Ziheng (Jack) Chen

zihengchen2000@gmail.com | zihengjackchen.com | github.com/zihengjackchen | US Permanent Resident

Education

University of Illinois Urbana-Champaign (UIUC) Aug 2023 – Expected May 2025
Master of Science in Computer Engineering GPA: 4.00/4.00
University of Illinois Urbana-Champaign (UIUC) Aug 2019 – May 2023
Bachelor of Science in Computer Engineering (Highest Honor) GPA: 3.99/4.00

Leadership: ECE Graduate Student Advancement Committee Member, CS/ECE 374: Introduction to Algorithms
Achievements: Ackmann Family Scholarship, I-Promise Scholarship, Top 5% in LeetCode Contests

Skills

Programming Languages: Python, C++, C, Go, SQL, JavaScript, TypeScript, Bash, Assembly, CUDA, Scala
Big Data and Cloud: Databricks, MySQL, MongoDB, Splunk, Envoy Proxy, Airflow, Kafka, Spark, Redis
Technologies and Tools: Git, Linux, Docker, Kubernetes, Rancher, Azure DevOps, pandas, PyTorch, pytest, CARLA, TCP/IP

Experience

Machine Learning Engineer (co-op), StoneX Group – Chicago, IL Jan 2023 – June 2023

- Created a proof of concept pipeline to benchmark prototypical commodity indices, enabling their potential product release
- Analyzed the profitability of commodity indices corresponding to 200+ expert-provided configurations based on historical prices, surpassing performance targets by 23.3% in 10-year total return with the highest-performing index
- Accelerated benchmarking by 3000% from code optimization in Python using pandas and parallel execution in Databricks
- Collaborated closely with marketing associates to refine evaluation criteria by simplifying technological concepts
- Devised a procedure to version track Databricks workflows in the repository and developed a Python script for CI/CD in Azure DevOps to automate workflow migration from production environments using APIs, eliminating manual processes
- Verified migration of Delta Live Tables and Spark SQL commands in complex scenarios through unit and integration testing

Data Engineer Intern, StoneX Group – Chicago, IL Aug 2022 – Dec 2022

- Enhanced and optimized a Streamlit web application through security integration and ETL pipeline improvements
- Researched for solutions to integrate Okta authentication into the web application, deploying Envoy Proxy microservice, achieving user access control using bearer tokens and MS SQL Server, enhancing data confidentiality
- Developed a dynamic usage analysis dashboard in Splunk, delivering associate usage and category-specific insights
- Migrated the data curation ETL pipeline from Apache Airflow to a continuous Databricks workflow, efficiently managing staging data with Azure Blob Storage, and updated the application using Docker to reduce load times by 30x to 5 seconds
- Managed microservices using Rancher for Kubernetes management and documented the overall system design for clarity
- Utilized Agile methodologies to ensure project flexibility and timely completion, adapting to evolving project needs.

Data Engineer Intern, Ecolab – Saint Paul, MN May 2022 – July 2022

- Profiled tables in Snowflake using SQL to examine key statistics, showing outliers and trends, enhancing data integrity
- Identified 201M invalid rows and created an interactive Snowsight dashboard to support business decisions
- Automated query generation with JinjaSQL in Python, utilizing Alation catalog, accelerating data quality evaluation
- Analyzed 19300 hours of Service Requests logs of dishmachines, cleansed using Python from Snowflake, processed using Azure Cognitive Service API to identify 6 common issues and their locations, informing potential refresh strategies

Projects

Traffic Risk Assessment and Mitigation – Autonomous Vehicles, Machine Learning, Safety Aug 2023 – Feb 2024

- Enhanced the resiliency of AVs in unfamiliar and accident-prone scenarios with a novel traffic risk assessment method
- Validated the method by unit-testing a prototype using designed experiments in CARLA Simulator and Argoverse dataset, including real-world geometric and semantic metadata, lane boundaries, geometric LiDAR, and ring camera information
- Created multi-threaded data generation and testing pipelines and boosted efficiency by 200% using subprocess in Python
- Constructed 6000 unfamiliar scenarios from NHTSA pre-crash typologies and trained lightweight Reinforcement Learning Agents in PyTorch to preemptively brake using the traffic risk as an indicator, reducing accidents by 72.7%

Friction Model Improvement in CARLA Simulator – Autonomous Vehicles, Simulation Sept 2022 – May 2023

- Improved the realism of the CARLA Simulator by refining the friction dynamics model as a function of weather parameters
- Generated 8000+ vehicle traces in diverse weather, uncovering 300% more accidents due to a vulnerable planning module