

Tianhao Gu

Email: gtianhao@umich.edu

EDUCATION

University of Michigan, Ann Arbor

Ann Arbor, MI

M.S. in Data Science (Computer Science Track), M.S. in Quantitative Finance

Aug. 2020 – Expected Dec. 2022

- Courses: Database System, Web System Development (Full Stack), Data Mining, Machine Learning, Information Retrieval, Data Science in Python, Advanced Programming Design, Statistical Learning in Regression, Operating System

Rensselaer Polytechnic Institute

Troy, NY

B.S. in Mathematics (Computational Mathematics Track); Minor: Finance

Aug. 2016 – May. 2020

- Cumulative GPA: 3.81 / 4.00 [Baccalaureate Honor: Magna Cum Laude Award]
- Courses: C++ Data Structures, Operations Research & Algos, Data Resource Mgt, Data Mathematics, Computational Optimization

TECHNICAL SKILLS

Programming Languages: C++, Python, SQL, Java, JavaScript, HTML, R, Matlab, Julia

Database & Web: PostgreSQL, MongoDB, JDBC, SQLAlchemy, Flask, ReactJS, Rest API, CSS, MapReduce

Other Tools: Linux, Shell Script, AWS, File System, Git, Latex, Wind

PRACTICAL EXPERIENCE

Appen Data Technology Co., Ltd

Shanghai office, China

Data Engineering Intern, R & D Department (Engineering China Team)

Jun. 2021 – Aug. 2021

- Develop Python scripts to build and maintain data pipeline of various projects in image, audio, video, etc, such as generating data collection reports (eg. extract image EXIF information), staging to Quality Assessments Platform via AWS S3 APIs (eg. traceback vendor rating of video project), completing sessions' delivery & packaging processes (eg. trim-pad audio files for format uniform).
- Use SQLAlchemy framework in Python for extracting data from PostgreSQL database. Improve concurrency and control file Read-Write via applying thread pool executor and file lock. Update codebase in SQLAlchemy data table definition for ORM abstraction.
- Google Map Project: Use Python to integrate different files of store info and polygon info into one final file containing JSON data plus other attributes and prompts. Correct data format, precision and characters to support uploading in mobile app development.
- Complete several troubleshooting tasks, such as writing SQL code to detect database issue and then update PostgreSQL database, using debugging tool to find mapping key problems for solving data flow issue, detecting disk storage issue from IO-Write error in log.
- Use Python to implement a bloom filter to check the duplicate rate of text messages in all history sessions of one project in file system, match files and specific messages that contain duplicate texts. Create binary file to store and test intermediate results.
- Use crontab in Linux to configure auto-run of scripts. Meet with PM in Australia, US and UK offices frequently to follow up demands.

Guotai Junan Securities Co., Ltd

Shanghai, China

Sales & Trading Intern, Fixed Income (FICC) Department

May. 2019 – Aug. 2019

- Use Wind to collect data for rate securities issued, and assist in securities biddings. Conduct credit rating materials and market reports.
- Perform statistical analysis on sales and inventories of trust products, followed by performance evaluation and target index adjustment.

RELEVANT PROJECTS

Web System Development | University of Michigan

Sep. 2021 – Expected Dec. 2021

- Write hand-coded HTML files & CSS files, and template HTML files from JSON data. Write Python program (with jinja2, click library) that takes as input HTML templates, JSON data and static files, and generates Instagram websites of static contents.
- Build server-side dynamic Instagram websites using Flask, Python and HTML. Create SQLite Database for backend connection. Support CRUD for user account registration, posts, comments, following, etc. Write Shell Scripts to test results and deploy to AWS.
- Use JavaScript, Rest API and ReactJS to build Instagram websites with client-side dynamic pages. Deploy to AWS EC2. (future)
- Use Python socket to build a multi-process & threaded MapReduce server that executes user-submitted MapReduce jobs. (future)
- Build a scalable search engine for users to enter queries by React and Flask. Build a MapReduce indexer by Hadoop pipeline script.

Database System Implementation | University of Michigan

Feb. 2021 – Apr. 2021

- Design ER-Model via relational database structure. Use SQL to create an Oracle SQL database to store data of Facebook users with various information from the public database, load out to public dataset format and create external views for displaying the data.
- Use JDBC in Java to build a Java application that executes SQL queries to identify various information about Facebook users along with relationship between users, and store results in specialized data structures. Optimize code to reduce runtime of querying.
- Use JDBC in Java (HashMap, etc) to extract all data from Facebook Database and export a JSON file containing all users' information. Use JavaScript (along with MapReduce model, etc) to make query collecting information about users and build MongoDB database.
- Use C++ to emulate disk, memory and bucket storage. Implement the partition phase and probe phase of Grace Hash Join (GHJ) algorithm, and index linear hashing technique of database internal structures.

Machine Learning Models for analyzing stock return | University of Michigan

Mar. 2021 – Apr. 2021

- Use Python to crawl S&P500 stock data and construct technical indicators like RSI, Stochastic Oscillator, Bollinger Band, etc.
- Use sliding windows method to perform train-test-split, apply Random Forest Regressor (GridSearchCV for mode improvement), GBRT, Lasso, NN2 to predict returns. Calculate and visualize MSE, R^2 of test-predicted comparison, and show feature importance.

RESEARCH EXPERIENCE

Binomial tree algorithm for pricing American Options

Troy, NY

Senior Research Project; Supervisor: Prof. Chjan Lim

Oct. 2019 – Dec. 2019

- Use Matlab to implement binomial tree algorithm for pricing American options and compare results with actual price by historical data.
- Compare and visualize results of call/put option via testing early exercise period, time to maturity, number of steps, price convergence.
- Write a 12-page report to analyze all cases and results above, and parameters setting like historical volatility, risk free rate, tree depth.