黄成勉

(+86)13051682766 | 335462631@qq.com
https://vicety.github.io/

教育经历

瑞典皇家理工学院 2021年08月 - 2023年06月

Software Engineering of Distributed Systems 硕士

- GPA：4.75 / 5.0
- 相关课程：分布式系统 / 搜索引擎 / 编译原理

北京邮电大学 2016年09月 - 2020年06月

通信工程 本科 信息与通信工程学院

- GPA：3.6 / 4.0
- 相关课程：通信原理 / 数据结构 / 计算机网络 / 数据库技术与应用 / 微处理器与接口技术 / 数字信号处理

工作经历

百度 2020年07月 - 2021年08月

后端研发工程师 机器学习平台

- 负责百度机器学习平台任务中心的设计研发工作，基于Kubernetes List&Watch机制实现通用的任务提交与查询接口，支持所有内置与CRD资源类型
- 负责百度机器学习平台模型仓库与预测服务的设计研发工作，对训练得到的模型进行管理、优化，支持将模型发布为在线/离线预测服务
- 设计并实现分阶段的在线预测服务上线功能，接入多版本服务流量切分、日志收集、服务监控等特性
- 开发平台资源管控服务，通过与Kubernetes的状态同步，提供集群各节点资源余量的查询；基于Mutating Admission Webhook监听Pod创建事件，通过修改节点亲和性实现负载向特定GPU类型节点的调度

项目经历

- 基于Raft一致性协议实现分布式KV数据库，提供线性一致的读写操作，支持快照、数据分片、分片迁移、分片清理等功能
- 基于Gossip协议实现分布式哈希表，使用逻辑时钟在读取阶段解决冲突，提供最终一致性
- 扩展流处理引擎，支持对运行时工作流进行线性一致的算子状态查询
- 实现Scala子集编译器前后端与REPL Shell，支持符号重定义与尾递归优化
- 实现简易搜索引擎，通过分块索引实现对大数据集持久化倒排索引的增量构建；基于PageRank与tf-idf分数对查询结果排序；基于LRU cache实现buffer pool，提升查询速度
- 基于Seq2Seq模型与Attention机制实现机器翻译模型，在BLEU分数上复现原论文结果

实习经历

爱立信研究院（瑞典） 2022年06月 - 2022年08月

云原生DPDK应用可观测性调研

- 基于K3s、SR-IOV等技术搭建实验环境，将VF注册为可分配的集群资源，实现可伸缩的网卡直通容器创建
- 调研并验证容器内DPDK应用基于Trace、eBPF、中断等技术的流量观测方案

腾讯 2019年05月 - 2019年09月

测试开发实习生

质量数据运营平台：整合多平台质量数据，提供多维度可视化呈现与分析，为部门间业务质量提供横向比较。采用敏捷模式开发，前后端分别采用Vue.js和SSM框架

- 开发数据采集与指标计算服务，实现指标计算与查询功能
- 开发任务调度模块，实现每日定时数据采集任务与指标离线计算任务的调度
- 基于JUnit与Mockito框架编写单元测试，核心模块代码行覆盖率达到90%

全链路日志监控平台

- 基于Vue.js + ElementUI搭建调用链日志查询前端
- 为内部使用的RPC框架添加HTTP调用类型的日志上报支持，通过性能测试验证此功能对性能无明显影响并编写测试报告

专业能力

- 编程语言：（熟悉）Golang、Java、Python（了解）C/C++ 、Scala、Erlang、JavaScript
- 框架与工具：Kubernetes、Docker、MySQL、Kafka、Flink、Gin、Gorm、Vue.js、Git

Chengyang Huang

Professorsslingan 17 Room 1101, 114 17 Stockholm
(+46)76-9608-506 | chehuang@kth.se

Education

KTH Royal Institute of Technology Aug. 2021 - June 2023

Master of Science in Software Engineering of Distributed Systems

- Overall GPA: 4.75/5.0
- Relevant Courses: Distributed Systems / Data-intensive Computing / Compiler Construction

Beijing University of Posts and Telecommunications Sept. 2016 - July 2020

Bachelor of Engineering in Telecommunication Engineering

- Overall GPA: 3.6/4.0
- Relevant Courses: Data Structures / Computer Networks / Computer Architecture / Databases

Work Experience

Baidu July 2020 – Aug. 2021

Software Development Engineer, Machine Learning Platform

- Designed and built a **cloud-native** web service that deploys machine learning models as Serverless APIs to provide real-time inference based on **Golang**, **Docker**, and **Kubernetes**.
- Managed and Led the abstraction of a general task service that allows other **microservices** to submit and monitor Kubernetes custom resources without the need to learn Kubernetes APIs, which resulted in **15% fewer lines of code** and **77% fewer task monitoring issues** in corresponding services.
- Introduced a GPU scheduling policy that refines the scheduling granularity to GPU type level, satisfying training jobs which require to be executed on certain types of GPU.
- Integrate the system with MPI-Operator, a **Kubernetes Operator** allowing users to run distributed training jobs on Kubernetes using Horovod.

Projects

Independent Study on Distributed Key-Value Database Sept. 2021 - Dec. 2021

KTH Royal Institute of Technology

- Implemented a strongly consistent **distributed key-value database** with LevelDB that supports snapshot and configurable sharding based on the Raft consensus protocol.
- Implemented a distributed hash table, maintains routing table using gossip protocol, resolves inconsistency from client side using vector clock, provides eventual consistency.

Internship

Ericsson June 2022 - Aug. 2022

Research Intern

- Conducted research on fine-grained latency measurement of **DPDK** application in a cloud-native environment.
- Built an experimental environment with K3s and SR-IOV network plugin, allowing automatic NIC passthrough for dynamically created containers.

Tencent May 2019 - Sept. 2019

Test Development Intern, Quality Data Operating Platform

- Designed and built a task scheduling service, which sorts tasks with dependencies using topological sort and runs available tasks in parallel, reducing task running time by up to 60%.
- Collaborated with two engineers in developing data collecting and processing services with **Java** and **Spring Boot**, which ingests data from multiple data sources and aggregates it into customizable indexes.
- Ensured code quality with unit tests using JUnit and Mockito, achieved 90% coverage for core features.

Technical Skills

Programming Language: Golang, Java, Python, Bash, SQL
Platform & Framework: Kubernetes, Docker, MySQL, Kafka, Flink, Git, Gin, Spring, Zookeeper