Haizhou **Sun**

Machine Learning Cloud Platform Engineer · Software Architect

(+86) 18510343050 | sunhaizhou.ai@gmail.com | haiker2011

"Take big bites.Anything worth doing is worth overdoing."

Summary

I am currently a Development & Architect Engineer at Beijing SmartMore, specializing in machine learning platform development. I am also a cloud-native technology enthusiast actively involved in various open-source projects. My contributions include designing and developing machine learning platforms and multi-cluster scheduling solutions.

Work Experience

SmartMore Inc.

Beijing, China

Machine Learning Cloud Platform Architect

Apr. 2022 - Present

- China Computing NET: Implemented a multi-cluster heterogeneous scheduler based on the Karmada scheduling framework. This includes the management of Kubernetes container platforms, supercomputing platforms, and OpenStack platforms. Supported multi-cluster scheduling strategies include priority scheduling, fair scheduling, algorithmic-based scheduling, affinity scheduling, and rescheduling.
- cubeml Machine Learning Platform: Involved in the architecture and core functionality development of the cubeml machine learning platform. Implemented the cubeml machine learning platform based on the Volcano batch scheduling system, covering the development of scheduling, AI container, distributed development and training, inference services, etc. Also addressed and resolved preemption scheduling issues in Volcano open source project.
- Hyper-Converged System: Engaged in the architecture design and development of the Hyper Converged system. Main responsibilities include system architecture design and technology selection; leading the research, development, and implementation of distributed training, container development, and large-scale model training.
- Machine Learning Platform Implementation Projects: Projects include the Peng Cheng Laboratory multi-cluster Scheduling, Shanghai University Machine Learning Platform, Tongji University Machine Learning Platform, Chengdu Internet Machine Learning Platform, and more.

Kingsoft Cloud Inc.

Beijing, China

Machine Learning Cloud Platform Development Engineer

Mar. 2020 - Apr. 2022

- KingAI Machine Learning Platform: Research and Development of the AI Container Development Module. This module primarily includes SSH login, integration and authentication of Jupyter notebooks, and distributed development and training. It seamlessly integrates with the entire KingAI platform, providing an all-in-one machine learning development experience.
- KingAI Machine Learning Platform: Inference Service Development. Redesigned and restructured the inference service module, overhauled online inference, significantly streamlined the number of online inference services, and improved the stability of the inference services.
- KingAI Machine Learning Platform: Fine-grained Authentication for Services, Service Traffic Governance, and more.

The Institute of Computing Technology, Chinese Academy of Sciences (ICT)

Beijing, China

Student

Jul. 2017 - Apr. 2020

- Public Opinion Monitoring System: Refactoring and developing a new version of the distributed service framework for keyword analysis, sentiment analysis, named entity recognition, segmentation, text classification, news extraction, etc. Through a plugin mechanism, a revamped framework is created to significantly enhance service scalability.
- Machine Learning Experiment Platform: Set up a continuous integration Jenkins + Harbor + Kubernetes cluster to achieve continuous integration and delivery of services, accelerating delivery. Automated the laboratory process from code development to deployment testing environment, enhancing development efficiency.

Beijing Certificate Authority Co., Ltd (BJCA)

Beijing, China

Android Software Engineer

Jul. 2015 - Jul. 2017

- Product Development: Responsible for defining and developing requirements for the Android SDK of the Xinshou Shu product.
- Project Customization: Customized development for projects such as China Life and Shanghai Bank.
- Project Implementation: Responsible for providing technical support for project implementation using SDKs.

Education

UCAS(University of Chinese Academy of Sciences)

Beijing, China

M.S. in Computer Science and Engineering

Mar. 2017 - Aug. 2020

- Research focus on Natural Language Processing (Fine-grained Sentiment Analysis of Text).
- Cloud-Native Development.

UESTC(University of Electronic Science and Technology of China)

Chengdu, Sichuan

B.S. in Software Engineering

Mar. 2011 - Aug. 2015

- Major in Embedded Systems (Software Engineering).

Open Source Projects

volcano.io *GitHub*

Apr. 2023

- Expanded functionalities for job scheduling in the Volcano project, integrated Volcano with a machine learning platform, and provided support for the machine learning platform.
- Bug fixes related to preemption functionality in the Volcano open source project.
- Utilizing the Volcano plugin mechanism for container development, inference, and other tasks, developed plugins such as sshpiper and ingress.

karmada.io *GitHub*

Dec. 2023

- Development and refactoring of certain features in Karmada, integration of a heterogeneous fusion system with Karmada, providing support for multi-cluster management and scheduling.
- Advancing multi-cluster scheduling support for the Karmada community, promoting support for machine learning platforms within the multi-cluster scheduling framework, and driving the development of batch scheduling systems for multi-clusters, including big data and machine learning workloads.
- Built upon the Karmada multi-cluster scheduling framework, developed a proprietary multi-cluster heterogeneous fusion scheduling system (containers, Supercomputing Center, virtual machines). Implemented the system's deployment on the China Computing NET scheduling platform.

kubeedge.io *GitHub*

Jul. 2023

- Contributed to the Sedna open-source project, introducing new features such as synchronization configuration and pull request templates.

kubeedge.io *GitHub*

Feb. 2022

- Contributed to the KubeEdge open-source project by expanding Helm configurations and providing more convenient configuration options.

istio.io *GitHub*

May. 2022

- Maintenance of Istio official documentation in Chinese, including the translation of Istio release 1.1 update documents.

awesome-nlp-sentiment-analysis *GitHub*

Jan. 2024

- Collecting datasets, papers, and open-source implementations in the field of Natural Language Processing (NLP), with a particular focus on sentiment analysis, emotion cause recognition, and the extraction of evaluation targets and terms, 576 stars

getting-started-with-knative *GitHub*

Feb. 2019

- Translation of Chapter 3 of "Getting Started with Knative (Chinese Version)" with reviews for other chapters, 224 stars, To be printed and published by Pivotal Corporation