZhou University of Chinese Medicine, Guangzhou 510006, Guangdong, China
E-mail: b536631@163.com
* Corresponding Author

Abstract

With the continuous innovation and development of modern computer science and mobile Internet and other information technologies, artificial intelligence (AI) is not a new thing. It has been widely studied and applied in many fields, and it is very important for people in modern society. The research fields of artificial intelligence mainly include: deep learning, natural language processing, computer vision, intelligent robot, automatic programming, data mining and so on. All kinds of industrial production and daily life will bring a very important practical significance and far-reaching influence. The rapid development and improvement of AI have effectively changed the daily life of modern people and improved work efficiency, and promoted the vigorous and healthy development of human economic and social civilization and the progress of information technology. When widely used, traditional network information and big data processing technologies are difficult to adapt to its development needs. Only by closely combining cloud computing technology with other technologies can it play a better role and give full play to AI technology and its development. The enthusiasm and promotion

of related application technologies have promoted the smooth progress of AI technology and related undertakings. With the development and improvement of cloud computing technology, more and more users tend to use the cloud to work. However, a large number of cloud service failures occurred, causing huge losses for enterprises and individuals. In order to prevent damage to the interests of enterprises and individuals, cloud service providers will provide high-quality services as much as possible. This paper aims to study the application of AI technology in cloud computing environment resources, research on the indicator of reliability, and propose a cloud service reliability verification method for the infrastructure-as-a-service layer. Experimental research shows that through the reliability detection method in this paper, users can easily and quickly obtain the reliability of the purchased cloud service, and can intuitively feel whether the performance of each server meets the promised situation in the cloud service provider's SLA.

1 Introduction

The main purpose of the rapid development of AI is to better serve industries in other fields. Cloud computing-based AI-as-a-service products are also the best presentation of this development purpose [1, 2]. In order to always maintain a leading position in market competition, more and more AI companies are trying to integrate AI technology into other related applications, products, services, and various big data analysis methods [3, 4]. And modern enterprises have begun to consider the use of AI technology to develop one of the easiest and most popular way is to consider the use of cloud computing-based AI as a service products [5, 6].

Cloud computing is a kind of distributed computing, which means that huge data processing programs are decomposed into numerous small programs through the network "cloud", and then processed and analyzed by a system composed of multiple servers. The results of these small programs are returned to users. Cloud computing is an emerging application model. Cloud service providers integrate a large number of computing, storage, and network resources, and provide these resources to users through virtualization and other technologies. These resources are allocated on demand and paid for by volume [7, 8]. With the continuous development of cloud computing technology, its application range is becoming wider and wider. In order to

reduce costs, enterprises and individuals are more inclined to migrate services to cloud platforms. Under this trend, people demand the availability and reliability of cloud platforms [9, 10]. In the early days of cloud computing, simply speaking, it was a simple distributed computing, which solved the problem of task distribution and merged the results. Therefore, cloud computing is also called grid computing. Through this technology, tens of thousands of data can be processed in a very short time (a few seconds), so as to achieve a powerful network service.

This paper proposes an IaaS-oriented cloud service reliability detection method. By mapping the three parameters of QPS value, CPU performance, and memory performance as indicators of cloud service reliability, the analytic hierarchy process is used to perform weighted summation, and from a trusted third party from the point of view, the reliability of cloud services is scored. Users can obtain the reliability of cloud services according to the reliability standards provided by trusted third parties. According to their own needs, they can choose suitable cloud service providers to meet their own dual performance and price.

2 Research on the Application of AI Technology in Cloud Computing Environmental Resources

2.1 Features of Cloud Computing

(1) Super large scale
Cloud computing provides users with a lot of resources. These resources cannot arise out of nothing. Therefore, there must be a place to store these resources. Take Google as an example. It already has more than 1 million servers that support the "cloud" it owns. And IT giants such as Amazon, Microsoft, and IBM also have hundreds of thousands of servers supporting their own clouds. Generally speaking, enterprise-level private clouds have hundreds or thousands of servers [11, 12].

(2) Virtualization
Virtualization technology is the foundation of cloud technology. Cloud computing allows users to use various services on their terminals at any place and at any time. The requested resources all come from the resource pool of the cloud platform. These services run somewhere on the platform. Then users do not need to know where their services are running. Therefore, it can be said that the entire cloud service is completely transparent to users.

(3) High reliability

In order to ensure high data security, various technologies will be used in the platform to ensure data security, and provide users with highly reliable, high fault-tolerant services, such as multiple copies of data, or technology such as service migration.

(4) Versatility

It's not just about providing services for a certain application, the cloud platform should be a platform that can provide different services for different users. For example, while providing user A with computing resources for scientific research projects, it also provides B with corresponding resources to complete other tasks.

(5) High scalability

The platform's resource pool will not have a fixed size, but can be dynamically adjusted. In this way, when a resource node is idle, it can be used as a part of the platform resource pool to provide resources for the resource pool to serve other users. When the node is used for other purposes, the management of the platform can be removed again to provide users with one-to-one services.

(6) High cost performance

When users use the platform, all services can be deployed on the platform. In the case of constant work efficiency, the maintenance cost of the server is greatly reduced, thereby increasing the input-output ratio.

2.2 Classification of Cloud Computing

Traditionally, cloud computing frameworks are divided into three categories. The order of division is from bottom to top, according to the level of service. They are: infrastructure as a service (IaaS) for underlying hardware, platform as a service (PaaS) for platform, and software as a service (SaaS) for software. Infrastructure as a service: it provides consumers with processing, storage, network and various basic computing resources to deploy and execute various software such as operating system or application program.

The acronym that combines the initials of the three phrases SaaS, PaaS, and IaaS is SPI. This is the SPI pyramid model. As shown in Figure 1, not all cloud platforms conform to this model. The cloud platform can be one or a combination of several.
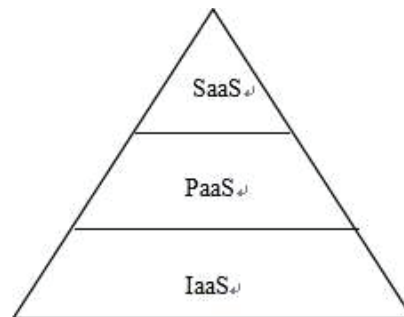
**Figure 1**   Cloud computing classification diagram.

2.3 Key Technologies of Cloud Computing

(1) Energy management technology

In large and medium-sized data centers, not only the power consumption of computer equipment such as servers, but also the power consumption of auxiliary equipment such as cooling. Generally speaking, the power consumed on a computing device is about the same as the power consumed on other auxiliary devices. In other words, if the power consumption of computing equipment in a data center is 1, then the power consumption of the entire data center is 2. For some very good data centers, using some advanced technologies, the power consumption can reach 1.7 at most, but Google has made some data centers reach the industry-leading 1.2 through some effective designs. In these designs, one of the most distinctive features is the high temperature of the data center, which means that the computing equipment in the data center runs at a relatively high temperature.

(2) Virtualization technology

Virtualization technology is the most basic technology to realize cloud computing, which realizes the logical abstraction and unification of physical resources. Using this technology can improve the use efficiency of physical hardware resources, according to the needs of users, flexible and rapid configuration and deployment of resources. Virtualization technology is completely different from multitasking and hyper threading technology. Multitasking means that multiple programs run in parallel in one operating system, while in virtualization technology, multiple operating systems can run at the same time, and there are multiple programs running in each operating system, and each operating system runs on a virtual CPU or virtual host; The hyper threading technology is just a single CPU simulation of two CPUs to balance the
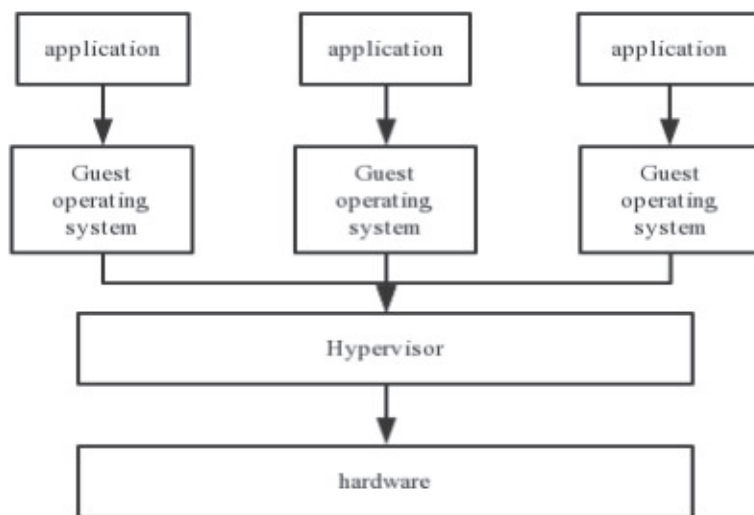
Figure 2   Schematic diagram of virtualization architecture.

performance of the program, the two simulated CPUs can not be separated, can only work together.

In cloud computing, virtualization is achieved by running multiple virtual machines on the physical host at the same time. In the cloud computing platform, the platform always maintains the monitoring of these virtual machines, as well as the allocation and deployment of resources, as shown in Figure 2, which is a schematic diagram of the virtualization architecture.

(3) Mass data management technology

Cloud computing systems need to process and analyze data efficiently, and colleagues also need to provide users with high-performance services. Therefore, in data management technology, how to find the required value in such a huge scale of data has become a core issue. The data management system must have the characteristics of high fault tolerance, high efficiency, and ability to operate in a heterogeneous environment. In traditional IT systems, technologies such as indexing, data caching, and data partitioning are commonly used. In the cloud computing system, because the amount of data greatly exceeds the amount of data owned by the traditional system, the technology used in the traditional system is difficult to handle. Database management system is a data operation software that can provide data entry, modification and query. It has the functions of data definition, data operation, data storage and management, data maintenance, communication, and can be

used by multiple users. In addition, the development of database management system is closely related to the development of computer technology.

(4) Distributed storage technology

The cloud computing system is composed of a large number of server parts and also provides services for a large number of users. In order to better ensure the reliability of data transmission, redundancy or spatial storage is adopted to process and store these massive data. Distributed file system is a way to achieve data transmission and storage by means of redundancy or idle storage. It mainly evolved from the file system and is suitable for various distributed file systems on the cloud platform. For a new type of data storage technology, its high reliability, i/0 throughput capacity and load balancing ability will undoubtedly become the core of it. In terms of the reliability of information storage, the platform system has a special function of supporting multiple copies of data between a node to improve and enhance the reliability of information. In terms of i/0 throughput control capability, according to the importance of data and access frequency, the system usually stores various data in multiple copies of a certain category, and parallel reads and writes to various hot data. Thereby improving the I/0 throughput capacity. In terms of load balancing, the system migrates node data to new or lower load nodes based on the current system load.

Traditional network storage system uses centralized storage server to store all data. Storage server becomes the bottleneck of system performance and the focus of reliability and security, which can not meet the needs of large-scale storage applications. The distributed network storage system adopts an extensible system structure, uses multiple storage servers to share the storage load, and uses location server to locate the storage information. It not only improves the reliability, availability and access efficiency of the system, but also is easy to expand.

3 Experimental Research on the Application of AI Technology in Cloud Computing Environmental Resources

3.1 Performance Testing Tools

For the QPS value, this article mainly uses Apache's own stress test tool AB, and for the CPU performance and memory performance, mainly uses the Geekbench tool.

3.1.1 QPS acquisition tool ab

ab is a stress test tool that comes with apache itself. Apache comes with its own ab testing tool, which can test servers such as apache, is, tomcat, and client. ab has an output description, it can simply perform some motor pressure tests on it, which is very convenient and quick, simple to operate, and users can directly install it at any time.

The ab command is to create concurrent access threads in multiple URLs to simulate multiple users simultaneously accessing the address location in a specific URL. The use of the ab command has very low power requirements for the issuing computer and the computer, and will not occupy very high cpu and memory in the computer. However, for a target server, it will cause extreme load to them, and its working principle is similar to the cc (challenge collapsar) attack. Once the concurrency set by the system is too large, it may directly cause the overload of the target server and even cause the system to be paralyzed.

3.1.2 Unixbench

Unixbench is a tool for testing the performance of Unix systems. It can not only test single-core systems, but also supports performance testing of multi-core systems. By default, it performs two tests, the first is a single-process test, and the second is a multi-process test. This is to be able to fully test the single-task and multi-task performance and parallel processing capabilities of the system. Unixbench not only tests the CPU situation of the system, but also includes tests such as memory and system read and write capabilities. Each test obtains a score, which is compared with the benchmark score to obtain an index value. The performance index of the entire system is obtained after weighting by integrating each index value.

UnixBench will perform a series of tests, including performance measurement of 2D and 3D graphics systems. The test results are not only based on CPU, memory or disk, but also depend on hardware, operating system version and compiler.

3.1.3 Geekbench

Geekbench is a comprehensive system performance testing tool and one of the best processor benchmark tests on the market. It supports multiple systems and mainly tests the performance of CPU and memory. In its test, integer performance is more important, with scores accounting for 50% of the overall value. By introducing some test items in image processing and database, the whole integer part of the test is more balanced; in the floating-point

workload test, the test is considered. The balance and diversity of the matrix multiplication and fast Fourier transform tests have been reduced from two sub-items to one, and some projects with more development prospects have been added. This part of the score accounts for 30% of the total. The test almost covers most of the performance parameters of the CPU and memory, and can accurately test the stability and reliability of the system.

3.1.4 Stream

Stream is a commonly used memory bandwidth performance test benchmark tool in memory testing. It can give full play to the performance of the memory, can test the actual maximum value of the system memory bandwidth, and is widely used. The test process is mainly based on the operations of four arrays, namely Copy (copy of the array), Scale (scale transformation of the array), Add (vector summation of the array), and Triad (compound vector summation of the array). The sample result is shown in Figure 3:

Function	Best Rate MB/s	Avg time	Min time	Max time
Copy:	19056.6	0.016880	0.016792	0.017026
Scale:	19249.3	0.016837	0.016624	0.017008
Add:	20265.2	0.023860	0.023686	0.023983
Triad:	20347.5	0.023796	0.023590	0.023953

Figure 3 Stream performance test example results.

3.2 Algorithm Analysis

Given that the observation set $X = \{x\}$, $i = 1,2\ldots n$, where each observation value is a d-dimensional real vector, k-means clustering is to divide the n observation values of the observation set X into the k set $C = \{c\}$, $k = 1,2,\ldots K$, and kSn, making the sum of squares within the group the smallest. Let μh be the mean value of c, then the variance of each sample data and the mean value in each category is defined as:

$$J(c_k) = \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \tag{1}$$

The goal of this algorithm is to minimize the variance of all K categories:

$$J(c) = \sum_{k=1}^{k} \sum_{x_i \in c_k} \|x_i - \mu_k\|^2 \tag{2}$$

4 Application Experiment Analysis of AI Technology in Cloud Computing Environment Resources

Iaas layer cloud service reliability testing verification and analysis.

Based on the analytic hierarchy process, this article weights the three parameter limit QPS values, CPU performance and memory performance of the reliability index. Its structural model is shown in Figure 4:
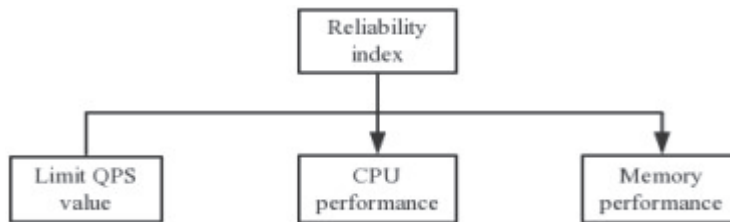
Figure 4 Reliability index model.

This paper verifies the validity of the reliability method of this paper by conducting hypothetical experiments. Assume that the 3rd server os03 is a true and reliable virtual machine configuration promised by the cloud service to the user, and the reliability index obtained by the test is the reliability standard; the 1st server os01 and the 2nd server os02 are two provided by the user. The hypothetical experimental environment is that the cloud service provider promises to provide users with two virtual machines, the configuration is 4 core 8G memory 3M bandwidth, in fact, the virtual machine configuration provided to the user is the 1 core CPU of the 1st server. 1GB memory, 1MB. Bandwidth; 2 core CPU, 2GB memory, 1MB bandwidth of No. 2 server. When a user applies to a trusted third party for reliability testing, the trusted third party applies to the cloud service provider for a virtual machine (ie os04) configured with a 4-core 8G memory and 3M bandwidth, and tests its reliability index as a reliability standard. Test the reliability indicators on the user's virtual machine (ie os01, os02, os03). Obtain the reliability of cloud services through the reliability formula.

Table 1 Reliability index parameters of No. 3 server

Server	Limit QPS			CPU Performance			Memory Performance		
	Max	Min	AVE	Max	Min	AVE	Max	Min	AVE
Os03	38321.99	36354.55	37962.97	34582	34491	34554.6	6591	6540	6567.8

It can be seen from Table 1 that the reliability index of server 3 os03 is 0.7606. Because the No. 3 server os03 is a detection tool provided by the cloud service provider to a trusted third party, and its information is true and reliable, the reliability index obtained by the No. 3 server os03 test is the reliability standard. Subsequently, the user reliability index test is carried out.

Table 2 Reliability index parameters of No. 1 server

Server	Limit QPS			CPU Performance			Memory Performance		
	Max	Min	AVE	Max	Min	AVE	Max	Min	AVE
Os01	9769.64	8785.71	9301.226	26868	26615	26727.2	10345	7535	9092.4

According to the data in Table 2, the reliability index of server 1 os01 is 0.5069.

Table 3 Reliability index parameters of No. 2 server

Server	Limit QPS			CPU Performance			Memory Performance		
	Max	Min	AVE	Max	Min	AVE	Max	Min	AVE
Os02	26592.94	18453.22	22850.08	33791	31624	33262.6	8553	6440	6951

In the same way, the single-core reliability index of server 2 os02 is 0.5643.

By repeating the reliability test 20 times, the normalized value obtained is weighted and summed through the analytic hierarchy process to obtain the reliability index of the cloud service, and the reliability of the virtual machine is obtained by comparing with the reliability standard. The reliability of server 1 os01 is 66.64%, and the reliability of server 2 os02 is 74.19%. Analyzing the reliability of three users, verifying the validity of the reliability test method of the IaaS layer cloud service in this paper.

Therefore, through experimental verification, through the reliability detection method in this article, users can easily and quickly obtain the reliability of the purchased cloud service, and can intuitively feel whether the performance of each server meets the promised situation in the cloud service provider's SLA.

5 Conclusions

With the help of China's existing mobile Internet and new-generation electronic information and other Internet technologies, through cloud computing,

mobile Internet, Internet of Things, and AI and other Internet technologies have been closely intersected and integrated on the basis of science and information technology. Can effectively solve the problem of constraints on the analysis and processing of large amounts of data in various networks, and improve the analysis and processing speed of large amounts of data in various networks, which has promoted and promoted the current revolution of China's network information science and technology Innovation and the development of science and technology play a very important role in active promotion and promotion. At the same time, related network technologies can better adapt to different types and diverse user needs in actual applications. Form a new business model that really helps, and really make data generate value References. Through a large number of experiments, the effectiveness and superiority of the proposed algorithm are verified.

References

[1] Sun L, Jiang X, Ren H, et al. Edge-Cloud Computing and Artificial Intelligence in Internet of Medical Things: Architecture, Technology and Application. IEEE Access, 2020, PP(99):1–1.

[2] Xin C, Wang X. Research on the Application of College Physical Education Teaching Mode in the Cloud Computing Environment. Journal of Physics: Conference Series, 2020, 1624(2):022068 (4 pp).

[3] Wang Q, Lu P. Research on Application of Artificial Intelligence in Computer Network Technology. International Journal of Pattern Recognition and Artificial Intelligence, 2019, 33(5):1959015.1–1959015.12.

[4] Bogue, Robert. Cloud robotics: a review of technologies, developments and applications. Industrial Robot, 2017, 44(1):1–5.

[5] Pandian A P. Artificial intelligence application in smart warehousing environment for automated logistics. Journal of Artificial Intelligence and Capsule Networks, 2019, 2019(2):63–72.

[6] Yu Z, Liu Y, Zhu C. Application of Propofol in Oral and Maxillofacial Surgery Anesthesia Based on Smart Medical Blockchain Technology. Journal of Healthcare Engineering, 2021, 2021(1):1–11.

[7] Lai Z, Yu L. Research on Digital Marketing Communication Talent Cultivation in the Era of Artificial Intelligence. Journal of Physics: Conference Series, 2021, 1757(1):012040 (5 pp).

[8] Wu Q, Xu H, Zhao Y, et al. Cloud-based smart emergency rescue system and its application in mine water disaster. Meitan Xuebao/Journal of the China Coal Society, 2018, 43(10):2661–2667.

[9] Qiao L, Zhang X. Frequency Management Method Based on Cloud Computing, Big Data and Artificial Intelligence. Journal of Physics: Conference Series, 2021, 1757(1):012106 (8 pp).

[10] Xu Z, Liu W, Huang J , et al. Artificial Intelligence for Securing IoT Services in Edge Computing: A Survey. Security and Communication Networks, 2020, 2020(1):1–13.

[11] Trong H B, Kim U. Application of Information and Technology in Supply Chain Management: Case Study of Artificial Intelligence – A Mini Review. European Journal of Engineering Research and Science, 2020, 5(12):19–23.

[12] Huang D. Application of Translation Technologies in the Translation of IMTFE Transcripts. English Language and Literature Studies, 2021, 11(1):51.

Biographies

Xiangbin Wen was born in Yantai, Shandong Province, in 1977. Master degree of Guangzhou University of Chinese medicine. Now engaged in higher education research, information planning, management and construction, the main research direction: Internet of things technology, cloud technology, network security, etc.

Yuan Zheng was born in Born in Shangrao, Jiangxi Province in 1979. Master's degree from South China University of technology. Senior engineer, now engaged in information planning and construction, the main research direction: computer technology, cloud technology, network security, etc.