Haoyang Zhang

EDUCATION

Master of Science, Georgetown University — **Washington D.C., U.S.**

Major: Computer Science | Current GPA: 3.96/4.0 | Anticipated Graduation: May 2023 — **Aug 2021 – Present**

Bachelor of Engineering, Beijing Jiaotong University — **Beijing, China**

Major: Computer Science and Technology | Major GPA: 3.34/4.0 — **Sep 2016 – Jun 2020**

SKILLS

- Solid expertise in **Python, Java, Object Oriented Programming, Algorithms, Data Structures, Design Patterns, Kubernetes, AWS, SQL, NoSQL, Git, Spring, Unix/Linux, Keras, Pytorch, Spark, Scala, PyQt, C/C++**.

EXPERIENCES

Research Assistant | McDonough School of Business, Georgetown University — **Jun 2022-Nov 2022**

Project: Disparities in 311 Cases and Its Impact on The Fairness of Data-driven Decision | Advisor: Emisa Nategh

- Counseled and budgeted the **cloud-based** project workflow, trimmed development cost by **50%** and time by **70%** compared to training and tuning models from scratch.
- Designed and built a **serverless** data processing workflow with **AWS Lambda** and **AWS S3**. Improved the processing time with parallelized serverless function calls by **1000x** than traditional cloud server instances, while keeping the cost similar, reduced data processing time by over **90%**.

Algorithm Intern | R&D Dept, Beijing Computing Center — **Jul 2019 – Sep 2019**

- Facilitated and helped the build of backend of proposal analysis system, including implementing text similarity, text classification, abstract generation algorithms, and corresponding backend **RESTful API** with **Flask** in **Python**, and **ORM SQL** frameworks.
- Refactored monolithic applications to a **Microservice** and Component based architecture, breaking components into Pods and containers for **Kubernetes** clusters.
- Implemented an image correction algorithm which reduced color RMSE by **60%**, which was integral to ensuring the product's performance.

Research & Development Intern | Baidu APP R&D Dept, Baidu Inc. — **Jul 2018 – Aug 2018**

- Top tier internet company in China.
- Completed optimizations and manufacturer customization requirements with self-testing for the Baidu Mobile App for Android in **Java**, of which 80% pushed to master, including adapting different notification push services for Chinese users, modifying default page styles for different phone models, adding or removing specific entries for preinstalled OEM versions, and so on.
- Expanded team technology stack by evaluating new UI toolkit Flutter.

SELECTED PROJECTS

Project: Toy-DB: Implementation of Relational Database Management System — **Feb 2022 – May 2022**

Main Contributor | Georgetown University, Washington D.C. | Advisor: Ophir Frieder

- Implemented an RDBMS in **Java**, with Visitor **Design Pattern** and composite **data structures**, supporting nearly **full SQL syntax**, including implicit join, expression updates, and arbitrary expression evaluation for WHERE conditions.
- Implemented cost-based and rule-based query optimization, integrity constraints and achieved **sub-1-second responsiveness** manipulating up to **one-million-record** tables, which is close to commercial grade.
- Now open source: github.com/CoreJa/ToyDB

Project: Music Data Analysis — **Aug 2021 – Dec 2021**

Group Leader | Georgetown University, Washington D.C. | Advisor: Lisa Singh

- Scraped and collected feature data for over 500K songs and artists from sources including Spotify, Wikipedia and allmusic.com, larger than any public dataset, stored in **MongoDB NoSQL** database. Worked on missing value imputation, outliers identification and duplication removal with statistical methods using **Pandas**, **Numpy** and **Scikit-learn**.
- Analysed influence and trends caused by popular artists with clustering, regression, ANOVA, classification, and network analysis using **Scikit-learn**, **Scipy** and **NetworkX**. Utilized **Matplotlib** and **Plotly** for visualization.

Project: Prediction on Taxi Drivers' Income Based on GPS Data — **Mar 2019 – Mar 2020**

Main Contributor | Institute of Network Science and Intelligent Systems, Beijing Jiaotong University, China | Advisor: Huaiyu Wan

- Built a high-quality dataset to describe behaviors of taxi drivers in Qingdao using multiple dimensions from hundreds of Gigabytes of raw GPS data with a **distributed system** in the workflow of **MongoDB** and **Spark**; extracted multiple features like the empty rate, work time, and profit from the spatial–temporal data.
- Designed a brand-new multi-input **RNN** model with human-related features, environment-related features, and income data input simultaneously using **PyTorch**; the RMSE for predicting drivers' income improved by **8.3%** using the dataset as compared to **LSTM**.