

Arhitekture računarstva u oblaku: modeli, evolucija i komparativna analiza savremenih distribuiranih sistema

Cloud computing architectures: models, evolution and comparative analysis

Lazar Stošić^{*1,2}, Olja Krčadinac¹

Faculty of Informatics and Computer Science, University Union—Nikola Tesla, Belgrade, Serbia¹;
Don State Technical University, Rostov-on-Don, Russian Federation²; lstosic@unt.edu.rs; okrcadinac@unionnikolatesla.edu.rs

Sažetak - Računarstvo u oblaku predstavlja dominantnu paradigmu za isporuku skalabilnih i fleksibilnih računarskih resursa putem Interneta. Brza evolucija cloud arhitektura—od tradicionalnih modela zasnovanih na virtualizaciji do savremenih paradigmi kao što su kontejnerizacija i serverless računarstvo—dovela je do povećane kompleksnosti u projektovanju i implementaciji sistema. Ovaj rad pruža sveobuhvatan pregled arhitektura računarstva u oblaku, sa fokusom na njihove modele, razvoj i komparativne karakteristike. Analizirani su modeli IaaS, PaaS i SaaS, kao i noviji pristupi poput kontejnera i serverless okruženja. Komparativna analiza zasnovana je na kriterijumima performansi, skalabilnosti, troškova i bezbednosti. Takođe su razmotreni aktuelni izazovi i budući pravci razvoja, uključujući edge računarstvo i primenu veštačke inteligencije u upravljanju resursima. Rezultati rada pružaju osnovu za razumevanje kompromisa između različitih arhitekturnih pristupa.

Ključne reči - računarstvo u oblaku; arhitektura; virtualizacija; kontejnerizacija; serverless; skalabilnost

Abstract — Cloud computing represents the dominant paradigm for delivering scalable and flexible computing resources over the Internet. The rapid evolution of cloud architectures—from traditional models based on virtualization to modern paradigms such as containerization and serverless computing—has led to increased complexity in system design and implementation. This paper provides a comprehensive overview of cloud computing architectures, focusing on their models, development and comparative characteristics. IaaS, PaaS and SaaS models are analyzed, as well as newer approaches such as containers and serverless environments. The comparative analysis is based on performance, scalability, cost and security criteria. Also discussed are current challenges and future development directions, including edge computing and the application of artificial intelligence in resource management. The results of the work provide a basis for understanding the trade-offs between different architectural approaches.

Keywords – *cloud computing; architecture; virtualization; containerization; serverless, scalability*

I. INTRODUCTION

Cloud computing represents a fundamental change in the way computing resources are provided, managed and used. Instead of relying on local infrastructure, organizations are increasingly using cloud solutions to increase flexibility, reduce costs and improve system scalability.

The development of cloud technologies has led to the emergence of different architectural models with specific characteristics and limitations. Early systems were based on virtualization, while modern approaches include microservices, containerization, and the serverless paradigm.

The aim of this work is:

- analysis of cloud architecture evolution
- identification of key models
- comparative evaluation of their characteristics

II. CLOUD SERVICE DELIVERY MODELS

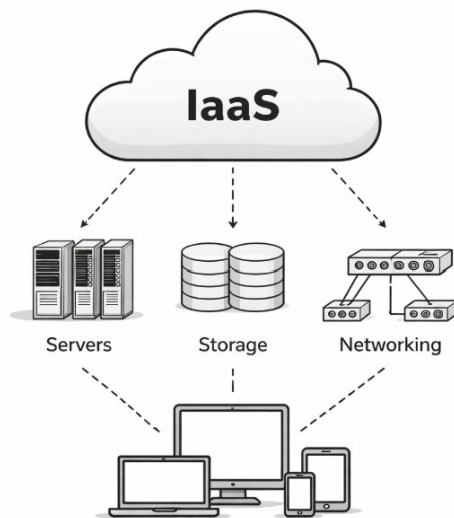
Cloud service delivery models define how IT resources and functionality are delivered to users over the Internet. Cloud computing is a model that uses a shared set of resources such as networks, servers, storage, virtual machines that can be quickly added and removed with minimal management overhead.[1] Basic models include IaaS (Infrastructure as a Service), which provides infrastructure (servers, networks, storage), PaaS (Platform as a Service), which provides a development environment and tools for applications, and SaaS (Software as a Service), where users directly use ready-made applications without infrastructure management.

These models allow for different levels of control, flexibility and accountability between providers and users,

thereby optimizing the use of resources and reducing IT system costs.

Service models

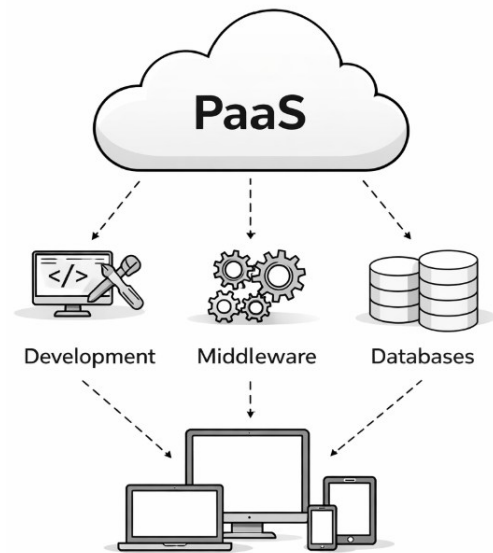
IaaS - (Infrastructure as a Service) represents the basic layer of the cloud model, which allows users to access virtualized computer resources, including virtual machines, storage and network infrastructure. The user has full control over the operating system, applications and configuration of the environment, while the cloud provider manages the physical infrastructure and virtualization layer. IT resources in the IaaS architecture are not preconfigured, which increases the administrator's responsibility. This model is best for cloud users who require more control over their environment. It happens that service providers sometimes use the services of other cloud servers in order to expand their own cloud. In the evolution of cloud architectures, the IaaS model marks the transition from physically bound systems to flexible, scalable and software-defined resources, while in a comparative analysis it provides the highest degree of control and responsibility compared to more abstract models such as PaaS and SaaS. The cloud-computing market is growing and the leaders in public- and private-cloud services can change. [2]



IaaS – Infrastructure as a Service

PaaS - (Platform as a Service) represents the middle layer of the cloud model that provides a development and execution environment for building, testing and implementing applications without the need to manage the basic infrastructure. The provider manages the operating system, middleware, databases and development tools, while the user focuses on developing the application logic. In the evolution of cloud architectures, the PaaS model enables acceleration of development and standardization of processes through automation and integrated services, while in a comparative analysis it represents a balance between flexibility (characteristic of IaaS) and ease of use (typical of the SaaS model). Briefly stated, PaaS is a "ready-to-use" model that contains already installed and configured IT

resources. PaaS also serves as a platform for the industrialization of the software industry. [3]



PaaS – Platform as a Service

SaaS - (Software as a Service) represents the highest level of abstraction in cloud models, where users access complete applications via the Internet, without the need for installation, maintenance or management of the infrastructure and platform. All system components, including application logic, data and security, are fully under the control of the service provider. The SaaS framework is a software distribution model that allows clients to access applications over a network, with hosting managed by the provider. This process is known as the most sophisticated manifestation of cloud computing. [4] In the evolution of cloud architectures, the SaaS model enables mass availability and standardization of software solutions, while in a comparative analysis it offers the highest level of ease of use with minimal control over the system compared to IaaS and PaaS models. In the market, SaaS is taking over cloud computing, and globally, industries are showing interest in creating their own SaaS add-ons and renting services to grow on their desktops.[5]



SaaS – Software as a Service

III. COMPARATIVE ANALYSIS OF CLOUD ARCHITECTURES

Performance is a key criterion in the evaluation of cloud architectures and refers to the system's ability to efficiently process requests with minimal delay and optimal use of resources. This criterion includes metrics such as latency, bandwidth, response time and throughput, which directly affect the quality of user experience and reliability of applications. In a comparative analysis, different architectures (e.g. virtualized, containerized, serverless) show variations in performance due to differences in abstraction layers and resource management. Modern cloud-native and serverless architectures enable performance optimization through automatic scaling and load distribution, but may introduce additional latencies in specific scenarios (e.g. cold start effect).

Scalability represents the ability of the cloud architecture to effectively adjust system capacity to growing or decreasing load requirements, without performance degradation. This criterion includes horizontal scaling (adding new instances) and vertical scaling (increasing the resources of existing instances), whereby modern cloud-native and serverless architectures favor automated horizontal scaling. In a comparative analysis, container and microservice architectures enable finer granularity and faster response to load changes compared to traditional and virtualized systems. Scalability is a key indicator of cloud system efficiency, as it directly affects availability, cost optimization and the ability of the system to respond to the dynamic requirements of modern applications.

Costs represent a key criterion in the comparative analysis of cloud architectures and refer to the total expenses associated with the implementation, maintenance and scaling of the system. In the cloud environment, the billing model based on actual consumption (pay-as-you-go) dominates, which enables the optimization of operating

costs compared to traditional infrastructures with high capital investments (CAPEX). In a comparative analysis, serverless and cloud-native architectures often enable more efficient cost management thanks to automatic scaling and elimination of unused resources, while IaaS models can generate higher costs due to the need for more detailed resource management. Effective cost management requires continuous monitoring and optimization of consumption, especially in complex and distributed cloud environments.

Security represents one of the most critical criteria in the evaluation of cloud architectures and refers to the system's ability to protect data, applications and infrastructure from various threats and vulnerabilities. This criterion covers aspects such as authentication and authorization, encryption, data integrity, network security and access management. In a comparative analysis, different cloud models offer different levels of control and responsibility within the shared responsibility model, where the private cloud allows a higher level of control, while the public cloud provides advanced security mechanisms through standardized and automated services. Modern architectures, such as cloud-native and serverless models, integrate security through the principles of security-by-design and Zero Trust, but require careful configuration management to avoid vulnerabilities due to system complexity.

Complexity refers to the level of technical complexity in the design, implementation, and management of cloud architecture, including the number of components, system interdependencies, and required maintenance competencies. This criterion includes factors such as infrastructure configuration, service orchestration, integration of different technologies, and application lifecycle management. In a comparative analysis, traditional and monolithic architectures have a lower level of initial complexity, but limited flexibility, while cloud-native, container and multi-cloud architectures introduce significantly greater complexity due to their distributed nature and the need for automation. Although increased complexity enables greater scalability and agility, it requires advanced DevOps practices, standardization, and appropriate tools for effective system management.

IV. FUTURE RESEARCH DIRECTIONS

Autonomous cloud systems represent an advanced research direction in which cloud infrastructure and services have the ability to manage themselves without direct human intervention, relying on artificial intelligence and machine learning. These systems integrate functionalities such as self-configuration, self-optimization, self-healing and self-protection, thus achieving a high level of adaptability and resilience. In an evolutionary context, the autonomous cloud represents the next stage of development after automated and AI-assisted systems. In a comparative analysis, this approach promises a significant improvement in performance, reliability and efficiency of resource management, but at the same time it raises issues of transparency, control and ethical implications of decision-making without human supervision.

Optimization of serverless environments represents an important research direction aimed at improving performance, reducing latency and more efficiently managing costs in Function-as-a-Service (FaaS) models. Key challenges include minimization of cold start latency, optimization of function scheduling, as well as efficient resource management under dynamic and unpredictable load conditions. In an evolutionary context, the focus is shifting towards intelligent mechanisms that use predictive models to proactively scale and cache execution environments. In a comparative analysis, an optimized serverless architecture can achieve a high level of efficiency and scalability with minimal operating costs, but requires advanced management techniques and balancing between performance and resource consumption.

AI orchestration represents an advanced approach to managing cloud and distributed systems in which artificial intelligence is used for automatic coordination, deployment and optimization of resources and services. This concept integrates machine learning into orchestration processes (e.g. in container and microservices environments), enabling adaptive decision-making based on real-time data and historical patterns. In an evolutionary context, AI orchestration marks the transition from static and rule-driven systems to dynamic and self-learning platforms. In a comparative analysis, this approach enables improvement of performance, reliability and resource utilization, but introduces additional complexity in terms of model interpretability, data management and control over automated decisions. Artificial Intelligence of Things (AIoT) relies on a service-oriented IoT architecture and at the same time supports decentralized, dynamic management of large-scale AI. With AI control, in AIoT, every component of a product or cost of knowledge is AI-based.[6]

Quantum computing in the cloud represents an innovative direction that enables access to quantum processors via cloud infrastructure, which eliminates the need for expensive and complex local quantum equipment. This model allows researchers and organizations to experiment with quantum algorithms (e.g. for optimization, cryptography and simulations) using hybrid classical-quantum approaches. In the evolutionary context, the quantum cloud represents a potential disruptive phase of computing development, which can significantly change data processing paradigms. In a comparative analysis, although the technology is still at an early stage and faces limitations such as quantum bit errors and a limited number of qubits, its potential for solving complex problems exceeds the capabilities of classical cloud systems in specific domains. Vendors should try to allocate machine resources as efficiently as possible to improve system availability, while customers should try to use efficient resource allocation strategies to schedule jobs and maximize uptime. To achieve all of this, it is crucial to understand the characteristics of quantum job execution, as well as those on cloud machines.[7]

V. CONCLUSION

The paper provides a systematic overview of cloud computing architectures and their evolution. The results indicate that modern models, especially container and serverless approaches, offer significant advantages in terms of scalability and flexibility, but introduce new challenges in the area of security and system complexity. Future research should be directed towards the automation and sustainability of cloud environments.

Analysis of models, architectures and modern trends in cloud computing confirms that this paradigm represents the key foundation of the digital transformation of modern information systems. The evolution from traditional IT infrastructure, via virtualization, to cloud-native, container and serverless architectures indicates a continuous shift towards greater flexibility, automation and efficiency of resource management. At the same time, the development of different service models (IaaS, PaaS, SaaS) and implementation approaches (public, private, hybrid and multi-cloud) enables system adaptation to specific organizational and technological requirements.

Comparative analysis shows that the choice of optimal cloud architecture depends on the balance between performance, scalability, costs, security and complexity. Although modern architectures offer significant advantages over traditional systems, challenges such as vendor lock-in effect, security risks and resource management remain relevant and require careful strategic planning.

Contemporary trends, including edge computing, the application of artificial intelligence in the management of cloud environments and the concept of the green cloud, point to a further transformation towards intelligent, distributed and sustainable systems. Future research directions, such as autonomous cloud systems, AI orchestration and quantum computing in the cloud, open up new possibilities, but also new questions related to control, security and ethical implications. In conclusion, cloud computing continues to develop as a dynamic and complex field, whose further progress will depend on the ability to integrate innovation with reliable and sustainable architectural solutions.

REFERENCES

- [1] S. Satpathy, B. Sahoo, and A. K. Turuk, "Sensing and actuation as a service delivery model in cloud edge centric internet of things," *Future Generation Computer Systems*, vol. 86, pp. 281–296, 2018, doi: 10.1016/j.future.2018.04.015.
- [2] N. Serrano, G. Gallardo, and J. Hernantes, "Infrastructure as a service and cloud technologies," *IEEE Software*, vol. 32, no. 2, pp. 30–36, 2015, doi: 10.1109/MS.2015.43.
- [3] D. Beimborn, T. Miletzki, and S. Wenzel, "Platform as a service (PaaS)," *Wirtschaftsinformatik*, vol. 53, no. 6, pp. 371–375, 2011, doi: 10.1007/s11576-011-0294-y.
- [4] S. Saeed, Z. M. Zainal, F. A. Ghaleb, and B. A. S. Al-Rimy, "Enhancing public cloud resilience: An analytical review of detection and mitigation strategies against economic denial of sustainability attacks," *Discover Internet of Things*, vol. 5, no. 1, p. 79, 2025, doi: 10.1007/s43926-025-00183-9.

- [5] M. Gupta, D. Gupta, and P. Rai, "Exploring the impact of software as a service (SaaS) on human life," *EAI Endorsed Transactions on Internet of Things*, 2024, doi: 10.4108/eetiot.4821.
- [6] A. Nadar and J. Härri, "An AI-as-a-service platform for an artificial intelligence of things (AIoT)," in *Proc. IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, Jun. 2024, pp. 1–7, doi: 10.1109/VTC2024-Spring62846.2024.10683357.
- [7] G. S. Ravi, K. N. Smith, P. Gokhale, and F. T. Chong, "Quantum computing in the cloud: Analyzing job and machine characteristics," in *Proc. IEEE Int. Symp. Workload Characterization (IISWC)*, Nov. 2021, pp. 39–50, doi: 10.1109/IISWC53511.2021.00015.