Chenghao Lyu

Ph.D. Candidate in Computer Science

Amherst, MA
chenghao@cs.umass.edu
(+1) 857-506-4527
https://chenghao.pages.dev/

Research Interests

My research interests lie in big data analytics systems, machine learning, and multi-objective optimization, focusing on (1) adaptive query execution and optimization, (2) query performance modeling via deep neural network, (3) automatic system configuration recommendations via efficient cost-performance optimization. My designed optimizers have demonstrated their advancements in Apache Spark and AliCloud.

Education

University of Massachusetts — Amherst, MA
Ph.D., Computer Science, GPA: 3.94/4.00 — (Expected) 01/2025
Advisors: Professor Yanlei Diao and Professor Prashant Shenoy
Thesis: "Intelligent Cost-Performance Optimization for Big Data Analytics Systems"

Fudan University — Shanghai, China
M.S., Computer Science — 06/2018
Advisor: X. Sean Wang, entrance exams waived
B.S., Electronic Engineering — 06/2015
Rank: 10/100, pre-admitted

Professional Experience

Ecole Polytechnique — Paris, France
Scientific Collaborator — 09/2021 - 03/2024

- Proposed the first adaptive query execution (AQE) optimizer for Spark SQL.
  - Improved latency and cloud cost by up-to 60% and 20%, with an overhead of 0.7-0.8s on TPC-H and TPC-DS.
  - Designed a hybrid parameter tuning framework to optimize Spark context at compile-time and adapt AQE parametric rules to runtime statistics during execution.
  - Broadcast with an oral presentation, a poster and an open-source release at VLDB'24, influencing over 20 research teams and diverse industry groups including Databricks, Microsoft, AWS, Alibaba, etc.

Alibaba DAMO Academy — Remote & Hangzhou, China
Research Intern — 02/2020 - 09/2021

- Integrated an intelligent resource optimizer to Alibaba's default scheduler Fuxi.
  - Reduced latency by 36-37% and cloud cost by 37-75% across three production workloads in Alibaba.
  - Built predictive models at the data partition level, achieving 9-19% weighted mean absolute percentage error.
  - Designed polynomial-time algorithms for task-machine placement and resource packing problems (NP-hard).
  - Gave oral and poster presentations at VLDB'22, exchanging ideas with over 100 attendees.

University of Massachusetts — Amherst, MA
Research Assistant — 01/2019 - 08/2021, 03/2024 - Present

- Developing uncertain-aware robust models to enhance the effectiveness of cloud optimizers. (ongoing)
- Improved and completed `UDAO`, a unified data analytics optimizer proposed by our team.
  - Achieved a 50x speedup by contributing a multi-objective gradient-based solver (accepted in ICDE'21).
  - Led the release of `UDAO` library on PyPi as the first contributor of a team of five.
  - Demonstrated `UDAO` at VLDB'19 and IP Paris'21, guiding over 100 attendees through the project.

Selected Publications

[1] **Chenghao Lyu**, Qi Fan, Philippe Guyard, Yanlei Diao: A Spark Optimizer for Adaptive, Fine-Grained Parameter Tuning. (PVLDB 2024)

[2] **Chenghao Lyu**, Qi Fan, Fei Song, Arnab Sinha, Yanlei Diao, Wei Chen, Li Ma, Yihui Feng, Yaliang Li, Kai Zeng, Jingren Zhou: Fine-Grained Modeling and Optimization for Intelligent Resource Management in Big Data Processing. (PVLDB 2022)

[3] Fei Song*, Khaled Zaouk, **Chenghao Lyu***, Arnab Sinha, Qi Fan, Yanlei Diao, Prashant Shenoy. Spark-based Cloud Data Analytics using Multi-Objective Optimization. (ICDE 2021)

[4] Khaled Zaouk, Fei Song, **Chenghao Lyu**, Arnab Sinha, Yanlei Diao, Prashant J. Shenoy. UDAO: A Next-Generation Unified Data Analytics Optimizer. (PVLDB 2019, demo)

Selected Awards

- Dissertation Writing Fellowship at UMass Amherst (Fall 2024)
- Tung OOCL Scholarship, 3x Outstanding student at Fudan University
- Pre-admission for B.S. and entrance examination waived for M.S. at Fudan University
- 4× Champion in Fudan Basketball Cup
- First Prize in China C9-league Dragon Boat Competition (Summer 2015)

Skills

- **Programming:** Proficient with Python; Comfortable with Scala, Java, Cpp; Beginning Rust.
- **Mentoring:** TA at UMass Amherst, guiding a class of 50+ MS students to build a knob-tuner.
- **Misc:** Apache Spark, PyTorch, Jupyter Notebook, Latex, etc.

Professional Service & Outreach

- **External Reviewer**: KDD 2022
- **Volunteer:** VLDB 2020

Graduate Coursework

- Systems for Data Science, *with Prof. Peter F. Klemperer*, Teaching Assistant — Spring 2024
- Algorithms for Data Science, *with Prof. Andrew McGregor*, Teaching Assistant — Spring 2023
- Database Design & Implementation, *with Prof. Yanlei Diao*, Teaching Assistant — Fall 2019
- Advanced Algorithms, *with Prof. Ramesh Sitaraman* (A) — Spring 2021
- Optimization in Computer Science, *with Prof. Madalina Fiterau* (A) — Spring 2021
- Reinforcement Learning, *with Prof. Philip S. Thomas* (A) — Fall 2020
- Distributed Machine Learning & Data Mining, *with Prof. Peter J. Haas & Don Towsley* (A) — Spring 2020
- Machine Learning, *with Prof. Benjamin Marlin* (A-) — Fall 2019
- Distributed & Operating Systems, *with Prof. Prashant Shenoy* (A) — Spring 2019
- Neural Networks: A Modern Introduction, *with Prof. Erik Learned-Miller* (A) — Fall 2018
- Database Design & Implementation, *with Prof. Gerome Miklau* (A) — Fall 2018

References

- **Yanlei Diao** (Professor, Ecole Polytechnique) — yanlei.diao@lix.polytechnique.fr
- **Prashant Shenoy** (Professor, UMass Amherst) — shenoy@cs.umass.edu
- **Peter Haas** (Professor, UMass Amherst) — phaas@cs.umass.edu
- **X. Sean Wang** (Professor, Fudan University) — xywangCS@fudan.edu.cn