Original Research Article

Innovative Research on the Public Cloud Computing Power Platform

Junbin Le, Hong Niu, Dong Lin

Zhejiang Wanwang Information Technology Co., Ltd.

Abstract: Three key issues faced by the public cloud computing power platform in its operation process are discussed: low resource utilization, potential safety hazards, and unstable service quality. Corresponding optimization measures and improvement directions are proposed for these issues. First, by optimizing the resource allocation algorithm, reasonable scheduling and allocation of computing resources can be achieved to improve resource utilization and platform performance. Second, safety protection measures should be strengthened, starting from multiple levels such as physical safety, network safety, system safety, and data safety, to enhance the platform's safety protection capabilities. Third, service quality should be improved, starting from three aspects of service reliability, availability, and responsiveness, to optimize service processes and technical levels, and enhance user satisfaction and loyalty. The implementation of these measures will help the public cloud computing power platform achieve sustainable development and meet the growing needs of users.

Keywords: Public cloud computing power platform; Resource utilization rate; Safety hazard; Service quality

1. Introduction

With the rapid development and widespread application of cloud computing technology, public cloud computing power platforms play a crucial role in providing users with efficient and convenient computing services. However, during actual operation, these platforms face issues such as low resource utilization, security risks, and unstable service quality. These problems not only affect user experience and satisfaction, but also hinder the sustainable development of the platforms. Therefore, it is necessary to discuss these issues and propose corresponding optimization measures and improvement directions, in order to provide useful references for the development of public cloud computing power platforms.

2. Characteristics of Public Cloud Computing Power Platforms

2.1. High Scalability

One of the core features of public cloud computing power platforms lies in its high scalability, a characteristic that stems from the core advantage of cloud computing technology - virtualization. Through virtualization technology, public cloud computing power platforms can abstract and pool physical computing resources, enabling them to be accessed and utilized by users on demand, just like public resources such as water and electricity. Specifically, high scalability manifests in the following aspects: First, from the perspective of computing resources, public cloud computing power platforms can support vast amounts of computing resources, including both large-scale data centers and edge computing nodes[1]. These computing resources are dynamically allocated and scheduled to meet users' computing needs in different scenarios and at different times. When users require more computing resources, the platform can quickly expand the computing resource pool to meet their needs. Conversely, when users' computing demands decrease, the platform releases excess computing resources

to improve resource utilization. Second, from the perspective of service capabilities, public cloud computing power platforms can provide multiple types and scales of computing services, including virtual machines, containers, function computing, and other forms, to meet the individual needs of different users. Whether it's scientific research requiring large-scale parallel computing or online business requiring rapid response, public cloud computing power platforms can provide corresponding computing power. Third, from the perspective of system architecture, public cloud computing power platforms adopt a distributed, loosely coupled architecture design, enabling the platform to support horizontal and vertical scaling. Horizontal scaling means that the platform increases its overall computing power by adding more computing nodes and storage devices, while vertical scaling involves upgrading the hardware configuration of individual computing nodes to improve their individual computing power. This flexible system architecture allows public cloud computing power platforms to adapt to various complex computing scenarios and maintain continuous development and innovation.

2.2. High-Efficiency Calculation

Another significant characteristic of public cloud computing power platforms lies in their high-performance computing capabilities. High-performance computing refers to the utilization of powerful computing resources to complete large-scale and highly complex computing tasks in a very short time. By integrating a large number of high-performance computing resources and adopting advanced computing technologies and algorithms, public cloud computing power platforms achieve the goal of high-performance computing. Firstly, public cloud computing power platforms possess a vast computing resource pool, including high-performance computers, graphics processing units (GPUs), tensor processing units (TPUs), etc., which can meet various complex computing needs. Through virtualization technology, the platform centrally manages and schedules these computing resources, realizing the sharing and reuse of computing resources, thus improving the utilization rate of computing resources. Secondly, public cloud computing power platforms adopt advanced computing technologies and algorithms, including parallel computing, distributed computing, big data processing, artificial intelligence, etc., which can optimize and accelerate different types of computing tasks. For example, for scientific research requiring large-scale parallel computing, the platform utilizes parallel computing technology to decompose computing tasks into multiple subtasks and execute them in parallel on multiple computing nodes, thus accelerating the computation speed[2]. For business scenarios requiring the processing of large amounts of data, the platform employs big data processing technology to quickly process and analyze the data, extracting valuable information. Thirdly, public cloud computing power platforms also provide rich computing service interfaces and tools, which enable users to conveniently access the platform's computing resources for efficient computing tasks. Users can submit computing tasks to the platform for execution by writing programs, calling APIs, and other methods, and obtain the computation results. The platform also provides visualization tools to help users monitor the execution status of computing tasks, enabling them to identify and resolve issues promptly.

2.3. Safety and Reliability

The security and reliability of public cloud computing power platforms in a cloud computing environment are the cornerstones of their widespread adoption. With the rapid development of cloud computing technology, data security and system stability have become the top concerns for users. Public cloud computing power platforms ensure the security of user data and the stable operation of the system through a series of security measures and technical means. First, public cloud computing power platforms adopt a multi-layered security protection mechanism, covering aspects such as physical security, network security, system security, and data

security. Physical security involves strict access control and physical isolation to prevent unauthorized access and damage. Network security utilizes firewalls, intrusion detection systems, and other measures to defend against external network attacks and threats. For system security, the platform employs identity authentication, permission management, and other measures to ensure that only legitimate users can access system resources. For data security, the platform adopts data encryption, backup, and recovery technologies to safeguard the security and integrity of user data. Second, public cloud computing power platforms possess strong fault tolerance and redundant design. To ensure the stable operation of the system, the platform utilizes technologies such as load balancing and fault-tolerant backup. When a computing node or storage device fails, the system can automatically switch tasks to other normal nodes for continued execution, ensuring that users' computing tasks are unaffected. Additionally, the platform employs data redundancy backup technology to back up user data on multiple storage devices to prevent data loss or damage. Third, public cloud computing power platforms have established a comprehensive security management system and emergency response mechanism. By formulating strict security management policies and operational norms, the platform ensures that all security measures are effectively implemented. Simultaneously, the platform has established an emergency response team to promptly respond and handle security incidents, minimizing their impact on users and the system.

3. Problems of the Public Cloud Computing Power Platform

3.1. Low Resource Utilization Rate

In the operation process of public cloud computing power platforms, a significant issue is the low resource utilization rate, which mainly stems from multiple aspects. First, from the perspective of platform management, due to the lack of effective resource scheduling and allocation mechanisms, public cloud computing power platforms often fail to achieve optimal status in the allocation of computing resources. This results in some computing resources being idle during specific time periods, while other users are unable to complete their computing tasks due to insufficient resources. This imbalance in resource allocation directly affects the overall resource utilization rate of public cloud computing power platforms. Second, from the perspective of user behavior, users' computing needs are diverse and dynamic[3]. Users choose different computing resources and modes based on the characteristics of their computing tasks. However, due to the lack of precise prediction and matching of user computing needs, public cloud computing power platforms often struggle to allocate resources based on users' actual needs, leading to low resource utilization. Additionally, public cloud computing power platforms face several technical challenges in resource management, such as how to achieve dynamic monitoring and scheduling of computing resources and how to ensure load balancing of computing resources. The existence of these technical challenges makes it difficult for public cloud computing power platforms to achieve the ideal state in resource management, further reducing resource utilization.

3.2. Safety Hazards

As an important component of cloud computing services, the security risks faced by public cloud computing power platforms cannot be ignored. These security risks not only concern the security of the platform itself, but also involve the confidentiality and integrity of user data. First, public cloud computing power platforms often carry a large amount of user data and computing tasks, including sensitive information such as personal privacy and business secrets. If the security protection measures of the platform are not rigorous enough and are compromised by hackers or malicious users, it will lead to data breaches and misuse, causing significant losses

to users and the platform. Second, there are also security risks in the data transmission and storage processes of public cloud computing power platforms. During data transmission, if insufficiently secure encryption and transmission protocols are not used, hackers can intercept and steal data. In terms of data storage, if the platform fails to provide robust encryption technology and access control mechanisms, users' data will be at risk of unauthorized access and tampering[4]. Furthermore, public cloud computing power platforms also face internal security threats. Due to the large number of computing resources and complex system architectures, the platforms are difficult to manage and maintain. If administrators or maintenance personnel fail to strictly comply with security regulations and operational procedures, or if there are oversights and errors, security risks may arise, such as misuse of administrator privileges and misoperations, which pose risks to the platform's security. At the same time, with the rapid development and widespread application of cloud computing technology, new security threats and challenges are constantly emerging. For example, with the popularization of the Internet of Things (IoT) and big data technology, more and more devices and data are being connected to public cloud computing power platforms, posing greater challenges to the platform's security protection. Ensuring the security of these devices and data and preventing malicious exploitation and attacks has become an important issue facing public cloud computing power platforms.

3.3. Unstable Service Quality

During the operation of public cloud computing power platforms, unstable service quality is an issue that cannot be ignored. The existence of this issue not only affects users' experience but also poses challenges to the platform's reputation and sustainable development. First, the instability of service quality manifests in fluctuations in computing performance. Since public cloud computing power platforms handle a large number of computing tasks, when there is a surge in concurrent computing tasks, the platform's processing capacity can be affected, resulting in slow or failed execution of some users' computing tasks. This fluctuation in computing performance makes it difficult for users to obtain stable and reliable computing services, thereby affecting their user experience. Second, the instability of service quality also reflects in the availability and reliability of the services. As a complex system, public cloud computing power platforms have intricate dependencies between their internal components and services. Once a component or service fails, it can trigger a chain reaction, leading to the unavailability or degraded performance of the entire platform's services. Such issues with availability and reliability make it difficult for users to trust the platform, reducing their willingness and loyalty to use it. Additionally, the instability of service quality is also influenced by external factors, such as fluctuations in the network environment and the aging of hardware equipment, which can lead to a decline in the platform's service quality. These factors are unpredictable and uncontrollable, making it difficult for the platform to completely avoid issues of unstable service quality.

4. Countermeasures of the Public Cloud Computing Power Platform

4.1. Optimize the Resource Allocation Algorithm

In the operation of public cloud computing power platforms, optimizing resource allocation algorithms is a crucial approach to improving resource utilization and platform performance. With the rapid development and widespread application of cloud computing technology, the demand for computing resources is increasing, and efficiently managing and allocating these resources has become an urgent issue. Traditional resource allocation algorithms often rely on static or simple dynamic allocation strategies, which are insufficient to meet complex

and variable computing demands. Therefore, optimizing resource allocation algorithms has become an important research direction. The goal of optimizing resource allocation algorithms is to achieve rational allocation and scheduling of computing resources based on users' computing needs, the system's load status, and the real-time status of resources. Firstly, optimizing resource allocation algorithms needs to consider users' computing needs. Different users have different computing needs and priorities. For example, some users require significant computing resources to process big data or perform high-performance computing, while others only need a small amount of resources to run simple applications. Therefore, optimized resource allocation algorithms need to accurately identify and assess users' computing needs and allocate resources reasonably based on these needs. Secondly, optimizing resource allocation algorithms also needs to consider the system's load status. Public cloud computing power platforms often handle a large number of computing tasks, and the system's load status changes constantly. Optimized resource allocation algorithms need to monitor and evaluate the system's load status in real-time and dynamically adjust resource allocation strategies based on the load situation. For instance, when the system load is high, the algorithm should prioritize allocating resources to tasks with higher priorities to ensure timely completion of important tasks. When the system load is low, the algorithm can allocate remaining resources to other tasks to improve resource utilization. In addition, optimizing resource allocation algorithms needs to consider the real-time status of resources. The resource status of public cloud computing power platforms also changes constantly. For example, some computing nodes may be temporarily unavailable due to failures or maintenance. Optimized resource allocation algorithms need to obtain resource status information in real-time and allocate resources based on the actual situation of the resources. This ensures that resources are fully utilized and avoids assigning tasks to unavailable resources.

4.2. Strengthen Safety and Protection Measures

In the operation of public cloud computing power platforms, strengthening security measures is a crucial aspect of ensuring the safety of platform and user data. With the widespread application of cloud computing technology, security threats such as data breaches and unauthorized access are increasing, making it particularly important for public cloud computing power platforms to enhance their security measures. Firstly, strengthening security measures requires a multi-faceted approach encompassing physical security, network security, system security, and data security. Physical security ensures the safety of computing devices, data centers, and other physical facilities, preventing unauthorized access and destruction. Network security involves using technologies such as firewalls and intrusion detection systems to defend against attacks and threats from external networks. System security involves measures such as user authentication and access control to ensure that only legitimate users can access system resources. Data security includes techniques such as data encryption, backup, and recovery to safeguard the confidentiality and integrity of user data. Secondly, strengthening security measures requires a comprehensive defense strategy employing multiple layers and methods. Single security measures are often insufficient to address complex and evolving security threats. Therefore, it is necessary to integrate multiple technologies and approaches to form a multi-layered defense system. For example, using identity authentication and access control mechanisms to ensure the legitimacy of user identities and the rationality of access permissions, while employing data encryption techniques to encrypt user data to prevent data breaches and unauthorized access. Additionally, leveraging security audit and log analysis technologies to monitor and alert the platform's security status in real-time, enabling timely detection and response to potential security threats. Moreover, strengthening security measures also necessitates a focus on security management and personnel

training. The improvement and implementation of security management systems are the foundation for ensuring platform security. It is essential to establish sound security management systems and operational norms, and to strengthen the training and supervision of security management personnel. At the same time, it is necessary to enhance users' security awareness education and improve their ability to identify and prevent security threats.

4.3. Improve Service Quality

In the operation of public cloud computing power platforms, enhancing service quality is a crucial way to increase user satisfaction and boost platform competitiveness. As the cloud computing market matures and user needs diversify, users' demands for service quality are also rising. Therefore, public cloud computing power platforms need to continuously optimize service processes and upgrade technical capabilities to meet the increasing demands of users. Firstly, improving service quality requires addressing three aspects: reliability, availability, and responsiveness. Service reliability refers to the platform's ability to provide stable and accurate computing services, avoiding service interruptions or performance degradation due to system failures or insufficient resources. The platform should strengthen the maintenance and management of system hardware and software to ensure system stability and reliability. Simultaneously, the platform should establish a comprehensive fault detection and recovery mechanism to quickly respond and handle system failures, minimizing service interruption time. Secondly, service availability refers to the platform's ability to provide continuous and uninterrupted computing services to meet users' computing needs. The platform should strengthen the monitoring and management of resources to ensure adequate availability. Additionally, the platform should offer diverse computing resources and flexible configuration options to meet different users' computing needs[5]. By providing high-performance computing resources and superior service support, the platform can increase user satisfaction and loyalty. Thirdly, service responsiveness refers to the platform's ability to promptly and accurately respond to user requests and issues. The platform should establish a comprehensive customer service system, providing multiple contact methods and channels to facilitate users' access to help and support. At the same time, the platform should enhance the analysis and handling of user issues, improving the speed and efficiency of problem resolution. By promptly responding to user requests and resolving issues, the platform can enhance users' trust and satisfaction. Furthermore, improving service quality also requires attention to user experience and feedback. The platform should continuously optimize user interfaces and operational processes to enhance users' convenience and comfort. Additionally, the platform should actively collect user feedback and suggestions, promptly improving and optimizing service processes and technical capabilities. By continuously enhancing user experience and satisfaction, the platform can win the trust and support of more users, thereby enhancing its competitiveness and market share.

5. Conclusion

During its operation, the public cloud computing power platform faces various challenges, including low resource utilization, security risks, and unstable service quality. To address these issues, strategies such as optimizing resource allocation algorithms, strengthening security measures, and improving service quality have been proposed. By optimizing resource allocation algorithms, the platform can achieve reasonable allocation and scheduling of computing resources, thereby improving resource utilization and platform performance. By strengthening security measures, the platform can ensure the safety of user data and systems, reducing security risks. By improving service quality, the platform can enhance user satisfaction and loyalty, further boosting its competitiveness.

References

[1] Hao Junhui. Shanghai Telecom Launches Public Computing Power Service [J]. Posts and Telecommunications Economy, 2022(3): 12-16.

[2] Yang Xin, Shen Wenhai. The Architecture of Meteorological Computing Power Network and Computing Power Service System under the Trend of "Eastward Data and Westward Computing" [J]. Frontiers of Data and Computing Development, 2022, 4(5): 50-59.

[3] Shen Shikuai, Wang Guangquan, Tang Xiongyan. All-Optical Computing Power Network Opens Up the Main Artery for Computing Resource Transportation [J]. Communication World, 2022(6): 31-32.

[4] Li Shaohe, Li Taixin, Zhou Xu. Computing Power Network: Network-Centered Integrated Resource Supply [J]. ZTE Communications, 2021, 27(3): 6.

[5] Jiang Lintao. Cloud Computing, Edge Computing, and Computing Power Network [J]. Information and Communications Technologies, 2020(4): 4-8.