才天一

18031521995 tianyicai@vip.126.com

教育背景

耶鲁大学 *2015 - 2016*

硕士，计算机科学，GPA: 4.0 / 4.0 *New Haven, CT*

亚利桑那大学 *2011 - 2015*

本科，计算机科学，GPA: 4.0 / 4.0 *Tucson, AZ*

工作经历

Cobo 2025/06 - 2025/08

MCP-Now, 大模型应用开发 北京

- 负责开发 MCP-Now 后端基础架构，通过 Supabase Edge Functions 实现 PostgreSQL 数据库访问。构建第一版用户账户系统，包含安全认证、注册流程和内容分享管理。使用 GitHub Actions 和 Supabase 分支集成功能建立 CI/CD 流水线，实现自动化测试和部署。
- 利用 AI 开发工具（Claude Code、Cursor），提升开发效率和代码质量。

摩尔线程 2025/02 - 2025/05

大模型训练平台, *GPU* 生态 北京

- 参与开发大语言模型训练平台，集成 Kubernetes、DL-Rover 和 Megatron-LM，提供一键式部署、自动扩展和实时监控，支持大模型训练工作流。
- 参与开发高性能分布式内存系统，利用共享内存、RDMA 和 RoCE 网络技术，为训练 checkpoint 在训练节点中提供缓存，减少节点故障恢复时间。

字节跳动 2021/10 - 2023/06

抖音推荐, 离线架构 北京

- 开发并维护抖音近线数据流、视频索引和候选库。开发基于 RocksDB 的 KV 存储索引写入模块，实现列式更新和行式读取的能力。
- 实现多租户场景下的全面资源监控和治理机制，防止个别业务高负载影响主推荐数据流，避免索引存储空间过度消耗。
- 改进抖音和 TikTok 视频数据库架构，将单体索引拆解为视频和用户索引。实现在线缓存命中率提升，增强用户信息一致性，同时简化社交等功能数据访问。
- 独立设计并实现了基于 Python 的轻量级流处理引擎，成功部署于抖音团购等多个垂类业务场景。核心能力：基于 *epoll* 的事件驱动架构，实现高吞吐 IPC 通信；利用 Kubernetes 原生的健康监控和自动故障恢复；向后兼容历史业务代码同时方便业务同学开发迭代，通过完全解耦垂类业务与主推荐信息流，实现特性迭代速度提升和可靠性。

幻方量化 2021/03 - 2021/08

AI 训练平台（萤火超算集群）, 存储与网络 杭州

- 开发基于 BeeGFS 的并行文件系统，通过 Linux Async I/O 优化异步写入性能，提升读写吞吐。构建自动化测试环境，基于 GitLab CI/CD 实现虚拟机自动部署和端到端测试，使用 InfluxDB 和 Grafana 实现性能测试结果可视化。

· 架构并实现基于 InfiniBand 网络的高性能 RDMA 异步通信库。提供兼容 Boost.Asio 的 API 接口，基于 *epoll* 实现事件驱动架构以获得最佳性能。开发跨平台 Linux 和 Windows 原生版本，实现 Windows 应用与 Linux 存储集群的无缝集成。

字节跳动 2019/08 - 2020/07

综合搜索, 在线架构 北京

· 维护并优化头条搜索在线服务，提供全网搜索和字节内部搜索平台功能。涵盖召回、扇出控制、排序、请求控制、内容安全等核心模块。

· 治理下游服务依赖，统一服务注册、访问和监控机制。通过多级缓存优化长尾服务响应，显著降低搜索延迟。

· 实施敏感内容标记流式更新机制，在检索早期阶段过滤风险内容，保障内容安全的同时提升搜索质量。

· 重构 Xapian 索引库的元数据存储，将文档属性聚合为单一 Protobuf 结构，减少检索阶段的元数据查询次数，使召回延迟降低约 10%。

· 负责多个垂类搜索从 Elasticsearch 到自研 C++/Xapian/Hadoop 架构的迁移，包括用户搜索、音乐搜索等。设计并实现支持大量异构小索引的平台架构方案。

Google 2018/03 - 2019/06

Google Cloud 微服务治理, 全球软件负载均衡（GSLB） *Sunnyvale, CA*

· 参与开发 GCP Traffic Director（托管式 Istio Pilot）产品。该产品作为控制层向数据平面的 Envoy 代理提供服务网格配置，实现服务发现、负载均衡和安全认证等核心功能。

· 设计并开发 Envoy 配置缓存、位置感知、Ingress 安全配置等核心组件。

· 设计 Envoy 代理在不同环境下的启动链接方式，涵盖 VM、GKE、原生 Kubernetes，设计 Envoy GCP 和客户数据中心的部署方案。

· 开发基于 Python 的集成测试框架，通过读取配置自动调用 GCP API 构建测试环境，大幅减少手工操作，提升产品迭代效率。

Datera, Inc 2016/09 - 2018/02

分布式块存储, 数据层 *Sunnyvale, CA*

· 参与开发基于 Linux B-Cache 和 iSCSI 协议的分布式块存储系统，提供存储卷的多副本读写能力，支持自动横向扩展，支持 SSD HDD 混合机型。

· 维护数据层核心功能，包括副本间一致性检查与修复、存储节点间负载均衡等关键特性。

· 参与设计数据加密、跨卷去重、本地数据中心到 AWS 远程备份等产品功能。

技术背景

编程语言	C/C++, Python, Java
系统架构	分布式存储, 流处理管道, 服务网格, 大模型训练
数据库	RocksDB, PostgreSQL, 键值存储
基础设施	Kubernetes, RDMA/InfiniBand, Linux 内核, CI/CD