

WEIYUAN WU

youngw@sfu.ca · wooya.me · github.com/dovahcrow

EDUCATION

-
- | | |
|--|-----------------------|
| Ph.D. Candidate in Database, Simon Fraser University, Canada | Sep. 2019 - |
| M.Sc. in Database, Simon Fraser University, Canada
Thesis: "Enabling SQL-ML Explanation to Debug Training Data" | May 2017 - Sep. 2019 |
| B.S. in Computing Science, UESTC, China | Sep. 2012 - July 2016 |

RESEARCH INTERESTS

SQL & Machine Learning Debugging, Data Intensive System

PROFESSIONAL EXPERIENCE

-
- | | |
|--|---|
| Database Engineer Intern
<i>Memory Optimized Distributed Database</i> <ul style="list-style-type: none">• Piloted the application of the userpagefault in database page management.• Addressed the out of memory issue by integrating the userpagefault page management using C++.• Implemented the coroutine support for the page management component. | Sept. 2021 - Dec. 2021
Tencent |
| Tech Advisor
<i>Pan-European Digital Derivatives Exchange</i> <ul style="list-style-type: none">• Researched different IPC methods, storage and recovery solutions based on the latency and reliability requirements.• Designed the system architecture from zero to one, including matching engine, order entry gateway and risk engine. | Mar 2021 -
D2X Group |
| Project Leader
<i>The Fastest Library to Load Data from DB to DataFrames in Rust and Python</i> <ul style="list-style-type: none">• Conducted related work investigation and performed extensive evaluations.• Implemented the core pipeline of the system. It accelerates data loading by 13x and reduces the memory footprint by 3x comparing to Pandas, the most popular data tool.• Designed the DSL for easily extending the library. The DSL allows ConnectorX to support 7+ mainstream databases and 4 most widely used dataframes. | Jan 2021 -
github.com/sfu-db/connector-x |
| Project Leader
<i>Data Preparation in Python</i> <ul style="list-style-type: none">• Designed and implemented the core system including a module for EDA, a module for data collection and a module for data cleaning.• Managed a team with 20+ members. Established the team processes for communication, code committing, code review, issue triage and release.• Achieved ~300k downloads and over 1k Github stars within the past two years. | May 2019 -
dataprep.ai |

PUBLICATIONS

-
- Xiaoying Wang*, **Weiyuan Wu***, Jinze Wu, Yizhou Chen, Nick Zrymiak, Changbo Qu, Lampros Flokas, George Chow, Jiannan Wang, Tianzheng Wang, Eugene Wu, Qingqing Zhou:
ConnectorX: Accelerating Data Loading From Databases to Dataframes VLDB 2023, In Submission
- Lampros Flokas, **Weiyuan Wu**, Yejia Liu, Jiannan Wang, Nakul Verma, Eugene Wu:
Complaint-Driven Training Data Debugging at Interactive Speeds SIGMOD 2022
- Yejia Liu*, **Weiyuan Wu***, Lampros Flokas, Jiannan Wang, Eugene Wu:
Enabling SQL-based Training Data Debugging for Federated Learning VLDB 2022

Brandon Lockhart, Jinglin Peng, **Weiyuan Wu**, Jiannan Wang, Eugene Wu:

Explaining Inference Queries with Bayesian Optimization

VLDB 2021

Jinglin Peng*, **Weiyuan Wu***, Brandon Lockhart, Song Bian, Jing Nathan Yan, Linghao Xu, Zhixuan Chi, Jeffrey Rzeszotarski, Jiannan Wang:

**DataPrep.EDA: Task-Centric Exploratory Data Analysis
for Statistical Modeling in Python**

SIGMOD 2021

Xiaoying Wang*, Changbo Qu*, **Weiyuan Wu***, Jiannan Wang, Qingqing Zhou:

Are We Ready For Learned Cardinality Estimation?

VLDB 2021

Weiyuan Wu, Lampros Flokas, Eugene Wu, Jiannan Wang:

Complaint-driven Training Data Debugging for Query 2.0

SIGMOD 2020

Weiyuan Wu, Lampros Flokas, Eugene Wu, Jiannan Wang:

Towards Complaint-driven ML Workflow Debugging

MLOps 2020, Demo

SKILLS

Frameworks: Tensorflow, Pandas, Numpy, Scikit-Learn, Dask

Programming Languages: Rust (7y), Python/Cython (4y), C++, Typescript, SQL

Platforms: Docker, AWS, Solana