

张志强

✉ zhangzhiqiangcs@outlook.com · ☎ (+86) 131-2120-0924 · 🌐 zhangzhiqiangcs

教育经历

南开大学	软件学院/软件工程，本科	2012 – 2016
中科院计算所	计算机体系结构/区块链，硕士	2017 – 2020

工作经历

旷视, 北京	软件工程师	2020 年 6 月 – 至今
--------	-------	-----------------

AI PaaS 平台

- 参与 AI PaaS 平台的研发，构建了一个集成**算力、算法、智能体**管理的统一平台，实现推理系统的标准化和扩展性。
- 扩展算法管理模块，支持多种**大模型**的托管和推理，基于 vLLM 优化推理流程，通过 AI PaaS Gateway 提供 OpenAI API 代理，实现高效的模型推理服务。
- 支持多种智能体的接入，提供统一的 API 服务，实现会话管理、历史消息管理功能，提升对话质量和并发能力。
- 适配 Kubernetes 运行环境，实现模型推理任务的弹性调度和负载均衡，提高系统的稳定性和资源利用率。

私有云平台 MDP

- 基于 Kubekey 设计并实现了一套标准化 Kubernetes 交付方案，简化了异构环境下的集群部署流程，使 Kubernetes 能够在各种环境中快速落地，提高了系统的可维护性和交付效率。
- 构建 Kubernetes 交付平台，集成镜像中心、控制面 HA、监报告警、应用回滚等关键能力。基于 Helm Chart 实现模块化交付，使产品可自由组合组件，简化现场交付流程，提高可复用性和灵活性。
- 提供具有最佳实践的**基础设施能力**，涵盖基础设施的可观测性、备份恢复、扩缩容等能力。助力业务快速落地私有云架构。

私有云平台 DevOps

- 基于 Docker, Supervisor 自研容器编排引擎、集群自动化运维核心功能。在 ToB 领域、私有化、重交付场景中实现标准化交付、自动化运维和最佳实践。
- 包管理器**：自研内部包管理器，包含**命令行工具**，**统一制品仓库**，**控制器**。命令行工具提供依赖解析、配置传递、模板渲染、提交资源的功能。统一制品仓库提供存储、检索、版本管理的能力。控制器监听 apiserver 的资源，调度并部署到 DevOps 平台上。
- 加密授权体系**：基于 CodeMeter 和 Sentinel 设计并实现加密授权系统，增强核心产品的安全防护能力，防止未授权使用，并支持灵活的授权管理机制。

腾讯, 容器云开发组, 实习生, 北京

2019 年 6 月 – 2019 年 8 月

- 参与腾讯云容器存储（Ceph 集群）的开发和维护，优化存储服务的可靠性和性能。
- 排查线上 Ceph 集群异常问题，通过日志分析和 ceph status 监控诊断问题，修复存储节点异常问题，提高系统稳定性。

技能

- 精通 Golang，熟悉 Python，具备大型分布式系统开发经验。
- 开发平台主要为 Linux，擅长性能调优（profiling、GC 调优），有丰富的线上故障排查经验。
- 深入研究 AI Infra（vLLM、TensorRT），云原生（K8s、Operator、Service Mesh），内核技术（eBPF、cgroup）。
- 持有 Kubernetes CKA 认证，具备 Kubernetes Operator 开发经验，能独立编写 CRD、Controller。