Compound User Scenarios on a Hybrid Cloud

Hui-Shan Chen
National Center for High-performance Computing
National Applied Research Laboratories
Taichung, Taiwan
email: chwhs@narlabs.org.tw

Yi-Lun Pan
National Center for High-performance Computing
National Applied Research Laboratories
Hsinchu, Taiwan
email: serenapan@narlabs.org.tw

***Abstract*—The Information Technology (IT) industry is increasingly adopting hybrid cloud environments, which can leverage the advantages of public and private clouds. The Hybrid Cloud Platform of National Center for High-performance Computing (NCHC) provides flexibility and scalability strategies to meet various cross-cloud scenarios and artificial intelligence applications. Furthermore, based on this cloud architecture, High-Performance Computing (HPC) services are integrated to provide colossal computing power. Therefore, the cloud platform supports various value-added cloud services and applications suitable for various user scenarios.**

Keywords-hybrid cloud; cross-cloud; artificial intelligence; high-performance computing.

I. INTRODUCTION

Based on a highly available, highly reliable, and scalable design, the resource management of NCHC Hybrid Cloud Platform [1] connects with different cloud providers to realize hybrid cloud management of private and public clouds. The cloud platform incorporates multiple cloud technologies, including hyper-converged infrastructure, multi-cloud management tools [2][3], software-defined data centers and private network technology such as Chief Cloud eXchange (CCX) [4]. In Section 2, we introduce the system architecture. Then Section 3, the implementations for user scenarios are described. Finally, we conclude this poster and plan future developments in Section 4.

II. SYSTEM ARCHITECTURE

The NCHC Hybrid Cloud Platform supports diverse cloud services and offers flexible cloud solutions, whether for HPC applications, AI education, disaster recovery, or sensitive application requirements. Figure 1 presents the system architecture of the cloud platform. It leverages cross-cloud resource management to invoke the computing and storage resources of the public cloud for expansion. In order to cope with sensitive needs, the protection of sensitive data and network access control are very important, so the cloud platform also marks out the privacy zone.

The cloud platform is designed to integrate HPC services by designing a Resource Broker [5], an integrated scheduling and management tool for physical and virtualized computing resources. The Resource Broker can leverage various service resources through an API Gateway, including HPC, hybrid cloud, and storage devices.

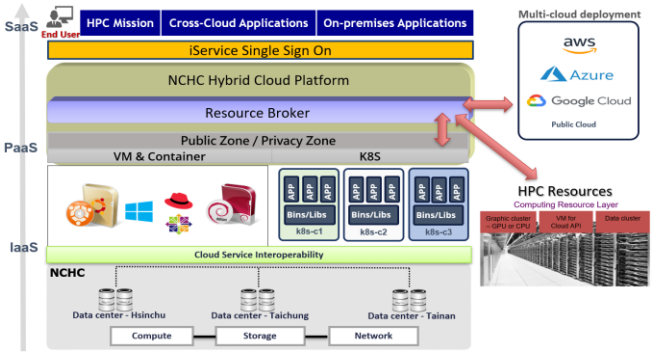

Figure 1. System Architecture

Moreover, Site-to-Site VPNs, private connections, or other network technologies will be established to connect the internal networks of virtual machine service and HPC services. Then, users just need to create a virtual machine as an interactive node to access HPC services. This will enable the cloud platform to perform computational simulations through seamless interaction with HPC resources and public clouds.

III. IMPLEMENTATION FOR USER SCENARIO

Leveraging the elasticity of the public cloud effectively not only achieves system scalability but also significantly reduces costs. The designed scheduling and management tools - Resource Broker can dynamically allocate appropriate computing resources based on demand.

A. *HPC Missions*

When on-premises cloud resources are fully loaded and reach quota limits, users can quickly submit requirements to HPC environments or public clouds to run HPC workloads [6]. We design and develop the integrated functions and interfaces of the cloud platform to bridge the HPC environment and connect HPC resources and storage devices to meet the needs of computing simulations. API Gateway dispatches and triggers all HPC Jobs for execution on computing nodes. The integration shortens researchers' development time and testing cycle, and achieves high-performance benefits.

B. *Cross-Cloud Services*

This research mainly integrates with AWS, so AWS services are used as examples. The Cross-Cloud Services [7] are connected through Site-to-Site VPN or AWS Direct Connect [8].

- **Cloud HPC Service.** The Cloud HPC Parallel Cluster seamlessly integrates on-premises and public cloud resources in hybrid cloud environments. AWS ParallelCluster [9] can automatically deploy and configure the SLURM for automated resource management and job scheduling. Additionally, it can dynamically expand or shrink the number of cloud hosts based on workload and support the job priority mechanism.
- **Disaster Recovery (DR).** An automated disaster recovery framework has been established, and it integrates AWS Elastic Disaster Recovery [10] to offer cloud-based disaster recovery services. Based on a multi-cloud management tool - Morpheus [11], scheduled jobs make it easier to automate system operation and management through workflow. The designed mechanism simplifies the complexity of implementing disaster recovery.
- **Site-to-Site VPN.** To ensure the high availability of cloud services, a secure and reliable private connection provides direct access to the public cloud and on-premises cloud resources when needed. On the hybrid cloud platform, PFSENSE [12] is adopted to replace CCX. Setting up tunnels separately is required for IPsec VPN to provide private networks, as shown in Figure 2. This enables on-premises cloud platforms to take full advantage of the elasticity and scalability provided by public clouds and also reduce costs.

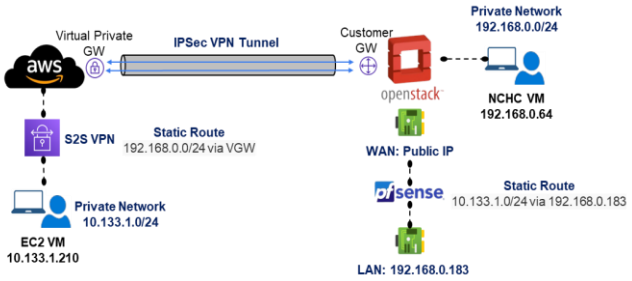

Figure 2. Site-to-Site VPN

C. *On-Premises Applications*

- **AI training and AI Education.** One such application is AI model training for making predictions on new data. Figure 3 shows running AI applications [13] [14] on GPU virtual machines will speed up the model training process and shorten model development time. In addition, it can also be used in the education field to provide teachers with a consistent or customized teaching environment. In this way, it can reduce the cost of IT infrastructure construction and management, and promote the popularization of cloud artificial intelligence education.

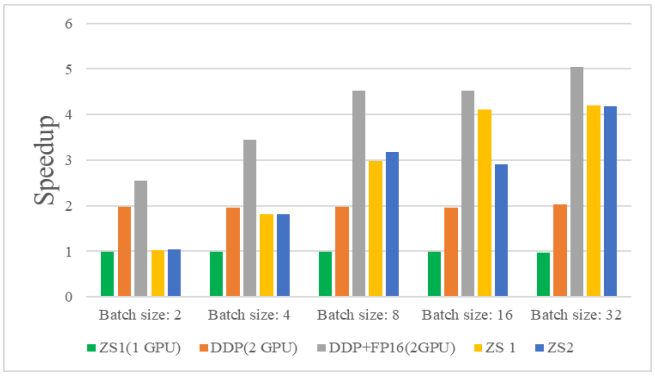

Figure 3. AI Applications

- **Privacy Zone.** This cloud platform provides a private portal, and users need to access and manage resources through a Virtual Desktop Infrastructure (VDI), such as Kasm workspace [15]. Privacy Zone is designed with a fully software-defined architecture to meet privacy requirements. The platform provides adequate network isolation and protection of highly sensitive data. This can ensure that non-approved data remains within the privacy zone, as shown in Figure 4.

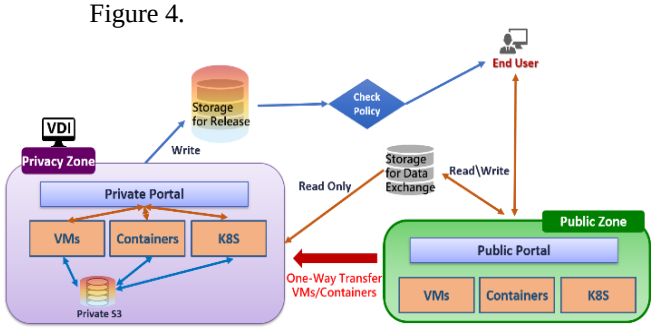

Figure 4. Privacy Zone

Whether it is for the medical field, government agencies, or research centers, information security risks will be significantly improved and avoided in the future.

IV. CONCLUSIONS

When the hybrid cloud platform needs to enlist HPC computing resources, the storage of the privacy zone needs to be synchronized to the computing nodes of the HPC environment. Therefore, our challenges include: (1) Ensuring that sensitive data does not leave the privacy zone. (2) Secure access control when requisitioning HPC resources. (3) Encrypted data transmission. These considerations are crucial for the future development of our platform. It can effectively enhance the security of cross-platform environments and ensure data transmission.

REFERENCES

[1] NCHC Hyper Kylin Hybrid Cloud Platform, URL:https://portal.hci-hybrid.nchc.org.tw/portal, August, 2024.

[2] H. A. Imran et al., "Multi-Cloud: A Comprehensive Review", 2020 IEEE 23rd International Multitopic Conference, Bahawalpur, Pakistan, pp. 1-5, doi: 10.1109/INMIC50486.2020.9318176, 2020.

[3] J. Hong, T. Dreibholz, J. A. Schenkel, and J. A. Hu, "An overview of multi-cloud computing." Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 33rd International Conference on Advanced Information Networking and Applications 33. Springer International Publishing, pp. 1055-1068, 2019.

[4] Chief Telecom Inc, "Chief Cloud eXchange Center[Computer software]", URL:https://en.chief.com.tw/exchange-center/, August, 2024.

[5] P. Subramanian et al., "Progressive Reservation of Cloud Services Using Multi-Cloud Broker System", Engineering Proceedings, 59(1):29, https://doi.org/10.3390/engproc2023059029, 2023.

[6] Q. He, S. Zhou, B. Kobler, D. Duffy, and T. McGlynn, "Case study for running HPC applications in public clouds", In Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing. ACM, pp. 395–401, 2010.

[7] A. Akram, "CRACS: Cross-Cloud Access Control Service for Multi-Cloud SaaS Applications", 2023.

[8] Network-to-Amazon VPC connectivity options, https://docs.aws.amazon.com/whitepapers/latest/aws-vpc-connectivity-options/network-to-amazon-vpc-connectivity-options.html, August, 2024.

[9] T. Dancheva, U. Alonso, and M. Barton, "Cloud benchmarking and performance analysis of an HPC application in Amazon EC2", Cluster Computer, vol. 27, pp. 2273–2290, 2024.

[10] AWS Elastic Disaster Recovery, URL:https://aws.amazon.com/tw/disaster-recovery/, August, 2024.

[11] Patel, G. and S. Patel. "Managing multi-cloud deployment with Morpheus", International Journal of Innovative Technology and Exploring Engineering, vol. 9, no. 3, pp. 1239-1243, 2020.

[12] S. T. Makani, B. P. Panchakarla, and S. R. Pulyala, "Enterprise-Grade Hosted VPN Services with AWS Infrastructure", Journal of Engineering and Applied Sciences Technology, 1-7, 10.47363/JEAST/2022(4)199, 2022.

[13] M. Marei, S. E. Zaatari, and W. Li, "Transfer learning enabled convolutional neural networks for estimating health state of cutting tools", Robotics and Computer-Integrated Manufacturing, Volume 71, 2021.

[14] S. Li et al. "PyTorch Distributed: Experiences on Accelerating Data Parallel Training", 2020.

[15] Kasm Workspaces, URL:https://kasmweb.com/, August, 2024.