“Beihai” RISC-V Cloud Computing Platform

Abstract

This paper introduces the "Beihai" RISC-V cloud computing platform, which integrates customized RISC-V hardware, a Kubernetes-based cloud management system, and optimized applications. The platform adopts a three-layer architecture consisting of heterogeneous computing infrastructure, a cloud-native management layer, and a container image repository. It manages a RISC-V cluster with several thousand cores. Based on this platform, we have completed the adaptation of core cloud-native components and end-to-end recompilation, thereby advancing the development of the RISC-V cloud computing ecosystem and providing support for related research and practical applications.

Introduction

With the continuous advancement of cloud computing technology, the need for a diversified range of underlying chip architectures has become increasingly evident. Against this backdrop, RISC-V, as an emerging open instruction set architecture, demonstrates significant potential for development due to its openness, flexibility, and modularity. Despite these advantages, the development of RISC-V in the field of cloud computing still faces many challenges. The lack of large-scale testing environments, a comprehensive cloud computing ecosystem, and integrated technical solutions spanning from hardware to underlying software and cloud applications are factors that severely hinder in-depth exploration and the promotion of applications of RISC-V in the cloud computing domain. To address these challenges, we have constructed the “Beihai” RISC-V cloud computing platform, committed to promoting innovative applications of the RISC-V architecture in cloud computing and the improvement of its ecosystem, providing an effective solution to the aforementioned issues, and propelling the development of RISC-V in the field of cloud computing.

Methodologies

“Beihai” RISC-V cloud computing platform adopts a layered architecture, integrating heterogeneous computing infrastructure, a cloud-native resource pooling and management platform, and an application container image repository. This design enables comprehensive support for the RISC-V instruction set and accelerates its adoption in cloud computing. The platform is built upon three key components:

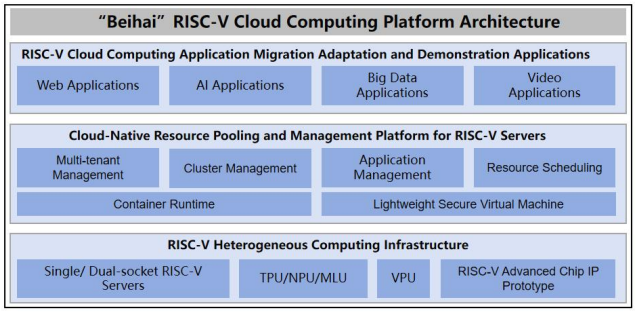

Figure1. The architecture of the “Beihai” RISC-V cloud computing platform

RISC-V Heterogeneous Computing Infrastructure

In building a heterogeneous computing infrastructure, the platform has successfully validated the adaptation of RISC-V servers with various AI accelerator cards, such as TPUs and NPUs. It has successfully deployed and run AI models, including small models like ChatGLM3-6B and large models such as Llama2-7B and Llama2-13B. We compare the inference performance of the Yolov5 small model on the RISC-V server with its inference performance on various other servers,demonstrating the efficiency of the RISC-V architecture in AI computing tasks.

Additionally,this platform manages our self-developed RISC-V VPU video transcoding card,TeleVPU,which features low cost,high concurrency,and powerful computing capabilities. This card supports parallel transcoding of 40 channels of 1080p/25fps video and is equipped with AI-powered intelligent video compression, reducing bandwidth and storage requirements by up to 80%. It provides an efficient and cost-effective solution for large-scale video processing scenarios, significantly enhancing the application value of RISC-V in the video processing domain.

Cloud-Native Resource Pooling and Management Platform for RISC-V Servers

In terms of the cloud-native pooling and management platform, it also integrates our self-developed cloud-native lightweight secure virtual machine, TeleVM,which supports the RISC-V H extension. By optimizing virtualization technology, this lightweight virtual machine reduces virtualization overhead by up to 80%, significantly improving resource utilization and system performance. Moreover, based on the OpenEuler operating system, we have completed the simulation and validation of this lightweight virtual machine on the Xiangshan open-source chip[1], verifying stability and compatibility of TeleVM in real-world operating environments.

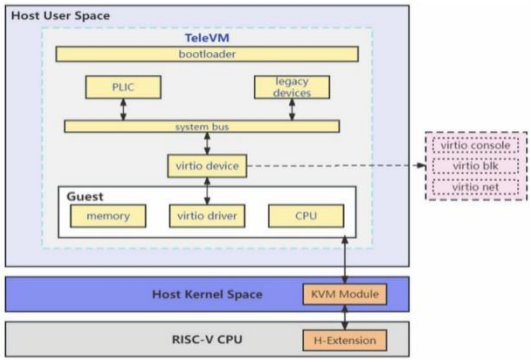

Figure2. Architecture of RISC-V Lightweight Virtual Machine TeleVM

Furthermore, “Beihai” has validated multi-tenant isolation, cluster management, application management, and an enterprise-grade application marketplace, supporting full lifecycle management for enterprise applications, including development environments, application distribution, and microservices management. These features establish a strong foundation for the cloud-native resource pooling and management of RISC-V servers, ensuring efficient execution and operation of cloud-native applications.

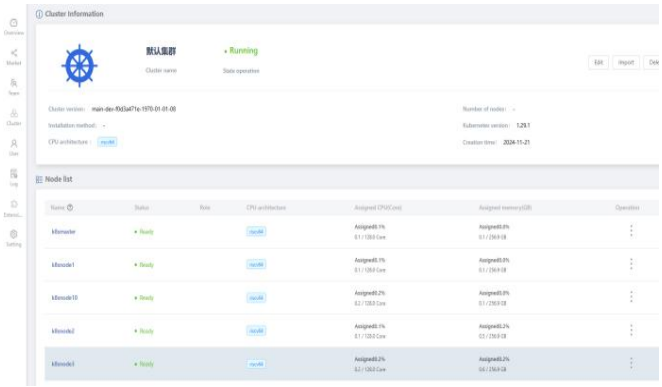

Figure3. Node List of the Cloud Computing Open Experimental Platform

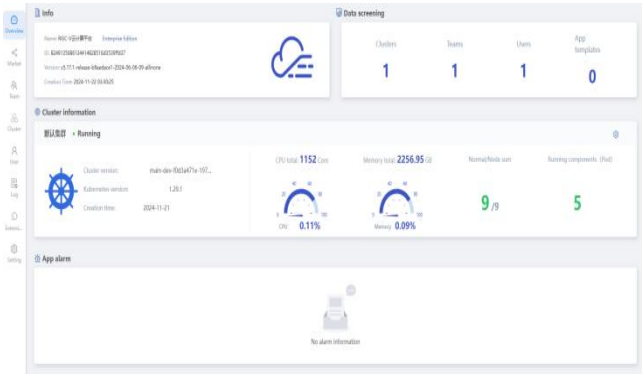

Figure4. Management Page of the Cloud Computing Open Experimental Platform

Application Container Image Repository for RISC-V Architecture

The platform has built a RISC-V application container image repository supporting over 100 container images. The repository includes images for operating systems, databases, middleware, cloud computing, big data, web services, and artificial intelligence, covering critical foundational software and core application domains.

Additionally, the platform has conducted real-world tests on various foundational application container images on the RISC-V cloud platform, including web servers, open-source search engines, message middleware, and load balancers. These tests validate the adaptability of this platform and its support for a wide range of applications, providing developers with extensive resources for application development and deployment.

By constructing and optimizing the application container image repository, the platform not only accelerates the growth of the RISC-V software ecosystem but also offers developers a convenient and efficient environment for development and deployment.

References

[1] OpenXiangShan. GitHub - OpenXiangShan/XiangShan: Open-source high-performance RISC-V processor[EB/OL]. (2025-01-01)[2025-02-06]. https://github.com/OpenXiangShan/XiangShan