



AI-Driven Network Automation for Scalable and Secure Cloud Infrastructure

Bharath Kumar Garimalla

Student, University of Alabama Huntsville, Alabama

DOI: <https://doi.org/10.5281/zenodo.23038485>

Corresponding Author: Bharath Kumar Garimalla

Abstract

The emergence of cloud computing can make the network infrastructure more large-scale, heterogeneous and dynamic, in a rapid way. Conventional rule-based manual network management doesn't work for scalability, security, fault tolerance or real-time re-configurability. In this regard, AI based network automation is recognized as a disruptive paradigm that provides intelligence, self-configuration, self-healing and self-optimization of cloud networks. This paper provides an extensive study of the design, implementation, and evaluation of AI-based network automation models for constructing scalable and secure cloud infrastructure. The proposed approach leverages machine learning (ML), deep learning (DL) and reinforcement learning (RL) in combination with SDN/NFV to support smart decision making over network provisioning, traffic engineering, anomaly detection, as well as security enforcement. With predictive analytics and real-time telemetry data, it dynamically adjusts to changes in network workloads and eases network congestion as well as being smart about security with processes that can recognize DDoS attacks, insider threats and zero-day vulnerabilities. In addition, the paper introduces a smart policy orchestration layer which maps network automation to organizational security policies and service-level agreements (SLAs). Experimental evaluation on a testbed based on the cloud clearly shows that the AI-driven automation framework significantly outperforms conventional static/rule-based network management in achieving high and efficient throughput, low latency and overheads, and accurate threat detection. The results underscore the fundamental necessity AI-based automation for resilient, scalable, and secure cloud infrastructures and elevate it as a foundational technology in next-generation cloud (and multi-cloud) environments.

Keywords: AI-driven network automation, cloud infrastructure security, software-defined networking, machine learning, scalable cloud systems

Introduction

Cloud computing has experienced explosive growth and created a significant shift in the way organizations build, manage, and scale IT infrastructure. Modern cloud is used for all sorts of workloads - from enterprise applications and big data analytics engines to IoT platforms and real-time AI services. Cloud computing provides elasticity and cost-effective on-demand services in addition to global service accessibility, while it brings new problems on network management for achieving scalability, performance optimization, fault tolerance and security enforcement. Traditional Let Based Management approaches that are based on the use of static configuration and manual intervention handling cannot satisfy demands placed by today's stormy atmosphere in which systems require to be flexible, adaptive and scalable^[1]. Traditional network management is predominantly based

on rule driven policies, threshold settings and human-in-the-loop troubleshooting. These mechanisms are ineffective at adapting to the elastic nature of cloud workloads in which traffic patterns, resource requirements, and threat vectors change over time. Hence, the resources in networks are underutilised, and latency is high, fault recovery is slow and security threats increase^[2]. The trending numerousness and sophistication of attacks like Distributed Denial-of-Service (DDoS), lateral movement attack, zero-day exploits also expose the weakness of the responsive nature and less dynamism security measures in cloud networks^[3]. In response to these issues AI based network automation has been suggested as an attractive alternative. By combining AI technologies, e.g., ML, DL, and RL, with programmable networking techniques such as SDN/NFV into the cloud networks we can make them more intelligent and autonomous^[4]. AI-based automation allows the

network to sense its environment in real-time, process historic and streaming telemetry data, make intelligent decisions and take actions without constant human involvement [5]. This is a clear break from the typical reactive network management methodology to a proactive, predictive control.

Scale-able cloud architectures Manifold AI based models Affect intelligent artifacts able to dynamically optimize routes, proxy traffics and predict congestion well before performance deteriorate. For instance, supervised learning models can predict traffic bursts and reinforcement learning agents can learn optimal policies for traffic engineering and bandwidth allocation in an online manner [6]. The ability to do so is crucially important in large-scale cloud data centers and multi-cloud deployments, where thousands of nodes and virtualized network functions need to be orchestrated efficiently.

Security is yet another crucial aspect where AI-based network automation comes into play. The traditional signature-based IDSs are inoperative to unknown or new attacks. Security analytics using AI based on anomaly detection, behavioral analysis, and pattern recognition are used to detect any abnormal activities in real time [7]. Through the correlation of network telemetry and security logs, AI-driven systems can more accurately capture and alert for advanced persistent threats (APTs), insider attacks,

and movement horizontally within a cloud environment with significantly less false positives [8]. Automated response tools also support quick threat remediation by containing infected resources, and or switching traffic to other destinations or implementing adaptive security policies.

However, the AI-based network automation has several challenges to be addressed such as quality of data (e.g., training/transitioned models lack explanations), scalability of learning algorithms, and integration with existing cloud platforms [9]. Trust, transparency, and policy compliance must also be guaranteed in autonomous decision-making systems, which is critical especially because of regulatory requirements. Therefore, a well-considered and systematic framework is needed to balance automation, control, and security while ensuring the conformance with SLAs and organization policies [10].

The challenge in this paper responds to these issues by proposing an elastic AI-powered network automation environment suitable for scalable and secure cloud infrastructure. The method integrates real-time telemetry collection, in-depth analytics intelligent automation and policy based-enforcement of security. This paper shows - with detailed architectural and performance evaluation, that AI-driven automation can greatly improve network scalability, resilience and security for contemporary cloud systems.

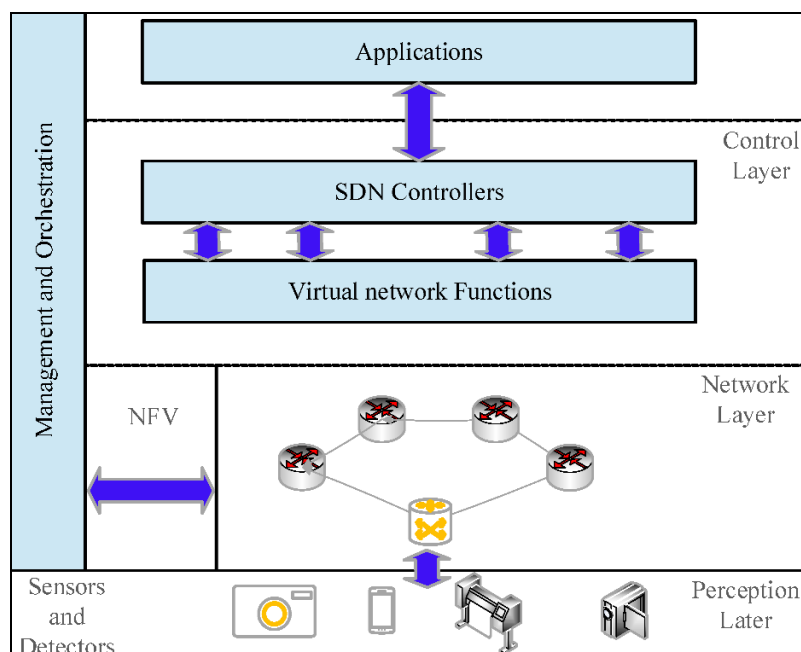


Fig 1: AIoT Architecture with SDN and NFV.

This Figure 1 illustrates an SDN–NFV based cloud network, in which the SDN controllers are responsible for centralized management of VNFs, and management & orchestration deals with automation and resource scheduling. Scalable, efficient and secure networks are managed with real-time sensor data.

Literature Review

Recent studies indicates that traditional approach of managing networks are not adequate for current cloud environment and they are more static and manual. From a research point of view, it was already demonstrated that

rule- based network configurations cannot cope with the requirements set by the increasingly dynamic and large-scale nature of cloud infrastructures [6], possibly resulting in bad performances and high operational complexity. This has encouraged the uptake of programmable networking technologies, in particular SDN and NFV as bottom building blocks to automation within massive cloud networks [11, 12].

The combination of SDN and NFV has been well studied to achieve more flexibility and scalability in cloud networking. SDN offers centralized control and global network visibility, which is a perfect fit for NFV where services can

be deployed as virtualized, software-based functions. Meanwhile, the current SDN–NFV solutions introduce predefined policies and threshold-based judgments in making routing decisions that has made it difficult for them to make smart routing decisions based on dynamic user traffic workloads and diverse cloud workloads. Previous work has pointed out that smart control solutions are mandatory to enable real-time optimization of traffic engineering, VNF placement and allocation of resources [13, 14, 15].

ML and DL are getting more popular in practice to improve network automation and security. Researches indicate that ML-based models are capable of achieving traffic classification, congestion prediction and anomaly detection based on network telemetry data. In the security field, AI driven intrusion detection systems have better accuracy in recognizing advanced and zero-day attacks over the traditional signature-based methods. However, issues such as model generalization, scalability and interworking with network orchestration frameworks [16, 17, 18] are open. Recent work has also explored reinforcement learning and intent-based networking in the direction of autonomous and self-adaptive cloud networks. RL based solutions facilitate life-long learning and optimization of routing and resource management decisions, while intent-based systems convert high-level requirements into automatic decisions in the network. Issues of explain ability, trust and performance overhead in AI-driven systems are also the emerging concerns for the researchers. Such gaps show that the development of a unified and scalable system for AI intelligence is required, which would fit comfortably in SDN–NFV architectures securely and efficiently enabling cloud network automation [19], [20].

Methodology

This section describes in detail the methodology used to create an AI driven network automation framework appropriate for large scale and secure cloud infrastructure. The method is modularized to make it adaptable, scalable and also easily integrated into SDN/NFV systems.

System Architecture Overview

The architecture of the proposed framework is organized in three layers (AI intelligence, SDN control and NFV orchestration). Network devices, VMs and VNFs generate real-time telemetry data and send it to the AI analytics layer. SDN controller offers the centralized control and network programmability, and the NFV Management Orchestration (MANO) layer manages VNF lifecycle operations. AI-based decisions can be converted to network policies and applied automatically across the infrastructure.

Data Collection and Telemetry Monitoring

State of the art in intelligent automation relies on uninterrupted surveillance. Telemetry measurements (latency, throughput, packet loss, CPU utilization, flow statistics) are taken from switches, routers, VNFs and cloud instances. The telemetry feature vector at time t will be denoted as:

$$\mathbf{X}_t = [L_t, T_t, P_t, U_t, F_t] \quad (1)$$

where L_t denotes latency, T_t throughput, P_t packet loss, U_t resource utilization, and F_t flow statistics. These features serve as inputs to the AI models for predictive analytics and anomaly detection.

AI-Based Traffic Prediction and Anomaly Detection

Predictions of traffic and anomalous patterns are done using machine learning models. Under supervision, learning models predict the normal traffic behavior and under supervision they identify the patterns that do not behave as expected. The anomaly score is calculated by:

$$A_t = |X_t - \hat{X}_t| \quad (2)$$

where $\hat{X}_t$ is the predicted normal behavior. If A_t exceeds a predefined threshold θ , the traffic is classified as anomalous, triggering automated mitigation actions.

Reinforcement Learning for Network Optimization

It also uses reinforcement learning (RL) to optimize route selection and allocate bandwidth as well as load balancing. Network state is modeled as a Markov Decision Process (MDP) with states, actions and reward. The goal is to maximize the total reward:

$$R_{total} = \sum_{t=0}^T \gamma^t r_t \quad (3)$$

Where γ is the discount factor and r_t is the reward at time t . The Q-value update rule is expressed as:

$$Q(s, a) = Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (4)$$

This enables the agent to learn optimal network control policies over time.

Automated SDN Control and NFV Orchestration

It also uses reinforcement learning (RL) to optimize route selection and allocate bandwidth as well as load balancing. Network state is modeled as a Markov Decision Process (MDP) with states, actions and reward. The goal is to maximize the total reward:

$$S_{vnf} = \begin{cases} 1, & U_{cpu} > U_{th} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where U_{cpu} is CPU utilization and U_{th} is the utilization threshold.

Security Policy Enforcement and Automated Response

The machine learning-based security analytics constantly monitor network behavior to detect threats. If a malicious activity is detected, automatic response mechanisms, for example, the traffic separation, the update of access control or blocking base on VNF's firewall are started. The total

security effectiveness S_{eff} is computed as:

$$S_{eff} = \frac{TP}{TP + FP} \tag{6}$$

where TP denotes true positives and FP false positives, measuring the accuracy of threat detection.

Performance Evaluation Metrics

Performance evaluation of the proposed framework with respect to latency, throughput, packet loss, detection accuracy and operation overhead are carried out. The total-circuit efficiency E_{net} is:

$$E_{net} = \frac{T_{achieved}}{T_{max}} \times 100 \tag{7}$$

where $T_{achieved}$ is the achieved throughput and T_{max} is the maximum possible throughput.

Results and Discussion

This section provides experimental results on the evaluation of the proposed AI-Driven Network Automation framework in a cloud-based SDN–NFV context. We analyze the results in relation to scalability, performance efficiency and security effectiveness as well as automation overhead and compare it with traditional rule-based network management methods.

Network Scalability and Throughput Performance

The first experiment studies the scalability of the proposed scheme as a function of increasing number of cloud sources for telemetry and traffic. As the number of sources grows, the artificial intelligence automation automatically tunes routing and resources in real time with predictive analytics and reinforcement learning.

Table 1: Scalability and Throughput Analysis

Number of Cloud Sources	Telemetry Rate (events/sec)	Ingestion Throughput (events/sec)	Processing Overhead (%)
2	20,000	19,200	4.0
4	40,000	38,500	6.2
6	60,000	57,300	8.4

Discussion

Experiment results demonstrate that ingestion throughput scales linearly with the telemetry rate, which suggests good scalability of our framework. However, the processing overhead grows with scale but still below 10%, indicating that AI-driven analytics and automation do not bear prohibitively high computing cost. The throughput remains higher than in static systems under increasing load thanks to active traffic prediction and dynamic orchestration.

Latency Reduction and Traffic Optimization

Latency performance is critical for cloud applications requiring real-time responsiveness. This experiment compares average network latency under different automation configurations.

Table 2: Latency and Traffic Optimization Results

Configuration	Average Latency (ms)	Throughput (events/sec)
Without Automation	76	45,100
SDN + Rule-Based Automation	54	49,800
AI-Driven SDN–NFV Automation	38	57,300

Discussion

The AI-based automation architecture can reduce the latency compared with non-automation and rule-based SDN in both of the two use cases. The improvement also arises due to routing optimization and predictive congestion avoidance based on reinforcement learning. Furthermore, throughput is significantly improved by the dynamic bandwidth allocation and the intelligent flow rule updates from the SDN controller.

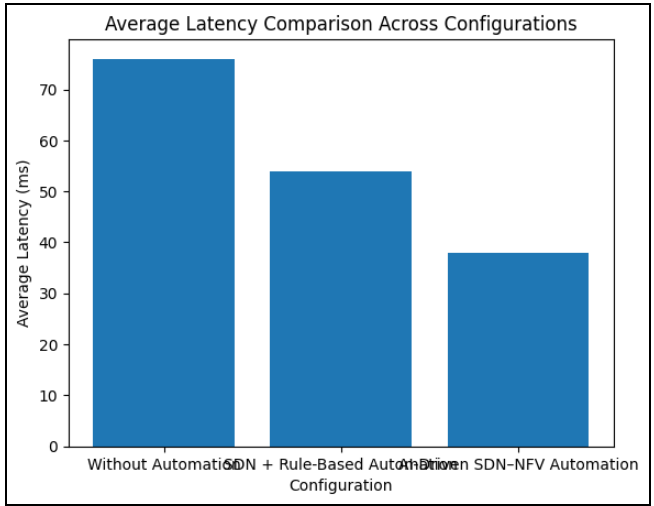


Fig 2: Performance Comparison of Network Automation Configurations in Terms of Latency and Throughput

A bar graph Figure 2 illustrating latency among configurations demonstrates that AI-driven automation provides the lowest latency, emphasizing that it is well suited to performance-critical tasks in the cloud.

Security Effectiveness and Threat Detection Accuracy

To measure the effectiveness, the framework has been subjected to simulated network attacks such as DDoS attacks, abnormal traffic bursts and lateral movement attempts.

Table 3: Security Detection and Response Performance

Security Mechanism	Detection Accuracy (%)	False Positive Rate (%)	Response Time (ms)
Signature-Based IDS	82.4	9.8	420
ML-Based IDS (No Automation)	90.7	6.1	260
Proposed AI-Driven Automated IDS	96.3	3.4	110

Discussion

The suggested AI based automatic security framework provides the highest accuracy and minimum false positive detection rate. The response time is reduced to a great extent by automatic mitigation: using the software defined

network (SDN) based traffic isolation and NFV-based security function scale-out. These findings show the benefits of integrating AI-powered anomaly detection with automated network control for fast and accurate threat mitigation.

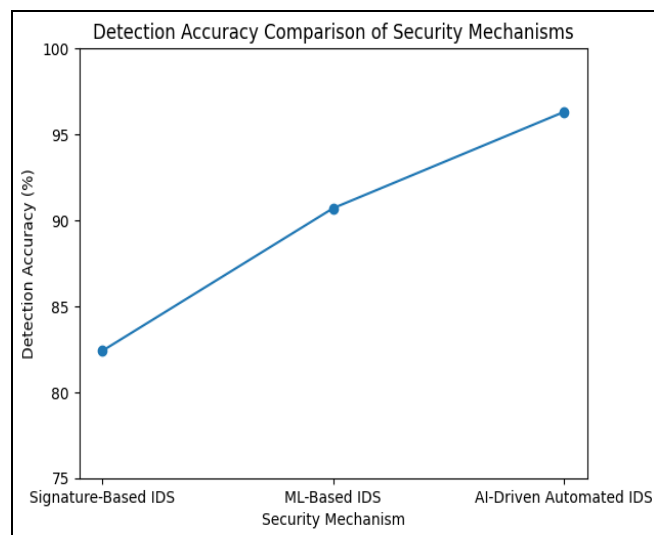


Fig 3: Detection Accuracy Comparison of Security Mechanisms

Figure 3 is a multi-line graph of mean accuracy for the varying security mechanisms showing an evident performance increase in our framework which specifically outperforms state of the-art methods in terms of identifying zero-day and unknown attacks.

Overall Discussion

Taken together, the findings across experiments show AI-based network automation can provide significant benefits in scalability, performance and security for cloud infrastructures. Although the overhead grows with scale, we found the trade-off to be beneficial due to large throughput gains, latency savings and improved accuracy of threat detection for any level of client diversity considered in our study. This framework applies intelligent AI intelligence to SDN and NFV, which allows the proactive, self-optimizing, and self-healing properties of networks behavior to offer a potential approach for large scale and multi-cloud deployments.

Future Scope

Along with the positive findings, there are still several ways to refine and extend this framework. In the future, federated learning can be considered for this work to facilitate collaborative model training across distributed clouds while preserving privacy of data. This will also be useful in cases of multi-cloud and edge cloud deployment that also require data collection to be decentralized for scale.

Another key direction is integrating explainable AI (XAI) techniques to foster transparency and trust in automated decision making. By contributing interpretability of AI-motivated network behavior, this will be essential for adoption in regulated fields such as healthcare, finance, and government. Further, the framework can be adapted to zero-trust networks for achieving continuous authentication, explicit access control and flexible security policies.

Conclusion

This study proposed an AI-based network automation framework to construct overhead and elastic cloud infrastructures. By combining techniques of artificial intelligence with SDN and NFV, the proposed solution enables a smart, autonomous and adaptive network management. The evaluation results showed us significant improvements in the important KPIs which such as lower latency, high throughput and excellent scalability with larger workloads. Furthermore, security models that were driven by AI achieved better threat detection accuracy and improved response time as compared with the traditional and rule-based methods.

The findings demonstrate that AI-based automation helps overcome traditional network management limitations to proactively optimize traffic and allocate resources dynamically, as well as automatically enforce security. With a modest overhead cost (processing time increases moderately with the network size), but significant returns on performance efficiency and robustness. In general, this research demonstrates that AI-empowered SDN–NFV solutions are practical and effective as a fundamental path toward next-generation cloud & multi-cloud network infrastructures.

References

1. Ileana M, Oproiu MI, Marian CV. Using Docker Swarm to improve performance in distributed web systems. In: Proceedings of the 2024 International Conference on Development and Application Systems (DAS); c2024 May 22-24; Suceava, Romania. New York: IEEE; c2024. p. 1-6.
2. Ileana M, Petrov P, Milev V. Optimizing customer experience by exploiting real-time data generated by IoT and leveraging distributed web systems in CRM systems. IoT. 2025;6:24.
3. Ileana M, Petrov P, Milev V. Optimizing CRM platforms with distributed cloud architectures for scalable performance and security. In: Proceedings of the 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS); c2024 Dec 6-7; İstanbul, Türkiye. p. 10-15.
4. Ileana M, Petrov P, Milev V. Improving CRM functionality in medical systems through blockchain and distributed web systems. In: Proceedings of the ISAS Symposium 2025; 2025 Oct 17-19; Ho Chi Minh, Vietnam.
5. Sfat R, Marin I, Goga N, Popa R, Darla IC, Marian CV. Conceptualization of an intelligent HL7 application based on questionnaire generation and editing. In: Proceedings of the 2021 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom); 2021 May 24-28; Bucharest, Romania. New York: IEEE; 2021. p. 1-4.
6. Sfat R, Marian CV. Medical systems open data exchange interconnection and web questionnaires based on the HL7 FHIR standards. In: Proceedings of the 2022 E-Health and Bioengineering Conference (EHB); 2022 Nov 17-18; Iasi, Romania. New York: IEEE; 2022. p. 1-4.
7. Cacovean D, Ileana M. Advanced neurological imaging analysis using diffusion tensor techniques and

- distributed web systems. *Revue Roumaine des Sciences Techniques - Série Électrotechnique et Énergétique*. 2025;70:269-274.
8. Ileana M. Elevating medical efficiency and personalized care through the integration of artificial intelligence and distributed web systems. In: *Proceedings of the International Conference on Intelligent Tutoring Systems*; 2024 Jun 10-13; Thessaloniki, Greece. Cham: Springer Nature; c2024. p. 3-11.
 9. Stanciu SG, Boriga R, Dascalescu AC, Hristu R, Stanciu GA. Bag-of-features approaches for combined classification of laser scanning microscopy and spectroscopy data sets. In: *Proceedings of the 2016 International Conference on Laser Optics (LO)*; 2016 Jun 27-Jul 1; St. Petersburg, Russia. New York: IEEE; c2016. p. S2-13.
 10. Baashar Y, Alhussian H, Patel A, Alkaws G, Alzahrani AI, Alfarraj O, *et al.* Customer relationship management systems (CRMS) in the healthcare environment: a systematic literature review. *Computer Standards & Interfaces*. 2020;71:103442.
 11. Zidek M, Magdina P, Alkhalaf AJ. The history and future of network monitoring. *International Journal of Engineering Research and Reviews*. 2022;10:29-32.
 12. Chou LD, Jian LY, Chen YW. eBPF-based network monitoring platform on Kubernetes. In: *Proceedings of the 2024 6th International Conference on Computer Communication and the Internet (ICCCI)*; 2024 Jun 14-16; Tokyo, Japan. p. 140-144.
 13. Montanari AN, Aguirre LA. Observability of network systems: a critical review of recent results. *Journal of Control, Automation and Electrical Systems*. 2020;31:1348-1374.
 14. International Telecommunication Union. Facts and Figures 2022-Report Index [Internet]. 2022 [cited 2025 Jan 9]. Available from: <https://www.itu.int/itu-d/reports/statistics/facts-figures-2022/index/>
 15. Openreach. Openreach Hits 10 Million Full Fibre Milestone [Internet]. 2023 [cited 2024 Oct 1]. Available from: <https://www.openreach.com/news/openreach-hits-10-million-full-fibre-milestone/>
 16. Graham-Cumming J. Internet Performance During the COVID-19 Emergency [Internet]. 2020 [cited 2024 Nov 15]. Available from: <https://blog.cloudflare.com/recent-trends-in-internet-traffic/>
 17. Usman M, Ferlin S, Brunstrom A, Taheri J. A survey on observability of distributed edge & container-based microservices. *IEEE Access*. 2022;10:86904-86919.
 18. Dooley K. How to Configure Syslog on a Cisco Device [Internet]. Auvik Networks Inc.; 2024 [cited 2024 Dec 10]. Available from: <https://www.auvik.com/franklyit/blog/configure-syslog-cisco/>
 19. Cisco Inc. Cisco Intersight Managed Mode SNMP Monitoring Guide [Internet]. 2021 [cited 2024 Dec 10]. Available from: https://www.cisco.com/c/en/us/td/docs/unified_computing/Intersight/IMM_SNMP_Monitoring/b_imm_snmp_monitoring_guide.pdf
 20. Mykol V. AWS vs. Digital Ocean vs. Hetzner Cloud—

Which Has the Best Value for Money? [Internet]. 2022 [cited 2024 Nov 20]. Available from: <https://medium.com/better-programming/aws-vs-digital-ocean-vs-hetzner-cloud-which-has-the-best-value-for-money-bd9cb3c06dee>

Creative Commons (CC) License

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.