

# Xinjie (Edward) Hu

xha102@sfu.ca | +1 672 866 7668 | English, Français, 汉语 | <https://edx-h.github.io/>

## EDUCATION

### Simon Fraser University (SFU)

Burnaby, Canada

3<sup>rd</sup> year Ph.D. candidate in Computing Science

Sep. 2024 – Present

- **Publications:**

1. **[VLDB 2026]** *Benchmarking the Full Pipeline of Materialized-View-Based Query Rewriting*. Xinjie Hu, Zhengjie Miao.

### Southern University of Science and Technology (SUSTech 南方科技大学)

Shenzhen, China

Bachelor of Economics, Finance major (GPA: 3.75/4.00)

Sep. 2016 – Jul. 2020

- **Honors:** University Scholarships for three consecutive years (Top 10%).

## RESEARCH INTERESTS, CERTIFICATION & SKILLS

- **Research Interests:** Materialized view-driven optimization; AI4DB task scheduling on GPU.
- **Certification:** AWS Certified Solution Architect.
- **Language:** Python, C++.

## RESEARCH & PROFESSIONAL EXPERIENCE

### Simon Fraser University (Database Systems Lab)

Burnaby, Canada

Research Assistant

Sep. 2024 – Present

- **Next-generation materialized view-driven query optimization with LLM:**

1. **[VLDB 2026]:** built the first end-to-end MV-driven query optimization benchmark to jointly evaluate all stages with equal attentions — candidate enumeration, view selection and query rewriting. The work benchmarked over 7 engines from the industry, and proved the necessity of studying holistic pipeline-style evaluation, and identified the existing bottlenecks which were never discovered before.
2. **Ongoing:** developing next-generation sound and safe MV-driven query optimization. Sound: Upgrading the pipeline with an LLM as a semantic reasoner to drive candidate enumeration and rewriting; Safe: the pipeline paired with a verifier that validates the equivalence of every LLM-generated rewrite before it reaches the optimizer.

- **AI4DB task scheduling on GPU:**

1. **Ongoing:** optimizing how AI4DB tasks (learned query optimizers, LLM UDFs, semantic query optimizers) share GPUs, where training and inference jobs are short, bursty and latency-sensitive; exploring workload-aware placement, request batching and GPU space/time sharing so that model serving meets query-latency targets without dedicating a device per task.

### Sohu 搜狐 (Intelligent Media R&D Centre)

Beijing, China

Data Engineer and Product Manager for Advertisement

Sep. 2020 – Mar. 2024

- **Data Warehousing:** designed schemas and the layered table framework, and wrote queries in Hive SQL; analysed query plans and tuned complex queries via partitioning, bucketing and memory/concurrency tuning; used Sqoop to exchange data between RDS and HDFS/Hive; adopted Impala for the application layer, including dashboards and visualisations.
- **Cloud Migration:** migrated 1 PB of data from on-premise to Tencent Cloud.
- **Real-Time Data Analysis:** automated the detection and treatment pipeline for fraudulent traffic using Apache Doris, Python and Redis, increasing efficiency by 90%.
- **Distribution Service:** built a C++ real-time data distribution service that pushed Kafka updates to downstream services with sub-second latency. Implemented in-memory state replication with Redis recovery and reliable fanout via ZeroMQ.

### SAP (Digital Aged Care Team)

Singapore

Consultant & Data Scientist

Jan. – May 2021

- Reviewed the literature on clinical quick methods for diagnosing Parkinson's and dementia.
- Used OpenCV to process hand-drawn spiral images from hospitals; applied histograms and dimensionality reduction to extract features capturing asymmetry and smoothness patterns.
- Built an AdaBoost model with hyper-parameter tuning and a training pipeline to predict Parkinson's probability, and combined clinical domain knowledge to interpret the model's decisions.

### TikTok (Headquarter)

Beijing, China

Digital Marketing & Ads Product Marketing Manager

Mar. – Jul. 2020

- Operated the ads measurement product Brand Lift Study (BLS) on TikTok and Helo; designed the “control-expose” experiment logic, analysed the change and significance of key conversion indices (awareness, purchase intent), and built dashboards in Tableau.
- Wrote HQL queries to select target audiences in the database.
- Researched competing products from YouTube and Facebook and supported the PM in product iteration, analysing strategies to improve the product's reliability and intelligence.
- Aligned product operation with third-party organisations (Nielsen, Kantar) and local customer solution teams entirely in English; advised customers on ads creative solutions to improve monetization rate.