

# 高策

才云科技 Tech Lead · Kubeflow Maintainer

(+86) 1592-1592-066 | ce.gao@outlook.com  
gaocegege.com(200k 阅读, 80k 访客) | gaocegege(1.4k followers) | gaocegege | @gaocegege

## 个人简介

目前是才云科技机器学习平台组 Tech Lead, 主要关注在云原生和 AI infra 领域, Kubeflow, coala Maintainer; 曾是上海交通大学东岳网络工作室开源与 MOOC 组组长, Google Code-In 2017, 2018 导师, Google Summer of Code 2018, 2020 导师, Google Summer of Code 2017 学生; 曾在蚂蚁金服, 触宝科技, 摩根士丹利等公司实习; 在 KubeCon China 2018, 中国 R 语言会议上发表过主题演讲。

## 工作经历

### 才云科技

Technical Lead

中国上海

2020 年 2 月至今

- 调研与设计产品新功能, 优化 Clever 产品架构
- 跟踪工业界趋势, 探索更多高级特性
- 负责 Clever 开源社区版的设计、实现与推广, 基于 caicloud/ormb 开源才云科技模型仓库系统
- 继续维护 Kubeflow 社区项目, 论文成果发表在 arxiv 上

### 才云科技

软件工程师

中国上海

2019 年 4 月至 2020 年 1 月

- 实现机器学习模型服务产品特性, 支持 GPU 共享、自动扩缩容等功能
- 基于 Kubeflow Katib 实现超参数搜索产品特性
- 继续维护 Kubeflow 社区项目

### 才云科技

合作研究

中国上海

2015 年 11 月至 2019 年 3 月

- 研究 Kubernetes 上对于机器学习基准测试的系统 kubeflow/kubebench, 研究成果发表在 IEEE AI4I'18 会议, 在 KubeCon China 2018 发表时长 30 分钟的演讲。
- 实现和维护 Kubeflow 在 Jupyter 上的内核项目 caicloud/ciao, 支持从 Jupyter 中发起分布式机器学习训练任务
- 为 Kubeflow 社区维护 TensorFlow 分布式训练支持 kubeflow/tf-operator 和超参数训练系统 kubeflow/katib
- 研究分布式机器学习任务在大规模机器集群上的调度, 研究成果发表在 ICA3PP'18 会议
- 实现基于 Docker 的持续集成与持续部署系统 Cyclone

### 摩根士丹利

项目实习生

中国上海

2017 年 2 月至 2017 年 8 月

- 为开源容器调度管理框架 treadmill 实现与 Kubernetes 类似的调度模型
- 调度延迟在 100 节点规模下与原本的调度器相比下降 12%, 但增强了其可配置性, 支持对节点上硬件资源的动态监控

### 上海触宝信息技术有限公司

实习大数据工程师

中国上海

2015 年 9 月至 2015 年 9 月

- 移植爬虫代码到新的平台, 优化重写部分过期的爬虫

### 蚂蚁金服(杭州)网络技术有限公司

实习 Java 研发工程师

中国浙江杭州

2015 年 7 月至 2015 年 9 月

- 在支付宝国际事业部创新业务组任职, 从事海外直购业务
- 实现部分包裹清关的逻辑和后台管理的逻辑

## 学术论文

### A Scalable and Cloud-Native Hyperparameter Tuning System

arxiv

Johnu George, Ce Gao, Richard Liu, Hou Gang Liu, Yuan Tang, Ramdoot Pydipaty, Amit Kumar Saha

2020.06

- In this paper, we introduce Katib: a scalable, cloud-native, and production-ready hyperparameter tuning system that is agnostic of the underlying machine learning framework. Though there are multiple hyperparameter tuning systems available, this is the first one that caters to the needs of both users and administrators of the system. We present the motivation and design of the system and contrast it with existing hyperparameter tuning systems, especially in terms of multi-tenancy, scalability, fault-tolerance, and extensibility. It can be deployed on local machines, or hosted as a service in on-premise data centers, or in private/public clouds. We demonstrate the advantage of our system using experimental results as well as real-world, production use cases. Katib has active contributors from multiple companies and is open-sourced under the Apache 2.0 license.

## GAI: A Centralized Tree-Based Scheduler for Machine Learning Workload in Large Shared Cluster

ICA3PP'18 (CCF-C)

Ce Gao, Rui Ren and Hongming Cai

2018.11

- 本文分析了机器学习模型的训练，识别了训练过程中的短板效应:与 CPU 训练相比，GPU 训练需要更高的网络带宽。这一观察启发了 GAI 的设计，GAI 是一个集中式的调度器，用于机器学习工作负载。它依赖于两种技术：1)树型结构。该结构分层存储集群信息，实现多层调度。2)扩展良好的优先级算法。我们全面考虑了模型培训工作的多个优先级，以支持资源退化和抢占。在 Kubernetes、Kubeflow 和 TensorFlow 上实现了 GAI 的原型。它是通过一个模拟器和一个真正的基于云的集群进行评估的。结果表明，在 DL 模型上，调度吞吐量提高了28%，训练收敛速度提高了21%

## Kubebench: A Benchmarking Platform for ML Workloads

IEEE AI4I'18

Xinyuan Huang, Amit Saha, Debojyoti Dutta and Ce Gao

2018.9

- Machine Learning (ML) workloads are becoming mainstream in the enterprise but the plethora of choices around ML toolkits and multi-cloud infrastructure make it difficult to compare their performance and costs. In this paper, we motivate the need for benchmarking ML systems in a consistent way, discuss the requirements of an ML benchmarking platform, and propose a design that satisfies the requirements. We present Kubebench, an example open-source implementation of an ML benchmarking platform based on Kubeflow, itself an open-source project for managing any ML stack on Kubernetes, a widely used container management platform.

## 演讲与分享

### 部分文章

#### 论文阅读笔记

GitHub

2016 年 11 月至今

- 记录了 100 余篇学术论文的阅读笔记，论文的方向多为分布式系统，虚拟化，安全，机器学习，超参数训练，网络模型结构搜索等领域

#### ormb: 像管理 Docker 容器镜像一样管理机器学习模型

gaocegege.com

2020 年 5 月 26 日

- 这篇文章介绍了如何利用 OCI 兼容的镜像仓库管理机器学习模型

#### 开源史海钩沉系列 [1] Ray: 分布式计算框架

gaocegege.com

2020 年 1 月 20 日

- 这篇文章介绍了 Ray 的设计与实现

#### Katib: Kubernetes native 的超参数训练系统

gaocegege.com

2018 年 3 月 7 日

- 这篇文章主要介绍了 Katib，一个 Kubernetes Native 的超参数训练系统

### 技术演讲

#### Kubeflow + Volcano 加速机器学习平台容器化进程

中国上海

华为开发者大会

2020 年 4 月

- 本次演讲介绍了 Kubeflow 与 Volcano 的集成能力

#### The Good, the Bad, and the Ugly of Managing ML Systems with Kubernetes

中国上海

人工智能产业发展联盟-开源开放组-第二次会议

2019 年 7 月

- 本次演讲，高策分享了才云科技在利用 Kubernetes 支持机器学习工作负载的方式，以及其中支持比较好的地方，以及会遇到的一些问题

#### 对 Kubeflow 上的机器学习工作负载做基准测试

中国上海

KubeCon China 2018

2018 年 11 月

- 本次演讲，介绍了基于 Kubeflow 的开源基准化工具 Kubebench，其帮助我们通过自动化和一致的规范，更好的理解 Kubernetes 上的 ML 工作量的性能特征。我们还说明我们可以怎样利用来自学术界和工业界的其他基准化工具，如 MLPerf 和 Dawnbench

#### Processing.R: 使用 R 语言实现新媒体艺术作品

中国上海

第十届中国 R 语言会议

2017 年 12 月

- 本次演讲介绍了 Processing.R。通过这一项目，用户可以利用 R 语言进行新媒体艺术作品的创作

## 其他经历

### Google Summer of Code 2020

中国上海

导师

2020 年 1 月至 2020 年 9 月

- 指导学生参与 Kubeflow 社区贡献

## Google Code-In 2018

导师

中国上海

2018 年 9 月至 2019 年 3 月

## Google Summer of Code 2017

学生

中国上海

2017 年 3 月至 2017 年 9 月

- 参与 Google Summer of Code 活动, 本次 GSoC 申请的接收率为 6%(1318/20651)
- 基于 R 语言在 Java 虚拟机上的解释器, 设计并实现了 Processing 在 R 语言模式 Processing.R, 获得 114 stars, 成为本次编程之夏 star 最多的项目

## 上海交通大学东岳网络工作室

MOOC & Open Source 组长

中国上海

2016 年 9 月至 2019 年 1 月

- 维护东岳网络工作室的技术博客以及知乎专栏: 东岳网络工作室团队
- 维护上海交通大学 XeLaTeX 学位论文模板: SJTUThesis
- 进行定期技术分享, 组织成员进行开源社区的贡献活动

## Google Summer of Code 2018

导师

中国上海

2018 年 1 月至 2018 年 9 月

- 指导来自印度的学生 Kilari Teja 完善与实现 coala Language Server 项目
- 在技术选型, 代码实现, 时间规划等方面进行讨论与指导, 定期进行会议交流以及代码 review
- 基于 Language Server Protocol 实现的 coala 在 Visual Studio Code 上的插件, 共被下载使用 686 次

## Google Code-In 2017

导师

中国上海

2017 年 9 月至 2018 年 3 月

# 教育经历

## 上海交通大学

软件工程专业硕士学位

中国上海

2016 年至 2019 年

- 研究方向为容器虚拟化, 集群调度

## 上海交通大学

软件工程专业学士学位

中国上海

2012 年至 2016 年

- 主修专业 (软件工程方向)

## 上海交通大学

法学专业学士学位

中国上海

2013 年至 2016 年

- 辅修专业 (知识产权方向)

# 荣誉与奖项

2017.10 Go 基金会中国黑客马拉松, 二等奖以及 PingCAP 专项奖

中国上海

2017.10 Intel Fellowship, 奖学金

中国上海