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: Data Structures, Database Fundamentals, NoSQL Technologies, OOP in Java, Information Retrieval

TECHNICAL SKILLS

Languages: Python, SQL, HTML, CSS, Sass, JavaScript, TypeScript, Golang, Java, Bash
Libraries/Frameworks: Pandas, Matplotlib, React, Jest, TensorFlow, Node.js, Express, Supertest, Airflow, Flask
Technologies: Git, Jupyter Notebooks, Postgres, MongoDB, Redis, Figma, Docker, Supabase, Excel, Neo4j, Tableau

EXPERIENCE

Metropolitan Transportation Authority   **New York, NY**
Tech Fellow (Data Engineer Intern)   *June 2023 - Present*
- Architected and built CDC data pipelines to **handle the ingestion and storage of over 50MB of transactional data per day** using Python, SQL, and Apache Airflow, resulting in synchronized data across multiple systems for 2+ internal teams
- Optimized the data ingestion and storage capabilities of CDC data pipelines by using different Python data libraries, **resulting in a 40% improvement in the completion times of the pipelines**
- Engineered a CDC pipeline that scrapes upstream data sources using Airflow and Python, **reducing data latency by 60%**
- Initiated and drove the development of a PDF data pipeline using Apache Tika, TheFuzz, and regex, capturing 25+ data pointers from scanned reports and **delivering data to stakeholders with an 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 of a RESTful API using Docker, resulting in **three nines availability to data** for stakeholders

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 web application's backend, utilizing PostgreSQL and Express.js to store role information and to create a RESTful API for client-server communication
- Engineered a role documentation automation feature by utilizing the cheerio library, **saving time for end users by 80%**
- Implemented GitHub Actions workflows for CI/CD, enabling rapid and reliable delivery of new features to production
- Constructed unit tests and integration tests with React Testing Library and Jest, achieving 70% code coverage and improving overall reliability and quality for end users

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

Emotion Detection Web App - GitHub
- Collaborated with a team of 3 to build a web application using Python, NTLK, and Streamlit, enabling users to conduct sentiment analysis on tweets and short-form text content with an accuracy of over 88%