Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

Application of Key Technologies of Cloud Computing Energy Saving in IT Support System

Bochao Zhang

Guangdong Telecom Planning and Design Institute Co., Ltd. Guangzhou 510630, Guangdong Province

Abstract: ***This article introduces cloud computing technology, analyzes the application principles of energy-saving key technologies in cloud computing for IT support systems, and dissects the practical energy efficiency of these key technologies. By examining cloud-based business scenarios and analyzing the basis and algorithms for resource scheduling, intelligent power management contributes to reducing host power consumption during data center operation. The computational demands of business operations are positively correlated with energy consumption, and these demands can vary due to business requirements. Creating an energy-saving scheduling model and implementing it within the IT support cloud platform helps address energy-saving and emission reduction issues in cloud computing. Furthermore, the key energy-saving technologies in cloud computing enable flexible implementation of resource scheduling.***

Keywords: IT Support System; Cloud Computing; Key Technologies for Energy Conservation; Application.

1. INTRODUCTION

In recent years, with the continuous optimization and improvement of cloud computing technology, the dependence of data center construction on IT equipment has also increased, and the scale of data centers has become increasingly large, making energy consumption a key issue that data centers need to consider. Since 2011, cloud computing management platforms have been used in IT support systems. Through cloud platforms and virtualization construction, cloud computing management platforms have also improved the operational efficiency of data centers. Based on this environment, the hardware scale that supports cloud platforms has also developed rapidly. How to scientifically and reasonably utilize energy is a key issue that the public needs to consider. There are many ways to save operating costs in the process of data center operation, which are generally applied to the following:

(1) Applying virtualization technology to improve resource utilization and save data center operating costs;

(2) Create a virtualized resource pool to decouple physical servers and their applications;

(3) Flexible deployment to improve the efficiency of data center operation and maintenance management.

This article introduces cloud computing technology, analyzes the application principles of key energy-saving technologies in IT support systems, and dissects the actual energy efficiency of key energy-saving technologies in cloud computing applications.

2. OVERVIEW OF CLOUD COMPUTING TECHNOLOGY

Cloud computing is the third IT industry revolution after the reform of computers and the Internet, which can generally divide the development of information technology into three stages.

The first stage is the era of computer transformation, which optimizes large mainframes that are only used in specific industries into affordable computers for individuals, achieving effective improvement in both enterprise efficiency and personal work efficiency.

The second stage is the Internet reform period. Gathering billions of information islands into a vast information network has improved the efficiency of human communication, collaboration, and sharing, while also enriching the entertainment and social channels for the public.

Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

The third stage is the cloud computing transformation stage, which involves transforming ITTI service objects and basic resources into social public infrastructure, turning cloud computing into an IT service and resource provider.

Wang et al. [1] developed a comprehensive cell atlas of the immune microenvironment in gastrointestinal cancers, providing insights into dendritic cell functions. Transitioning to urban logistics, Wang et al. [2] explored AI applications for optimizing last-mile delivery efficiency in smart cities. Healthcare innovations continue with Li et al. [3]'s machine learning approach for adverse event monitoring in drug trials and Yuan et al. [4]'s GPT-4 application for processing multimodal medical data in EHR systems. The intersection of human-computer interaction and industrial systems is represented by Song et al. [5]'s work on optimizing warehouse management interfaces through speech recognition technology. Data management challenges are addressed by Chen et al. [6] through a quantized framework for gig economy platforms, while Wang et al. [7] examined legal aspects of enterprise naming rights. Risk management solutions appear in Gong et al. [8]'s ensemble machine learning approach for enterprise decision support systems. Computer vision applications are advanced by Bohang et al. [9] through their active learning method for image steganalysis. Industrial optimization is demonstrated in Zhao et al. [10]'s deep learning approach for steel production scheduling and Yao et al. [11]'s innovative drone-3D printing system for post-disaster construction. Economic applications include Yang et al. [12]'s big data-based prediction method for economic cycles and Ji et al. [13]'s AI-driven retail strategy optimization. Financial technology innovations feature Yang et al. [14]'s LLM-integrated system for cross-asset risk monitoring. Healthcare research continues with Li et al. [15]'s optimized clinical trial strategy for anti-HER2 drugs and Peng et al. [16]'s investigation of IoT-enhanced cognitive training for middle-aged adults.

3. THE IMPACT OF CLOUD COMPUTING ON THE CONSTRUCTION OF IT SUPPORT SYSTEMS

3.1 Impact on IT system construction mode

In the past, IT support systems were built in a top-down model, purchasing system software landscape opinions based on business needs, keeping business requirements limited, and creating a "chimney style" system. Cloud computing construction, on the other hand, adopts a bottom-up construction model. The construction of cloud computing resource pools should be prioritized over system requirements and no longer bundled with business applications. Application system upgrades, construction, and expansion are mainly for software development. If some hardware physical resources are needed, an application needs to be made to the resource pool. The cloud computing resource pool is shown in Figure 1.

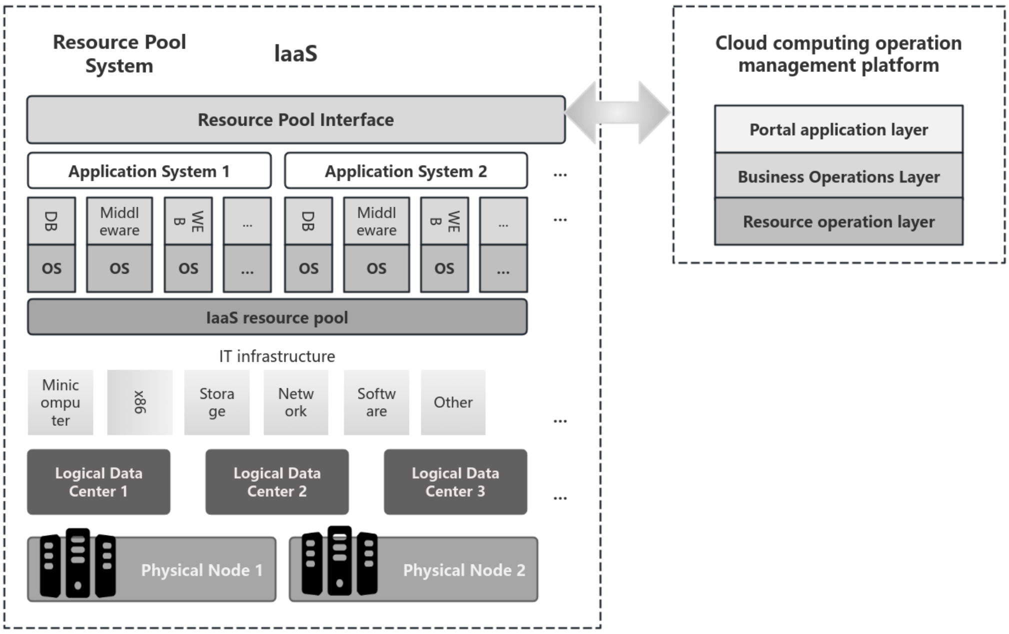

Figure 1: Cloud Computing Resource Pool

Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

3.2 Impact on departmental job functions

The traditional function of the IT department is mainly to understand business requirements, propose business implementation solutions, and install and deploy new equipment before the system limit to achieve real-time operation monitoring of system equipment. The IT department based on cloud computing focuses on service and resource operation management, proposing installation and deployment of new equipment resources and platform expansion plans. The issues of concern include: the degree of satisfaction of business requirements, the situation of newly deployed business applications, service quality assurance, the application of resources for business applications, and whether resource reserves meet the needs of business development.

4. KEY ENERGY-SAVING TECHNOLOGIES FOR CLOUD COMPUTING IN IT SUPPORT SYSTEMS

Firstly, integrate physical resources and establish a shared resource pool. Then, based on the current system load and resource application, activate intelligent power management and dynamic resource mobilization to achieve energy-saving and emission reduction goals in cloud computing.

4.1 Integrating servers

Because the development speed of hardware is significantly faster than that of software systems, in most cases, the utilization rate of application servers is very low. However, with the application of server virtualization technology, this situation has been improved. Generally, virtualization can be divided into two forms: multi virtual one and one virtual multiple. One virtual multiple refers to the abstraction of physical servers as multiple logical servers isolated from each other, while multiple virtual one refers to several physical servers cooperating with each other to process the same business, that is, distributed computing. The server virtualization technology in this article mainly refers to one virtual multi form.

Traditional physical servers mainly run a system, while virtualized physical servers can virtualize independent systems of each other. After virtualization integration is completed, it will greatly reduce the number of physical servers and IT equipment energy consumption. Research has found that reducing physical servers not only helps to save electricity, but also helps to conserve resources such as air conditioning and cooling in data centers, data center space, and equipment maintenance. In the process of integrating servers, full consideration should be given to the system's processing capacity, that is to say, to avoid a single virtual machine resource reaching the upper limit of the host machine resource, in order to meet the high availability requirements of the system, a certain amount of resources should be reserved in advance in the resource pool.

4.2 Scheduling dynamic resources

When resources are actually applied, capacity and demand will change over time, resulting in issues such as resource overload and idle resources. After optimizing resource scheduling management, resources can be reallocated. Generally, resource scheduling can be divided into two types: infrastructure level scheduling at the IaaS layer and job level scheduling at the PaaS layer. Job level scheduling is a comprehensive process of resource allocation and scheduling based on the running program at the PaaS layer. Facility level scheduling refers to using underlying resources as infrastructure and providing them to users, at which point users can effectively utilize the resources according to their actual situation.

The resource management and scheduling scheme can provide important support for saving IT support system resources. At present, IT support system resource management and scheduling mainly include load strategy, decentralized strategy, elastic expansion algorithm, fill strategy, load balancing adjustment algorithm, high availability strategy, and preload algorithm.

(1) Based on load strategy.

Usually, virtual machines are first installed on lightly loaded physical servers to improve application running efficiency. The main factor that affects host deployment priority based on load policy is host load.

(2) Based on a decentralized strategy.

A large number of physical servers will install virtual machines in a dispersed manner, which helps to reduce the impact of physical server failures and improve the efficiency of application running. The factors that affect host deployment priority based on a decentralized strategy include host priority, host CPU capacity, and host memory capacity.

(3) Based on the filling strategy.

Installing virtual machines in a small number of physical servers ensures that the servers can maximize the utilization of virtual machines, which not only reduces resource fragmentation, but also dynamically starts and shuts down the servers according to demand to achieve energy-saving and emission reduction goals.

(4) Based on high availability strategy.

Allocate physical server resources to critical businesses to improve resource availability. The main factor that affects host deployment priority under high availability is server type.

(5) Elastic expansion algorithm.
The so-called elastic expansion refers to the increase or decrease in the number of virtual machine instances that support cloud applications during the service provision process. In other words, under high business load conditions, a large number of virtual machine instances are started, while under low business load conditions, virtual machine instances are paused.

(6) Load balancing adjustment algorithm.
Fundamentally speaking, the main purpose of balancing and adjusting resource load is to ensure that computing in cloud computing can be distributed to various resources to improve resource utilization and save task execution time. The specific implementation method is to migrate the virtual machine with the lightest absolute load among the physical hosts with a load higher than the "balanced load threshold" in the pool to a physical host with a load lower than the "balanced load threshold", and ensure that the physical host load avoids being higher than the "balanced load threshold".

4.3 Intelligent power management

In the IT support system resource pool, in order to save energy consumption, dynamic resource scheduling is used to achieve resource load balancing, or combined with dynamic power management and association rules, virtual machines that comply with energy-saving strategies are migrated from the current host to another host, and then with the help of intelligent power management, idle hosts are shut down. Based on rules, if the existing host cannot support the current business, the host can automatically wake up. This process should be achieved in collaboration with technologies such as online migration of virtual machines, power management, and association rules. According to resource utilization, the resource pool can dynamically shut down or turn on the host power to avoid resource waste. Dynamically monitor the load of virtual machine resources in the resource pool, compare the available resources of hosts in the resource pool, and if the additional capacity of resources is sufficient, shut down the host to ensure it is in standby mode. Transfer the virtual machine of the host to be shut down to an available resource host, and then turn off the power. In the case of insufficient capacity, the host will be turned on through the dynamic resource scheduling module.

5. THE PRACTICAL APPLICATION AND EFFECTIVENESS OF ENERGY-SAVING SKILLS TECHNOLOGY

5.1 Practical application of energy-saving technology

The key technology of cloud computing is used in data centers to achieve energy-saving and emission reduction goals. Its actual application process is as follows:

(1) Integrate data center services that have expired or have low configuration servers, and replace them with high-performance servers;

(2) By integrating high-performance servers, resource pooling can be achieved;

(3) Set rules for associating hosts and virtual machines in resource pool settings;

(4) Selecting dynamic resource scheduling for resource pools and calculating high and low threshold values for resources;

(5) Enable power management for the resource pool host;

(6) Develop power management and resource scheduling strategies.

In the idle state of the business, through real-time migration, all virtual machines of low load hosts are migrated to other hosts that have not reached the load limit, and idle hosts are shut down. When the business is busy, wake up the standby host and migrate the virtual machine to the newly started host. This process can be manually controlled or automatically executed according to the plan settings.

5.2 Energy Efficiency Analysis

Intelligent power management helps to save host power consumption during data center operation. Business computing requirements are positively correlated with energy consumption, and business computing requirements may vary due to business needs. By analyzing the test results in Table 1, it was found that it was difficult to fully utilize resources when DMP was not started. The total power consumption of the server was 1132W. After setting the DMP threshold and reallocating the virtual machine, the memory, power, and CPU utilization of the server in idle state were all improved, but the total power consumption decreased by 33%. Figure 2 shows the specific comparative data. So, dynamic power management of the system helps to save the total power consumption of the host, while achieving an improvement in server resource utilization and achieving the goal of energy conservation and emission reduction. The above data is only for power consumption savings. The energy savings generated by servers also include transmission cables, server room space, network routing and switching equipment, cabinet and rack space, etc.

Table 1: Dynamic Power Management Test Results

Host identification	CPU and memory utilization	DMP shuts down host power consumption	DMP threshold setting host power consumption
1	CPU：31.62% Memory：41.25%	225W	240W
2	Shut down	228W	0W
3	CPU：50.97% Memory：60.28%	227W	272W
4	Shut down	225W	0W
5	CPU：31.56% Memory：34.89%	227W	240W
Total power consumption		1132W	752W

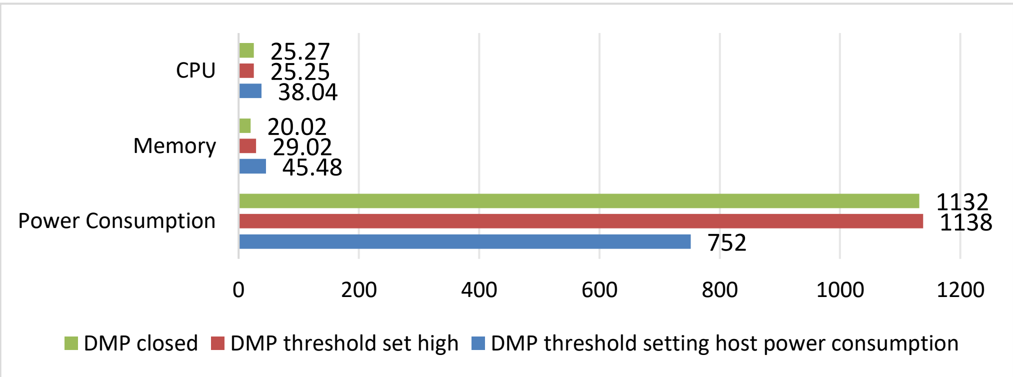

Figure 2: Energy saving effect of dynamic resource management

Woody International Publish Limited
An Multidisciplinary Academic Journal Publisher
Journal of Artificial Intelligence and Information, Volume 2, 2025
https://www.woodyinternational.com/

6. CONCLUSION

In summary, the effective application of cloud computing skills and key technologies supported by IT cloud platforms can contribute to energy conservation and consumption reduction. Sorting out cloud business scenarios, analyzing resource scheduling criteria and scheduling algorithms, creating an energy-saving scheduling model, and applying this model to IT supported cloud platforms can help solve the problem of energy conservation and emission reduction in cloud computing. Moreover, key energy-saving technologies in cloud computing can flexibly achieve resource scheduling, especially in business support systems, which have strong practicality. Especially when the server room space is insufficient, the number of systems is large, and energy consumption is high, it can help improve resource utilization, scientifically control electricity consumption, and achieve economic benefits.

REFERENCES

[1] Wang, Y., Yang, T., Liang, H., & Deng, M. (2022). Cell atlas of the immune microenvironment in gastrointestinal cancers: Dendritic cells and beyond. Frontiers in Immunology, 13, 1007823.
[2] Wang, J. (2025). Smart City Logistics: Leveraging AI for Last-Mile Delivery Efficiency.
[3] Li, T. (2025). Enhancing Adverse Event Monitoring and Management in Phase IV Chronic Disease Drug Trials: Applications of Machine Learning.
[4] Yuan, J. (2024). Exploiting gpt-4 for multimodal medical data processing in electronic health record systems. Preprints, December.
[5] Song, X. (2024). Optimizing the human-computer interaction interface of warehouse management systems using automatic speech recognition technology.
[6] Chen, J. (2025). Data Quality Quantized Framework: Ensuring Large-Scale Data Integration in Gig Economy Platforms.
[7] Wang, H. (2024). The Restriction and Balance of Prior Rights on the Right of Enterprise Name.
[8] Gong, C., Lin, Y., Cao, J., & Wang, J. (2024, October). Research on Enterprise Risk Decision Support System Optimization based on Ensemble Machine Learning. In Proceeding of the 2024 5th International Conference on Computer Science and Management Technology (pp. 1003-1007).
[9] Bohang, L., Li, N., Yang, J. et al. Image steganalysis using active learning and hyperparameter optimization. Sci Rep 15, 7340 (2025). https://doi.org/10.1038/s41598-025-92082-w
[10] Zhao, H., Chen, Y., Dang, B., & Jian, X. (2024). Research on Steel Production Scheduling Optimization Based on Deep Learning.
[11] Yao, T., Jian, X., He, J., & Meng, Q. (2025). Drone-3D Printing Linkage for Rapid Construction of Sustainable Post-Disaster Temporary Shelters.
[12] Yang, W., Zhang, B., & Wang, J. (2025). Research on AI Economic Cycle Prediction Method Based on Big Data.
[13] Ji, F., Zheng, X., Xue, H., & Wang, J. (2025). A Study on the Application of Artificial Intelligence in Personalized Go-to-Market Strategy in Retail Industry.
[14] Yang, J., Tang, Y., Li, Y., Zhang, L., & Zhang, H. (2025). Cross-Asset Risk Management: Integrating LLMs for Real-Time Monitoring of Equity, Fixed Income, and Currency Markets. arXiv preprint arXiv:2504.04292.
[15] Li, T. (2025). Optimization of Clinical Trial Strategies for Anti-HER2 Drugs Based on Bayesian Optimization and Deep Learning.
[16] Peng, Y., Zhang, G., & Pang, H. (2025). Exploring the effects of IoT-enhanced exercise and cognitive training on executive function in middle-aged adults. Alexandria Engineering Journal, 120, 106-115.