World Journal of Innovation and Modern Technology, Vol. 7, Issue 4, (Aug) 2024
ISSN 2682-5910
DOI: 10.53469/wjimt.2024.07(04).07

Design of the Data Center Network Architecture under the Background of Cloud Computing: A Review

Xiong Ying

Independent Researcher

Abstract: ***This article explores the importance and challenges of data center network architecture design in the context of cloud computing. The traditional data center network architecture has some limitations in meeting the needs of cloud computing, such as network topology bottlenecks and inefficient resource utilization. Moreover, data security and privacy issues, as well as performance and low latency requirements, have become more critical. To address these challenges, new data center network architecture design principles have been proposed, including the application of software-defined networking (SDN), hyperconverged infrastructure, and the integration of security and privacy.***

Keywords: Cloud computing; Data center; Network architecture.

1. INTRODUCTION

In the era of cloud computing, the design of a data center network architecture has become crucial, as it directly affects the performance, availability, and security of cloud services[1-6]. The traditional data center network architecture is no longer suitable for high elasticity, high availability, and large-scale scalability requirements in cloud computing environments. These traditional architectures have limitations in terms of network topology design and hardware device selection, resulting in low resource utilization efficiency[7]. Therefore, on the basis of the background of cloud computing, this article explores a new data center network architecture design.

2. OVERVIEW OF THE DATA CENTER NETWORK ARCHITECTURE

2.1 Traditional Data Center Network Architecture

The traditional data center network architecture is a complex system consisting of multiple levels, typically with a layered structure. In this structure, the core layer is responsible for handling large-scale data transmission, whereas the aggregation layer directs traffic from the core layer to the distributed layer, ultimately reaching the access layer[8-9].

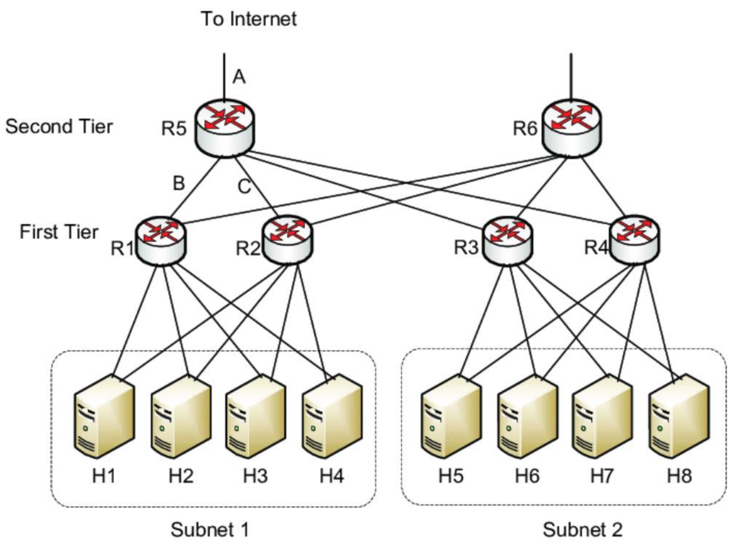

Figure 1: Traditional Data Center Network Architecture

As shown in Figure 1, the hardware equipment of the data center network includes core switches, aggregation switches, access layer switches, firewalls and virtual private networks (VPNs) used for network security. Data centers are typically connected to multiple telecommunications operators, such as China Telecom and China Unicom, to achieve dual-line redundancy. In addition, the data center also has a VPN for connecting various branch offices and a UTM (Unified Threat Management) for network security, including firewall, VPN, intrusion detection and defense (IDP) and other functions[10-12]. There is also a dedicated VPN DMZ area for isolating external websites, email servers, and other public servers to ensure security. Moreover, the data center also includes dual core switches for implementing hot backup to ensure high availability of the network. The overall network architecture can provide a 1000 Mbps LAN internet speed and includes security and performance features such as behavior management, traffic control, vulnerability scanning, and antivirus servers. Access layer switches and network management servers are used to manage terminal devices, internal servers, and network maintenance and monitoring. This traditional DCN architecture provides enterprises with a stable and high-performance DCN foundation[13-16].

2.2 Network Requirements for Cloud Computing

2.2.1 Elasticity

In the cloud computing environment, resilience has become a critical network requirement. Elasticity means that the network must to adapt to constantly changing workloads and demands. In cloud computing, the dynamic creation and destruction of virtual machines, as well as the elastic scaling of applications, require networks to adapt quickly to these changes. This requires the network to have the ability to automatically adjust bandwidth and resources to ensure sufficient performance during high loads and effective resource utilization during low loads. Network devices and topology structures need to support dynamic changes to adapt to the continuous evolution of workloads[17].

2.2.2 High availability

High availability is another important requirement for cloud computing networks. Owing to the critical importance of applications and data in cloud computing for enterprises, network unavailability may lead to productivity loss and data loss. Therefore, cloud computing networks must have high availability to ensure continuity and reliability. This requires the use of redundant equipment, links, and data centers to prevent single-point failures. Moreover, the network requires real-time monitoring and fault switching mechanisms to quickly switch to backup paths in the event of a failure, reducing the interruption time. High availability also requires consideration of network security to address potential network attacks and threats[18].

2.2.3 Large-scale scalability

Cloud computing networks must have large-scale scalability to meet the growing demands of users and applications. Large-scale scalability requires the network to support simultaneous access by many users and large-scale data transmission. This requires the adoption of distributed architecture and load balancing technology to ensure smooth network expansion without performance bottlenecks [2]. Moreover, large-scale scalability also requires the consideration of automation in network management and configuration to reduce maintenance costs and improve network scalability[19].

3. CHALLENGES AND PROBLEMS

3.1 Limitations of traditional architecture

3.1.1 Network topology bottleneck

Traditional data center networks typically adopt a layered network topology, such as a three-layer structure (core Distribution, Access) or two-tier structure (Core, Access)[20]. This structure may perform well in small-scale data centers, but there are bottleneck issues in large-scale data centers. With the dynamic migration of virtual machines and the deployment of high-density servers in cloud computing, the traffic in the network has increased sharply, and traditional topology structures cannot effectively cope with it. The devices at the core layer become bottlenecks, making it difficult to provide sufficient bandwidth and low latency, resulting in performance degradation and network congestion. This limits the scalability and performance of data centers, and the requirements for cloud computing applications are no longer applicable.

3.1.2 Inefficient Resource Utilization

The resource utilization in traditional data center networks is usually inefficient, which is reflected mainly in the waste of hardware equipment and links[21]. Because network devices are typically deployed in fixed configurations, some devices may be in a low load state, and resources may not be fully utilized. In addition, the resources of the link may not be fully utilized, as traffic is usually allocated according to the worst-case scenario, resulting in waste of link resources. This not only increases the construction and operation costs of data centers but also limits the scalability of resources.

3.2 Security and Privacy Issues

3.2.1 Data protection

Data protection is an important issue in cloud computing. Because cloud computing typically involves large-scale data storage and processing, data protection has become crucial. Cloud data centers need to take a series of measures to ensure the confidentiality and integrity of the data[22]. This includes security policies such as data encryption, authentication, access control, and data backup. Moreover, cloud service providers need to establish highly secure data centers and adopt physical and logical security measures to protect data from unauthorized access and threats. The network architecture of data centers must have the ability to encrypt data and ensure transmission security to meet customers' compliance requirements for data privacy.

3.2.2 Isolation

In a multitenant cloud computing environment, different users and tenants share the same data center network infrastructure. Therefore, it is necessary to ensure isolation between different tenants to prevent horizontal penetration and resource competition. Network isolation is the key to ensuring security between different tenants[23]. The network architecture of cloud data centers must support technologies such as virtual network isolation, access control lists, and virtual local area networks (VLANs) to ensure isolation and security between tenants.

3.3 Performance and low-latency requirements

In cloud computing environments, performance and low latency requirements are key challenges in data center network architecture design. Cloud computing applications have extremely high requirements for network performance and low latency, as users expect fast data transmission and efficient running of real-time applications. This means that the data center network must provide sufficient bandwidth and reduce the data transmission latency to ensure a smooth user experience[24]. In addition, the rise of virtualization and cloudification trends has made networks more complex, requiring support for rapid migration and resource allocation of virtual machines, which further increases the demand for performance and low latency. In response to this challenge, the data center network architecture needs to integrate high-speed network devices and intelligent routing and switching technologies to meet the high-performance and low-latency requirements of cloud computing applications, ensuring that data center networks can support large-scale data transmission and real-time application operations, providing users with excellent experience.

4. DESIGN OF THE DATA CENTER NETWORK ARCHITECTURE IN THE CONTEXT OF CLOUD COMPUTING

4.1 Software-Defined Networking (SDN)

4.1.1 Basic principles of SDN

The core idea of software-defined networking (SDN) is to separate the network control plane from the data forwarding plane to achieve more flexible, programmable, and centralized network management[25]. As shown in Figure 2, the controller in the SDN architecture is the brain of the network and is responsible for formulating routing policies, managing network traffic, and responding to various network events. The controller can be centralized or distributed, depending on the requirements of the network. Network devices at the data layer, such as switches and routers, execute network policies on the basis of controller instructions to transmit data packets from the source to the destination. The northbound interface is the interface between the controller and the application layer, allowing applications to communicate with the network to program and control network policies. The southbound interface is the interface between the controller and network devices, through which the controller can issue configuration and policy information to the network devices to control network traffic.

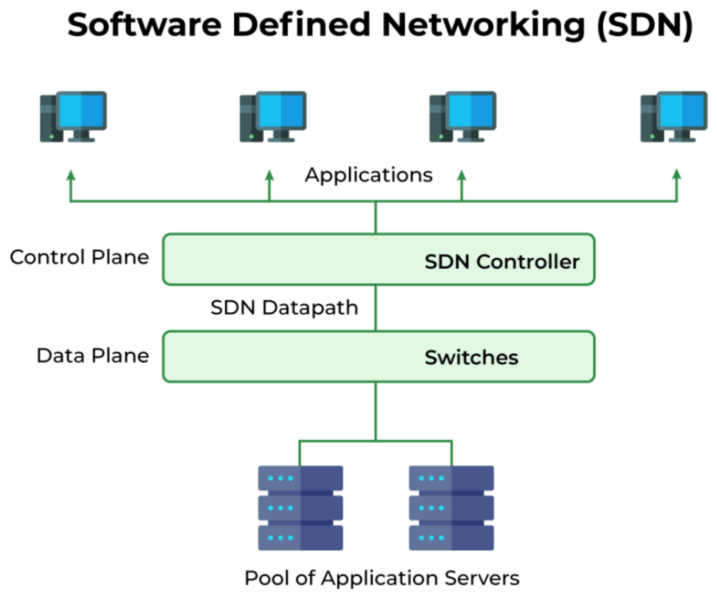

Figure 2: SDN Architecture Framework

4.1.2 Advantages of SDN

The application of SDN has brought significant advantages in cloud computing data center network architecture. First, SDN provides centralized network control, making network management more flexible. Network administrators can adjust network policies in real time according to their needs, thereby achieving optimal allocation of network resources. Second, the programmability of the SDN architecture makes the network highly customizable, allowing for a precise definition of network behavior on the basis of specific application requirements and improving network adaptability. In addition, the centralized control of SDN also reduces management complexity, reduces the difficulty of network maintenance, and thus lowers operating costs [26]. Most importantly, the advantage of SDN lies in its open nature, which allows third-party development of applications and introduces more innovation to data center networks, accelerating the evolution of network technology.

4.2 Superconverted infrastructure

4.2.1 Computing, Storage, and Networking Integration

The hyperconverged infrastructure (HCI) is a key innovation in cloud computing data center network architecture, which highly integrates computing, storage, and networking on a unified hardware platform[27]. In this architecture, computing resources, storage resources, and network resources are intertwined to achieve collaborative work. Computing nodes can directly access storage resources, whereas network resources provide highly flexible connections and communication methods. This integrated design makes the operation of data centers more efficient, reduces latency, and improves performance. Moreover, hyperconverted infrastructure also reduces the space requirements of data centers, lowers energy consumption, and provides support for sustainable development.

4.2.2 Collaboration of Hardware and Software

In the context of cloud computing, the architecture design of data center networks needs to closely integrate hardware and software to achieve collaborative work to meet the growing demands for computing, storage, and networking. The core idea of hyperconverted infrastructure (HCI) is the deep collaboration of hardware and software, which is the key to building efficient and flexible data center networks.

The synergy between hardware and software is fully demonstrated in the hyperconverted infrastructure. First, in terms of hardware, HCI adopts highly integrated hardware nodes, which typically include computing, storage, and network components. The interconnection between these hardware resources is achieved through software-defined methods, making resource allocation and scheduling more flexible and efficient. The communication between hardware nodes no longer relies on traditional hardware routers but is managed and scheduled by software controllers, thereby improving the scalability and performance of the network.

Second, in terms of software, hyperconverted infrastructure adopts virtualization technology to abstract hardware resources into virtual resources, achieving better resource utilization. This means that data center administrators can easily manage and allocate resources through software interfaces without the need to directly operate physical hardware. In addition, software-defined networking (SDN) technology is widely used, which allows administrators to dynamically adjust network topology and policies as needed to adapt to constantly changing workloads [5].

The collaboration between hardware and software is also reflected in the automation management of data centers. Through software-defined methods, data centers can achieve automated resource scheduling and load balancing, ensuring efficient completion of various tasks. Administrators can use monitoring and analysis tools to understand the performance of networks and applications in real time, thereby better responding to issues and needs.

In short, the collaboration of hardware and software is the core of hyperconverted infrastructure design, providing high flexibility, efficiency, and scalability for data center networks. Through deep integration and automated management, hyperconverted infrastructure is expected to meet the constantly changing demands of cloud computing, providing important references for future data center network architecture design.

4.3 Security and privacy integration

4.3.1 Encrypted Communication

In the design of a cloud computing data center network architecture, security and privacy are crucial considerations. Among them, encrypted communication is key for ensuring the protection of data during transmission. Encryption communication converts data into ciphertext to ensure that only authorized users can decrypt and read the data.
The specific design of encrypted communication usually adopts a public key infrastructure (PKI), which includes a pair of public keys and private keys for encrypting and decrypting data. Each node and terminal device in the data center network needs to generate and manage these key pairs. At the beginning stage of communication, both parties exchange their public keys to establish a secure communication channel.
The data are encrypted during transmission to ensure that even if unauthorized access occurs during the data transmission process, the content of the data cannot be interpreted. Once the data reach the recipient, the recipient will decrypt the data via their private key to access the original data. This encrypted communication process is particularly important in the multitenant environments of cloud computing, as data from different users may be stored on the same hardware device. Encrypting communication can prevent unauthorized visitors from accessing the data of other tenants.

In addition, encrypted communication can prevent data from being eavesdropped during transmission, ensuring the confidentiality and integrity of the data. By using strong encryption algorithms, cloud computing data centers can provide highly secure communication channels, ensuring optimal protection for user data. In short, encrypted communication is a key component of cloud computing data center network architecture design. By adopting a public key infrastructure and strong encryption algorithms, data transmission protection, privacy, and security can be achieved. This helps ensure that user data are properly protected and secure in cloud computing environments.

4.3.2 Authentication and Access Control

In the design of the cloud computing data center network architecture, authentication and access control are key aspects for ensuring data security and privacy. The design of this section aims to verify user identity and restrict access to data and network resources.

Identity verification typically involves users providing credentials such as usernames and passwords to verify their identities. However, in cloud computing environments, more advanced authentication methods can also be adopted, such as multifactor authentication (MFA) or biometric technologies such as fingerprint recognition or facial recognition. These advanced authentication methods provide greater security, ensuring that only authorized users can access cloud computing resources.

Access control is used to restrict users' access to specific resources. This can be achieved through role-based access control (RBAC), which assigns specific roles and permissions to each user to ensure that they can access only the resources necessary for their work. In addition, fine-grained access control can also be implemented, allowing administrators to define user access permissions to resources more specifically, thereby improving security.

In the design of the data center network architecture, these two aspects of detail are crucial. They are typically used in conjunction with encrypted communication to ensure that only authenticated users can decrypt and access data. In addition, audit trails for authentication and access control can help detect potential security threats, thereby enhancing overall security.

In summary, authentication and access control play important roles in the network architecture design of cloud computing data centers. By adopting advanced authentication methods and fine-grained access control, data security and privacy can be properly managed and maintained. This helps to protect user data, prevent unauthorized access, and improve overall security.

5. CONCLUSION

This article explores and analyzes the network architecture design of data centers in the context of cloud computing, emphasizing the changes in network requirements caused by cloud computing and the limitations of traditional architectures. A new data center network architecture design has been proposed in the face of challenges and problems, mainly including the application of software-defined networking (SDN), hyperconverged infrastructure, and security and privacy integration. In the future, the design of the data center network architecture will continue to adapt to the constantly changing demands of cloud computing, resulting in higher performance, better security, and greater flexibility. Multitask learning for data-efficient spatiotemporal modeling of tool surface progression in ultrasonic metal welding Predicting ICU admissions for hospitalized COVID-19 patients with a factor graph-based model using emotion recognition technology to enhance the user experience in real time.

REFERENCES

[1] Qi, H., Shiraz, M., Liu, J. Y., Gani, A., Abdul Rahman, Z., & Altameem, T. A. (2014). Data center network architecture in cloud computing: review, taxonomy, and open research issues. Journal of Zhejiang University SCIENCE C, 15, 776-793.

[2] Yao, J., & Yuan, B. (2024). Research on the Application and Optimization Strategies of Deep Learning in Large Language Models. Journal of Theory and Practice of Engineering Science, 4(05), 88-94.

[3] Ma, Y., Shen, Z., & Shen, J. (2024). Cloud Computing and Hyperscale Data Centers: A Comparative Study of Usage Patterns. Journal of Theory and Practice of Engineering Science, 4(06), 11-19.

[4] Xu, Y., Lin, Y.-S., Zhou, X., & Shan, X. (2024). Utilizing emotion recognition technology to enhance user experience in real-time. Computing and Artificial Intelligence, 2(1), 1388. https://doi.org/10.59400/cai.v2i1.1388

[5] Ji, H., Xu, X., Su, G., Wang, J., & Wang, Y. (2024). Utilizing Machine Learning for Precise Audience Targeting in Data Science and Targeted Advertising. Academic Journal of Science and Technology, 9(2), 215-220.

[6] Wu, X., Wu, Y., Li, X., Ye, Z., Gu, X., Wu, Z., & Yang, Y. (2024). Application of adaptive machine learning systems in heterogeneous data environments. Global Academic Frontiers, 2(3), 37-50.

[7] Wang, Z. (2024, August). CausalBench: A Comprehensive Benchmark for Evaluating Causal Reasoning Capabilities of Large Language Models. In Proceedings of the 10th SIGHAN Workshop on Chinese Language Processing (SIGHAN-10) (pp. 143-151).

[8] Chen, M., Jin, H., Wen, Y., & Leung, V. C. (2013). Enabling technologies for future data center networking: a primer. Ieee Network, 27(4), 8-15.

[9] Lu, Q., Guo, X., Yang, H., Wu, Z., & Mao, C. (2024). Research on Adaptive Algorithm Recommendation System Based on Parallel Data Mining Platform. Advances in Computer, Signals and Systems, 8(5), 23-33.

[10] Lee, G. (2014). Cloud networking: Understanding cloud-based data center networks. Morgan Kaufmann.

[11] Ren, Z. (2024). VGCN: An Enhanced Graph Convolutional Network Model for Text Classification. Journal of Industrial Engineering and Applied Science, 2(4), 110-115.

[12] Adami, D., Martini, B., Sgambelluri, A., Donatini, L., Gharbaoui, M., Castoldi, P., & Giordano, S. (2017). An SDN orchestrator for cloud data center: System design and experimental evaluation. Transactions on Emerging Telecommunications Technologies, 28(11), e3172.

[13] Qi, H., Shiraz, M., Gani, A., Whaiduzzaman, M., & Khan, S. (2014). Sierpinski triangle based data center architecture in cloud computing. The Journal of Supercomputing, 69, 887-907.

[14] Zhu, Z., Wang, Z., Wu, Z., Zhang, Y., & Bo, S. (2024). Adversarial for Sequential Recommendation Walking in the Multi-Latent Space. Applied Science and Biotechnology Journal for Advanced Research, 3(4), 1-9.

[15] Cao, Y., Cao, P., Chen, H., Kochendorfer, K. M., Trotter, A. B., Galanter, W. L., ... & Iyer, R. K. (2022). Predicting ICU admissions for hospitalized COVID-19 patients with a factor graph-based model. In Multimodal AI in healthcare: A paradigm shift in health intelligence (pp. 245-256). Cham: Springer International Publishing.

[16] Jadeja, Y., & Modi, K. (2012, March). Cloud computing-concepts, architecture and challenges. In 2012 international conference on computing, electronics and electrical technologies (ICCEET) (pp. 877-880). IEEE.

[17] Xu, J., Jiang, Y., Yuan, B., Li, S., & Song, T. (2023, November). Automated Scoring of Clinical Patient Notes using Advanced NLP and Pseudo Labeling. In 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 384-388). IEEE.

[18] Yuan, B., & Song, T. (2023, November). Structural Resilience and Connectivity of the IPv6 Internet: An AS-level Topology Examination. In Proceedings of the 4th International Conference on Artificial Intelligence and Computer Engineering (pp. 853-856).

[19] Yao, J., & Yuan, B. (2024). Optimization Strategies for Deep Learning Models in Natural Language Processing. Journal of Theory and Practice of Engineering Science, 4(05), 80-87.

[20] Lin, Z., Wang, Z., Zhu, Y., Li, Z., & Qin, H. (2024). Text Sentiment Detection and Classification Based on Integrated Learning Algorithm. Applied Science and Engineering Journal for Advanced Research, 3(3), 27-33.

[21] Lyu, H., Wang, Z., & Babakhani, A. (2020). A UHF/UWB hybrid RFID tag with a 51-m energy-harvesting sensitivity for remote vital-sign monitoring. IEEE transactions on microwave theory and techniques, 68(11), 4886-4895.

[22] Yuan, B. (2024). Design of an Intelligent Dialogue System Based on Natural Language Processing. Journal of Theory and Practice of Engineering Science, 4(01), 72-78.

[23] Xu, X., Yuan, B., Song, T., & Li, S. (2023, November). Curriculum recommendations using transformer base model with infonce loss and language switching method. In 2023 5th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 389-393). IEEE.

[24] Pengwei, Z., & Qiang, L. (2021). Design of Data Center Network Architecture for Cloud Computing in Forestry Information. Forest Chemicals Review, 52-62.

[25] Ye, K., Huang, D., Jiang, X., Chen, H., & Wu, S. (2010, December). Virtual machine based energy-efficient data center architecture for cloud computing: a performance perspective. In 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing (pp. 171-178). IEEE.

[26] Cai, L., Xing, C., & Deng, Y. (2021). Research on digital urban architecture design based on cloud computing data center. Environmental Technology & Innovation, 22, 101543.

[27] Yuan, B., Song, T., & Yao, J. (2024, January). Identification of important nodes in the information propagation network based on the artificial intelligence method. In 2024 4th International Conference on Consumer Electronics and Computer Engineering (ICCECE) (pp. 11-14). IEEE.