European Journal of Business, Economics & Management

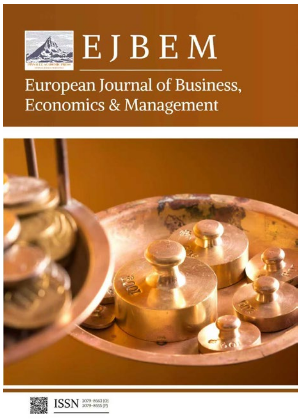
Vol. 1 No. 4 2025

Article **Open Access**

Cost Control and Stability Improvement in Enterprise Level Infrastructure Optimization

Yixin Zhou [1,*]

[1] Ads API Infra, Amazon, New York, 10001, US
* Correspondence: Yixin Zhou, Ads API Infra, Amazon, New York, 10001, US



Abstract: Optimizing enterprise-level infrastructure plays a key role in enhancing operational efficiency and saving costs. This article deeply analyzes the key factors and implementation strategies of infrastructure from the aspects of hardware resource optimization, automated operation and maintenance, and intelligent resource scheduling. With the deepening application of virtualization technology and cloud computing, enterprises can greatly improve resource utilization efficiency and stable system operation, while reducing maintenance costs and energy consumption. By implementing centralized monitoring and intelligent warning mechanisms, system issues can be quickly identified and effectively prevented, thereby enhancing the stability and reliability of infrastructure. With the rapid progress of cloud native architecture, artificial intelligence, edge computing, and green IT, the optimization of future infrastructure has more innovative possibilities. These strategies and Q trends provide a reference for efficient management of enterprises in the process of informatization and digitization, promoting the achievement of sustainable development goals.

Keywords: infrastructure optimization; cost control; system stability; virtualization technology; cloud native architecture

Received: 17 August 2025
Revised: 01 September 2025
Accepted: 25 September 2025
Published: 01 October 2025

1. Introduction

The infrastructure of an enterprise relies on a solid underlying architecture, and the performance and stability of this architecture are fundamental to forming the core competitive advantage of the enterprise. With the deepening of informatization and digital reform, traditional infrastructure is unable to cope with the constantly growing business demands and rapid market changes. In the context of limited resources, how to improve the operational efficiency of infrastructure is an urgent problem that enterprises need to solve. Enterprises can achieve innovation in the rational allocation and maintenance of resources by integrating advanced technology with reasonable operational management methods, thereby forming a competitive advantage. This article will provide a detailed analysis of the core strategies for improving infrastructure efficiency and look forward to future development trends.

2. Key Elements of Optimizing Enterprise-Level Infrastructure

2.1. Hardware Resource Optimization

In the improvement of enterprise infrastructure, the optimization of hardware resources plays a crucial role, with a focus on integrating resources, improving performance, and planning capacity. The integration of resources involves merging scattered hardware and server resources, reducing unnecessary equipment, and thus improving the effective utilization of resources. Performance optimization focuses on selecting and matching hardware, selecting high-performance CPUs, high-bandwidth memory, and solid-state drives based on specific business needs to maximize the potential of the hardware [1]. The planning of capacity needs to be combined with the estimation of enterprise business volume, making reasonable arrangements for storage, network bandwidth, and power supply to ensure sufficient supply of resources during peak periods.

2.2. Automated Operation and Maintenance

The key factors to improve the efficiency of automated operation and maintenance include automated task scheduling, automated management of system settings, and automated problem-solving processes. Task scheduling automation centrally controls and executes batch tasks through an operation management platform, significantly reducing the frequency of manual operations. The automated management of system settings relies on automation tools (Ansible, Puppet), which use configuration templates to uniformly manage server and application settings, reducing configuration errors caused by improper manual operations. The automated process of problem solving, combined with a log analysis and monitoring system, can quickly identify problems and perform automatic recovery operations, such as ensuring the continuous operation of the system through automatic restart and load balancing [2].

3. Cost Control and Stability Improvement Strategies in Optimizing Enterprise-Level Infrastructure

3.1. Intelligent Resource Scheduling and Automated Management

Efficient infrastructure operation relies on intelligent scheduling and automated management of resources, which requires the integration of dynamic task allocation strategies and mathematical models for resource optimization. The system responds to fluctuations in business demands by adjusting the layout of computing resources in real-time, prioritizing the execution of important tasks and reducing ineffective resource consumption. In practice, flexible partitioning techniques can be used to partition server nodes while adopting containerization technologies (Docker, Kubernetes) to deploy applications in lightweight containers, thereby improving resource isolation and utilization efficiency. By using resource optimization techniques such as particle swarm optimization or genetic algorithm, the optimal path for task allocation can be calculated, thereby reducing resource consumption [3]. By combining intelligent prediction algorithms, the system can predict resource demand based on historical data and trends, and adjust and allocate resources in advance to cope with fluctuating demands. For example, in the allocation of power resources, the task scheduling that minimizes costs can be calculated using the following formula:

$$MinimizeC = \sum_{i=1}^{n}(E_i \cdot P_i) + \sum_{j=1}^{m}(R_i \cdot U_j) \quad (1)$$

Among them, E_i is the energy consumption of task i, Pi is the unit energy consumption cost, R_j is the usage of resource j, and U_j is the unit resource usage cost. Through the above methods, intelligent resource scheduling and automated management can achieve dual improvements in infrastructure stability and cost efficiency for enterprises.

3.2. Application of Virtualization and Cloud Computing Technology

With the help of virtualization and cloud computing technology, significant improvements in infrastructure efficiency have been achieved, which involves centralized management of resources, flexible resource adjustments, and precise optimization of services. In terms of resource centralization, virtualization technology is used to merge actual physical servers, storage, and network resources into a comprehensive resource library, which can be dynamically allocated to various virtual machines according to different task requirements [4]. In terms of flexible adjustment of resources, cloud platforms can automatically increase or decrease resources based on changes in actual business volume, thereby automatically increasing computing resources during business peaks and releasing excess resources when demand decreases, in order to achieve optimal utilization of resources. The optimization of services is achieved through intelligent tools managed by cloud platforms (load balancers and container orchestration systems), which can accelerate the deployment process of applications and ensure high availability of applications. For example, a retail enterprise has adopted virtualization technology to integrate scattered server resources into a unified virtualization management, while combining the dynamic resource adjustment capability of cloud platforms to flexibly configure computing and storage resources according to business changes. The enterprise has also implemented container orchestration technology to further optimize the allocation process of resources, and through a centralized monitoring system to monitor the real-time operation status of the system, greatly improving the efficiency of infrastructure management and the effective utilization of resources. Table 1 summarizes the specific strategies and optimization effects of virtualization and cloud computing technologies in optimizing enterprise infrastructure.

Table 1. An extended table of optimization effects of virtualization and cloud computing technology.

technical means	Enhancement strategy	Save hardware costs (%)	Increase in average resource utilization rate (%)	System downtime rate reduced (%)
Resource pooling	Integrating physical servers and storage resources into a unified resource pool through virtualization technology	fifty	seventy-five	eighty-five
Elastic Expansion	Dynamic adjustment of computing and storage resources based on the cloud platform's on-demand deployment capability	forty-five	seventy	eighty
Service optimization	Utilizing cloud platform management tools to achieve load balancing and resource orchestration	forty	sixty-five	seventy-five
Centralized monitoring	Real-time monitoring of resource status and generation of automated alerts using monitoring tools	thirty-five	sixty	seventy

This table provides a detailed display of the specific methods and implementation effects of virtualization and cloud computing technology in optimizing enterprise infrastructure, providing a reference for enterprise optimization practices.

3.3. Centralized Monitoring and Intelligent Warning System

The centralized monitoring and intelligent warning system integrates scattered monitoring data to achieve global visualization and precise management. Using distributed monitoring software, such as Prometheus and Zabbix, to collect operational information of various hardware devices and display real-time data on an integrated platform, thereby enhancing information transparency. Among the many core processes, log processing is particularly critical. By using log analysis tools such as ELK Stack, the system automatically processes log data, detects abnormal patterns, and issues alerts promptly. The intelligent warning mechanism establishes a predictive algorithm model through machine learning, analyzes historical operating data to predict resource usage trends and potential risks, such as predicting possible high load situations through time series analysis. Combined with dynamic resource scheduling technology, the monitoring system can allocate resources in advance based on real-time changes in load to prevent system overload. To further optimize the warning mechanism, the system can use the following formula to measure warning efficiency:

$$E = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100 \quad (2)$$

Among them, E represents the warning efficiency, TP is the number of correct warnings, TN is the number of correctly judged no faults, FP is the number of incorrect warnings, and FN is the number of true faults that have not been triggered. Through this formula, the accuracy and comprehensiveness of the system's fault warning can be comprehensively evaluated, thereby optimizing the stability and operational reliability of infrastructure [5].

3.4. Regular Audit and Continuous Optimization

Regular audits and continuous optimization constitute key strategies to ensure the efficient operation of enterprise-level facilities. In the audit phase, it is necessary to conduct in-depth evaluations of the performance indicators of servers and other hardware, the stability of network transmission, and the effective use of storage resources, to identify ineffective resource consumption and performance-limiting factors. For example, using an automated monitoring system to regularly output detailed reports on resource consumption to identify inefficient areas. In the performance optimization phase, the focus is on adjusting the allocation plan of resources, adjusting system settings based on fluctuations in actual business volume, and preventing resource surplus or shortage. In terms of security, the risks faced by the system can be reduced through regular security vulnerability detection, permission setting review, and in-depth analysis of logs. The comprehensive use of these methods and the introduction of capacity management tools can help predict future hardware and storage resource requirements proactively, effectively reducing operational pressure caused by emergency needs. The following formula is used to evaluate the changes in resource utilization efficiency after optimization, providing a quantitative basis for continuous optimization strategies:

$$E = \frac{(R_a - R_i) \cdot T}{R_t} \quad (3)$$

Among them, E represents the efficiency of resource optimization, R_a is the actual resource usage after optimization, R_i is the actual resource usage before optimization, T is the optimization cycle, and R_t is the total amount of resources. This formula can comprehensively reflect the changes in resource utilization efficiency and the impact of optimization cycles, providing a reference for adjusting optimization strategies.

4. The Development Trend of Cost Control and Stability in Enterprise-Level Infrastructure Optimization

4.1. Popularization of Cloud Native Architecture

Cloud native architecture is gradually becoming popular in enterprise infrastructure, mainly reflected in the widespread application of containerization technology, microservices, and automated operations and maintenance. Containerization technologies (Docker and Kubernetes) have become standard tools for enterprises to publish and manage applications. Currently, over 70% of enterprises have adopted containerization technology in their production processes to optimize resource allocation efficiency and accelerate application deployment speed. In many industries, especially in the financial, retail, and Internet fields, microservices decompose a single application into multiple independent functional modules, which enables enterprises to expand on demand and improve maintenance efficiency. Service meshes such as Istio are increasingly being adopted by enterprises to handle communication issues between microservices and ensure system stability and performance. At the level of automated operations and maintenance, the cloud native Continuous Integration and Continuous Deployment (CI/CD) mechanism has been widely applied, greatly reducing the time from software development to deployment. Many enterprises have also achieved comprehensive monitoring and timely warning of system operation status by introducing cloud native log monitoring tools such as Prometheus. With the help of serverless architecture, some enterprises are trying to further simplify operation and maintenance work, only activating corresponding computing power when resource requirements arise. To evaluate the resource scheduling efficiency under containerization technology, the following formula can be used:

$$F = \frac{(\sum_{i=1}^{n} U_i^2)}{(\sum_{i=1}^{n} Ui)^2 \cdot n} \tag{4}$$

Among them, F represents the equilibrium coefficient of resource allocation, Ui is the actual resource utilization rate of the i-th service, and n is the total number of services. Through this formula, enterprises can evaluate the balance of resource allocation among services and optimize resource utilization strategies. This method helps to promote efficient management of enterprise infrastructure and provides directional support for subsequent adjustments and optimizations.

4.2. Application of Artificial Intelligence and Machine Learning in Optimization

In the process of optimizing enterprise architecture, artificial intelligence and machine learning technologies have been widely applied, especially in the fields of automated resource management, performance estimation, and fault diagnosis, where their role is particularly significant. By relying on deep learning methods, enterprises can extract patterns from large amounts of data, predict workloads, and dynamically adjust resources. These machine learning algorithms can monitor resource consumption in real-time, automatically calculate and allocate storage resources, thereby improving resource efficiency and preventing resource surplus or shortage. The application of artificial intelligence in anomaly monitoring and fault warning is equally important. It can predict possible problems by analyzing logs and performance parameters, so as to take measures in advance and reduce the risk of system downtime. For example, a global e-commerce platform has improved the allocation of its server resources using machine learning technology. By analyzing historical traffic data, it predicts future traffic and allocates resources accordingly. The enterprise also utilizes an AI-driven automated operation and maintenance system to monitor performance indicators, adjust load balancing, thereby accelerating system response time and improving user interaction experience. Figure 1 shows the popularity trend of artificial intelligence and machine learning technologies in enterprise optimization from 2015 to 2023, reflecting their process of becoming key factors in enterprise optimization.

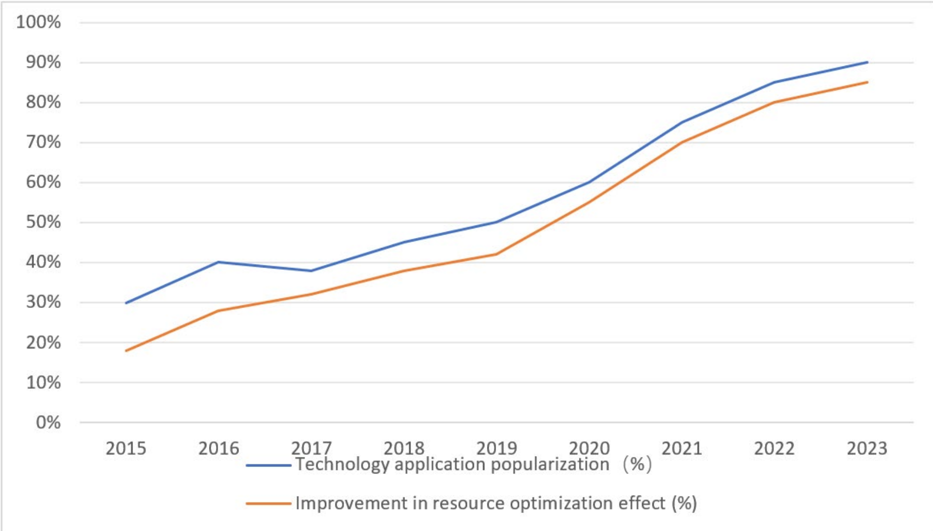


Figure 1. The prevalence trend of AI and machine learning technologies in enterprise optimization for a global e-commerce platform from 2015 to 2023.

From the graph, it can be seen that the application of artificial intelligence and machine learning in enterprise optimization has gone through an initial fluctuation stage, and then gradually showed a steady upward trend after 2020, indicating that enterprises' dependence on these technologies and their application effects have been increasing year by year.

4.3. The Rise of Edge Computing

With the rise of edge computing, the optimization strategy of enterprises for infrastructure is undergoing a revolutionary change, especially in terms of reducing data processing delay, improving immediate response capability, and achieving a distributed architecture, which highlights its advantages that cannot be ignored. Enterprises utilize edge devices such as IoT gateways and edge servers to process data locally, thereby reducing reliance on remote data centers. This is suitable for scenarios that require real-time analysis and quick decision-making. In industrial manufacturing, cargo transportation, and monitoring systems, edge computing improves processing efficiency and enhances the flexibility and adjustment ability of the system by deploying computing resources near data sources. For example, a logistics enterprise has introduced edge computing technology into its distribution network. It performs traffic data analysis and route planning tasks through edge devices throughout all nodes, realizing direct processing of computing tasks on localized devices, thus forming a rapid response scheduling system. The formula for calculating the resource allocation efficiency of edge computing is:

$$E_{opt} = \frac{D\cdot(\alpha\cdot P_e+\beta\cdot\frac{1}{T_e})}{(T_e\cdot N)+\gamma(L+C)} \quad (5)$$

Among them, E_{opt} is the comprehensive optimization efficiency of edge computing, D is the total data volume, P_e is the computing capacity of edge nodes, T_e is the time of single node processing tasks, N is the number of edge nodes, L is the transmission delay, C is the communication overhead α, β, γ is the weight coefficient, which is used to measure the importance of computing capacity, processing time and communication delay This formula provides an accurate evaluation method for the optimization effect of edge computing in complex systems by comprehensively considering the data volume, node performance, delay and communication overhead.

4.4. Green IT and Energy Optimization

In the process of upgrading and optimizing enterprise-level infrastructure, the application of green IT and energy optimization is particularly crucial, with the core goal being to reduce energy consumption and improve the effective utilization of resources. By deploying energy-efficient hardware and implementing virtualization strategies, enterprises have effectively reduced the number of server entities, thereby reducing energy demand at the root. With the help of Dynamic Power Management (DPM) technology, the system can automatically adjust power output based on actual workload, thereby reducing ineffective energy consumption. In the daily management of data centers, using liquid cooling technology instead of traditional air cooling systems greatly improves cooling efficiency and helps extend the service life of equipment. For storage systems, implementing a tiered storage scheme stores frequently accessed data in high-performance devices, thereby reducing storage energy consumption for infrequently accessed data. Building an energy monitoring system is also a key step, which can collect and analyze real-time energy consumption data of equipment, and then combine intelligent algorithms to optimize the allocation of equipment running time and load. These measures have jointly promoted the deep integration of green IT and energy optimization, achieved green and sustainable development of infrastructure, and ensured the stable and efficient operation of the system.

5. Conclusion

Optimizing enterprise-level infrastructure occupies a core position in the process of informatization, which directly affects the effective management of costs and the stability of system operation. Enterprises can achieve optimal allocation of resources and sustained stability of systems in ever-changing and complex business scenarios through strategies such as resource integration, automation management, virtualization technology, and green IT. With the wide application of cloud computing native architecture, artificial intelligence technology, edge computing, and other advanced technologies, infrastructure optimization is moving towards intelligence and sustainability. In the future, enterprises must constantly track technological updates and management innovations, flexibly adjust according to actual business needs, and adapt to market fluctuations and maintain competitive strength.

References

1. M. Rosenberg, "Which significant others?," *American Behavioral Scientist*, vol. 16, no. 6, pp. 829-860, 1973, doi: 10.1177/000276427301600603.
2. P. R. A. T. H. Y. U. S. H. A. Nama, "Cost management and optimization in automation infrastructure," *Iconic Research and Engineering Journals*, vol. 5, no. 12, pp. 276-285, 2022.
3. I. Rahman, P. M. Vasant, B. S. M. Singh, M. Abdullah-Al-Wadud, and N. Adnan, "Review of recent trends in optimization techniques for plug-in hybrid, and electric vehicle charging infrastructures," *Renewable and Sustainable Energy Reviews*, vol. 58, pp. 1039-1047, 2016, doi: 10.1016/j.rser.2015.12.353.
4. M. Kumar *et al.*, "Comprehensive review of electric vehicle technology and its impacts: detailed investigation of charging infrastructure, power management, and control techniques," *Applied Sciences*, vol. 13, no. 15, p. 8919, 2023, doi: 10.3390/app13158919.
5. N. Campos-Vasquez, R. K. Manturano-Chipana, J. L. Neyra-Torres, J. V. A. N. Jovanovic Aguirre, A. Carmona-Arteaga, and A. G. Maza-Chumpitaz, *et al.*, "Influence of the PMBOK in the optimization of budgets in road infrastructure projects in Lima," *LACCEI*, vol. 1, no. 8, 2023, doi: 10.18687/LACCEI2023.1.1.1342.