Timothv Gu

+1 (226) 978 3698 | 2013tim.g@gmail.com | timgu.me | github.com/timchenggu123 | linkedin.com/in/tim-chenggu

EXPERIENCE

AMD, Senior SDE - AI Software | Markham, ON, CAN

- Developed and maintained the ROCm code base for driving cutting-edge ML workloads on the Radeon and Instinct GPUs C++, Python, HIP
- Demonstrated system level understanding of over 40 components of the ROCm stack and performed end-to-end debugging and optimization.
- Collaborated across multiple internal organizations to evaluate and propose new ROCm features, along with designing stack-level integration strategies.
- Ensured outstanding relationships with clients; drove open-source community engagement and satisfaction, as well as managing user expectations.

iQua Group, Distributed Cloud Computing Researcher | Toronto, ON, CAN

- Engineered high-performance distributed communication system to enhance wide-area distributed ML, achieving a $3 \times$ increase in throughput compared to existing solutions. Rust, C++
- Optimized multi-process co-routine scheduling in Rust with in-depth profiling using tools such as Tokio-Console and htop in a Linux environment, unlocking additional 30% performance boost. Linux, Profiling, Optimization
- Implemented a centralized controller in TypeScript with Node.js, facilitating up to **10,000** current connections via WebSocket and managing a PostgreSQL database for data-oriented API access. Node. js, TypeScript, PostgreSQL
- Packaged the components of the overlay network into Docker for streamlined deployment, significantly reducing setup time by 80% and improving scalability across diverse computing environments.

Huawei Canada, Software Engineering Intern - ML | Remote (Toronto, ON, CAN)

- Innovated a data-driven datacenter network traffic optimization algorithm with machine learning, achieving 100-10000× speed boost while reducing SDN control cycle time by up to an estimated 90%. Python, Pytorch, C++
- Built and trained a custom 30-billion-parameter model for data-driven optimization, achieving 95% of the theoretical optimum. PyTorch, PyTorch Geometrics
- Deployed distributed training pipelines to $8 \times$ NVIDIA V100 GPU cluster, improving training speed by over $10 \times$. Pytorch Distributed, DeepSpeed
- Actively led and organized a 4-person research team, hosted daily meetings, and presented research progress weekly in team-wide meetings. Leadership, Communication

Amazon Web Services, Software Engineering Intern | Remote (Vancouver, BC, CAN) Sep 2020 - Jan 2021

- Integrated business intelligence (BI) collection for AWS Cloud Formation Modules, delivering key metrics that help drive 20-million dollar business decisions. Kotlin, Java
- Created over 50 unit tests to ensure to ensure compliance with enterprise code quality standards. Java, Mockito
- Composed native AWS ETL pipelines, delivering 20000+ records daily to 3+ data warehouses. AWS Lambda, Apache Spark, AWS Glue, AWS Redshift, AWS DynamoDB, AWS Athena.
- Managed and deployed code to **21 AWS regions** and created key maintenance pipelines. Java, Kotlin, Ruby

Wayfair, Data Engineering Intern | Boston, MA, USA

- Innovated internal automatic big-data model verification pipelines, accelerating data verification process by up to 400×, saving an estimated 8 hours of manual labor per week. Python, MySQL
- Managed internal data ETL pipelines between multiple data storage solutions for over **20-million** sales and product records. MySQL, PostgreSQL, VerticaSQL, Microsoft Access, GCP
- Developed and maintained complex data schema with over 200 relations. MySQL, PostgreS QL

SKILLS _

Software Development	Python, Rust, TypeScript, Java, C, C++, Kotlin, JavaScript, Ruby, Paxos, Raft
Data Engineering	Spark, PostgreSQL, MySQL, MongoDB, Redis, AWS DynamoDB, AWS Reshift, AWS Athena, AWS S3
Web Development	Next.js, React, TypeScript, Html5/CSS, Ruby on Rails, AWS
Machine Learning	LLM, GNN, Transformers, PyTorch, HuggingFace, Scikit-Learn, Horovod, DeepSpeed, NCCL
System and Tools	CUDA, ROCm, HIP, Linux/Unix, MacOS, Vim, VsCode, Docker, Github
EDUCATION	

University of Toronto, M.ASC in Computer Engineering | GPA: 4.0 / 4.0

- Thesis: "Accelerating Internet-Scale Distributed Machine Learning". Distributed Systems, Cloud Computing, Network *Communications, Distributed Machine Learning*
- University of Waterloo, B.ASC in Computer Engineering | CGPA: 3.73 / 4.0
- Graduated Honors Co-op program with Distinction; ranked top 20% in class.

PUBLICATIONS

• (First Author) "An ML-Accelerated Framework for Large-Scale Constrained Traffic Engineering", IEEE International Conference on Distributed Computing Systems, 2024

Sep 2016 - Jun 2022

Sep 2022 - May 2024

Sep 2022 - Aug 2022

Sep 2021 - Aug 2022

Jan 2020 - Mar 2020

Aug 2024 - current