

Alex Chen

Brooklyn, NY | itsalexchen@gmail.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

EDUCATION

New York City College of Technology, CUNY - Brooklyn, NY

June 2024

Bachelor of Science in Data Science | GPA: 3.8

Relevant Coursework: Database Fundamentals, Machine Learning for Physics, NoSQL Technologies, Information Retrieval

TECHNICAL SKILLS

Languages: Python, SQL, JavaScript, TypeScript, HTML, CSS, Sass, Golang, Rust, Java, PHP, Bash

Libraries/Frameworks: Pandas, React, Jest, TensorFlow, Node.js, Express, Supertest, Airflow, Flask, Next.js, Django

Technologies: Git, Postgres, Docker, dbt, MongoDB, Redis, Supabase, Terraform, Ollama, Excel, RabbitMQ, Neo4j, Figma

EXPERIENCE

Metropolitan Transportation Authority

New York, NY

Tech Fellow (Data Engineer Intern)

June 2023 - Present

- Architected and built data pipelines for **capturing and storing over 100MB of transactional data per day** about transit reports using Python, dbt, and Airflow, resulting in synchronized data across multiple systems for 2+ teams
- Engineered a CDC data pipeline to create a data source for an analyst team using Debezium, Kafka, DuckDB, and dbt, **leading to a 75% reduction in data latency** and ensuring up-to-date reporting for business analysts.
- Led the development of a PDF data pipeline using Apache Tika, TheFuzz, and Regex, capturing 25+ data points about work train requests from scanned reports and **delivering data to stakeholders with accuracy surpassing 70%**
- Collaborated with engineers and stakeholders to develop a mobile-friendly web application using Python, Flask, and Postgres to streamline the work train request process, **increasing workflow efficiency by 200%**
- Orchestrated the deployment and maintenance of dockerized data-related services with **three-nine availability**, minimizing downtime and ensuring stakeholders have access to critical data, documentation, and applications
- Optimized the data ingestion and storage capabilities of data pipelines through testing and using different Python data libraries, **resulting in a 40% improvement in the completion times of the pipelines**

Develop for Good

Remote

Software Engineer Volunteer (Data)

April 2023 - Aug 2023

- Collaborated with 8 engineers and the client to spearhead the development and execution of a CDC data pipeline using Python, SQL, and Apache Airflow, **resulting in a 40% reduction in the client's computing and storage costs**
- Extracted data from BigQuery using Python and the BigQuery API, ensuring efficient and timely transfer to a virtual machine for downstream processing, resulting in improved data availability for the data pipeline

The City University of New York

New York, NY

Energy Technology Intern (Backend)

July 2022 - Aug 2022

- Led End-to-End testing for an Excel workbook application by developing and executing comprehensive test plans using Python, **resulting in faster processing times by over 30%** and a significant reduction in application crashes
- Developed a Python script that automates data processing of raw data with over 100,000 data points into a report that can be referred back to later, **saving time manually querying energy consumption data by 50% for 2 analysts**
- Improved software reliability by implementing comprehensive unit tests using Python, covering critical functionalities of the Django application, **resulting in a 100% increase in code coverage from 50% to 100%**

PROJECTS

Job Tracker Web Application - [GitHub](#)

- Developed a full-stack web app using TypeScript React and Express that allows users to manage items in their job tracker while seeing the data about the roles through data visualizations
- Built a RESTful API backend with Express.js to help store role information within the Postgres database
- Engineered a role documentation automation feature by utilizing the cheerio library, **saving time for end users by 80%**

Wikipedia Information Retrieval System - [GitHub](#)

- Engineered a full-stack web application using Python, Docker, and Streamlit that allows users to efficiently retrieve information from Wikipedia articles for users through an intuitive search engine and a large language model
- Implemented microservices utilizing Flask, Docker, and Ollama to optimize the web application's core functionality, resulting in improved scalability for requesting data from the information retrieval system and the llama2 model

Portfolio - [Project Page](#)

- Designed and developed a portfolio website using React, and Styled-Components showcasing current and past projects
- Utilized GitHub Actions to automate the process of building and deploying a portfolio website, enabling faster and more efficient updates to the website