

LI XIANG

✉ lixiang99410@gmail.com · 📞 (+86) 18652936169

🎓 EDUCATION

Institute of Compute Technology, Chinese Academy of Sciences, Beijing Sept. 2020 – Jun. 2023

State Key Laboratory of Computer Architecture - Center for Advanced Computer System

Master in Electronic Information (Direction: Computer Architecture) GPA: 3.7/4.0 Rank: Top 10%

Nanjing University of Aeronautics and Astronautics, Nanjing

Sept. 2016 – Jun. 2020

B.S. in Computer Science and Technology (Honor Class) Rank: 6/100

🔧 PROJECTS

Research on replacement policy in last level cache of smartphone SoC

Jan. 2020 – Now

Main Contributor Mentor: Yuhang Liu

Brief introduction:

- Based on DRAMSim2 Simulator, use the trace captured in the real SoC when it running an application
- Propose and implement several replacement policy in the simulator, the result shows that cache hit rate increase by 5% using my policy.

Performance tuning of the GPT-2 model with billions of parameters

Jul. 2021 – Aug. 2021

Main contributor HPC Group - 2021 ByteCamp-Summer 2 / 30

- A GPT model with 15 billion parameters (20x baseline on PyTorch) was successfully trained on a machine with 8x V100s through gradient accumulation calculation, memory offload, parameter/gradient/optimizer state segmentation and mixed precision, and the training speed is only reduced by 30%.

3th DB Competition Innovation Cloud Performance Challenge of Tianchi Jun. 2021 – Jul. 2021

Main Contributor Preliminary round of High-performance analytical query track 39 / 1446

- Use data bucketing + two-level index, and R&W multi-threaded asynchronous methods to perform TopK queries on disordered data on persistent memory.

💖 HONORS AND AWARDS

Honorable Mention - The 2020 ICPC World Final

Oct. 2021

Outstanding Graduates - College of Computer Science and Technology, NUA

Jun. 2020

Gold Medal - The 2019 ICPC Asia Shenyang Regional Contest (ICPC Regional)

Nov. 2019

Gold Medal - The 2019 CCF Collegiate Computer System & Programming Contest (CCF-CCSP)

Oct. 2019

Alumni Association Scholarship - Nanjing University of Aeronautics and Astronautics

Nov. 2018

👥 EXPERIENCE OF INTERNSHIP

Bytedance Inc. Beijing, China

Nov. 2021 – Now

ML Platform Software Engineer Intern. Data-AML-Engine Mentor: Dongyue Wang

Brief introduction:

- Perform performance analysis and optimization of the distributed offline training engine for the advertising/recommendation model in Bytedance, and fine-tune the training speed of the online model.

Bytedance Inc. Beijing, China

Dec. 2019 – May. 2020

Backend Software Engineer Intern. Data-System-Platform Mentor: Jingtao Jie

Brief introduction:

- Build multiple data-flows and data warehouse with Kafka and HDFS using Python .

⚙️ SKILLS

- Programming Languages: C == C++ > Python > Java = shell
- Languages: English - Fluent, Mandarin - Native speaker

李想

✉ lixiang99410@gmail.com · ☎ (+86) 18652936169

🎓 教育经历

中国科学院计算技术研究所, 北京, 中国 2020 年 9 月 – 2023 年 6 月
计算机体系结构国家重点实验室-先进计算机系统研究中心 (预计)
专业硕士 电子信息 (方向: 计算机系统结构) GPA: 3.7/4.0 Rank: 前 10%

南京航空航天大学, 南京, 中国 2016 年 9 月 – 2020 年 6 月
学士 计算机科学与技术 (培优班) Rank: 6/111

📌 项目

智能手机 SoC 中末级高速缓存替换策略研究 2020 年 1 月 – 至今
主要负责 导师: 刘宇航, 中科院计算所先进计算机系统研究中心

- 实验基于 DRAMSim2 模拟器, trace 采用智能手机真实运行时抓取的访存 trace; 提出实现了多种替换策略, 命中率相较于 LRU 绝对提升 5%, 与前人研究结合后 DRAM 平均功耗相对降低 3%。

基于 DeepSpeed 的 NLP 超大模型训练调优 2021 年 7 月 – 2021 年 8 月
主要负责 高性能计算组 - 2021 年 ByteCamp 夏令营 2 / 30

- 通过梯度累积计算、显存 offload 内存、参数/梯度/优化器状态切分以及混合精度等方法将一个 150 亿参数的 GPT 大模型在有一台有 8 张 V100 的机器上成功训练起来 (参数量是 PyTorch 训练极限的 20 倍), 并且训练速度仅下降 30%。

第三届数据库大赛创新上云性能挑战赛 2021 年 6 月 – 2021 年 7 月
主要负责 高性能分析型查询赛道初赛 39 / 1446

- 使用数据分桶 + 二级索引, 以及读写多线程异步等方法在 pmem 上对无序数据进行 TopK 查询

♡ 获奖与荣誉

荣誉奖 - 2020 年 ICPC 世界总决赛 2021 年 10 月
优秀毕业生 - 南京航空航天大学计算机科学与技术学院 2020 年 6 月
金奖 - 2019 年 ICPC 亚洲区域赛沈阳站 2019 年 11 月
金奖 - 2019 年 CCF 大学生计算机系统与程序设计竞赛 (CCF-CCSP) 2019 年 10 月
校友会奖学金 - 南京航空航天大学 2018 年 11 月

👨‍💻 实习经历

北京字跳网络技术有限公司 北京, 中国 2021 年 11 月 - 至今
机器学习平台研发实习生 Data-AML-Engine Mentor: 王东岳

- 对于字节内部针对广告推荐模型的分布式离线训练引擎进行性能分析及优化, 并对业务线上模型进行训练速度精细调优。

北京字跳网络技术有限公司 北京, 中国 2019 年 12 月 – 2020 年 5 月
后端研发实习生 Data-系统-平台 Mentor: 介静涛

- 基于 Python 的后端 api 开发, 用 Kafka, HDFS 等基础设施, 完成了若干数据仓库的搭建工作。

📌 其他

- 编程语言: C == C++ > Python > Linux Shell
- 其他能力: 算法, 数据结构, Git, SQL, 等等。
- 曾担任南航微软学生俱乐部技术部部长, CCF 南航学生分会执行委员