

MINGYU GUAN

(+1)470-334-6144 · alice.my.guan@gmail.com

EDUCATION

Georgia Institute of Technology
Ph.D in Computer Science

Aug. 2019 - May. 2024
(GPA:4.0) Atlanta, GA

The Chinese University of Hong Kong(CUHK)
B.S. in Computer Science with Honours, First Class

Aug. 2017 - May. 2019
(GPA:3.5) Hong Kong, China

Sun Yat-Sen University(SYSU)
B.S. in Electronic Information Science (2+2*)
*A joint program offered by SYSU and CUHK.

Sep. 2015 - Jun. 2017
(GPA:4.0) Guangzhou, China

SKILLS

Programming Languages:

C/C++, Python, Java, SQL

Research Interests:

Systems for Deep Learning in Graph, Graph Processing, Big Data Analytics

PROJECTS

System for Dynamic Graph Neural Networks at Scale

May. 2021 - present

- Supported efficient dynamic GNN training in both single machine and distributed settings;
- Leveraged the computational structure in the GNN-RNN approach to propose cross-layer optimizations;
- Used time-step fusion to reduce the underutilization of GPUs and cached message passing to eliminate redundant neighborhood aggregations;
- Designed a simple distributed data parallel dynamic graph processing strategy that enables scalable dynamic GNN computation;
- Incorporated the above techniques on DGL and PyG, two widely popular GNN processing frameworks.

Processing Billion-scale Dynamic Graphs on a Single Machine

Jan. 2020 - Jul. 2021

- Introduced the design of cell abstraction, allowing a significant reduction in overall storage space as well as enabling a simple, yet effective load-balancing strategy;
- Proposed an API and execution model tailored for streaming graphs by incorporating a hybrid edge- and vertex-centric API coupled with the *edgeChanged* API to allow a timely reaction to graph changes;
- Designed a technique for concurrent analytics on streaming graphs, which fully exploits the similarities in data access among concurrent graph processing jobs.

Automating Massively Parallel Heterogeneous Computing

Jan. 2020 - May. 2021

- Modeled input program as a hierarchical data flow graph (HDFG) to perform a set of graph-based operations and transformations for automatic optimization and parallelization;
- Performed purity checking automatically by traversing abstract syntax tree(AST) module;
- Inferred types of variables and objects automatically with both static analysis and dynamic analysis.

System for Serving ML Inference Pipelines

Sep. 2019 - Dec. 2019

- Implemented serving models in docker containers across ML frameworks such as TensorFlow, PyTorch, Caffe;
- Used Ray Serve as the base platform to communicate with containers via RPC;
- Supported pipeline and batched queries functionality.

Distributed Online Analytical Processing (OLAP)

May. 2018 - Apr. 2019

- Supported OLAP on Husky, which is a general-purpose distributed computing system developed by the system laboratory at CUHK;
- Used the platform of Husky to implement the By-Layer cubing algorithm in Apache Kylin;
- Implemented SQL engine and customized query optimization rules on Husky using Apache Calcite.