

Yun Zhou

yz2685@cornell.edu | [linkedin.com/in/annyunzhou/](https://www.linkedin.com/in/annyunzhou/) | github.com/annsway | annsway.vercel.app/

Summary: 3+ years of experience in software development, specializing in distributed systems, real-time data pipelines, backend, and infrastructure development. Seeking a **Mid-Senior level Software Engineer** role.

Skills

- **Programming & Scripting:** Java, Python, JavaScript (ES6+), TypeScript, Bash, SQL
- **Frameworks & Libraries:** Node.js, React.js, Spring Boot, Spring MVC, GraphQL, RESTful APIs
- **Databases & Storage:** PostgreSQL, DynamoDB, Snowflake, NoSQL, S3, MySQL, Oracle
- **Infrastructure & Tools:** AWS, Docker, Kubernetes, Terraform, Argo CD, Helm, CI/CD, GitOps, Apache Kafka, Apache Spark, Apache Airflow, dbt, Grafana, Loki, Datadog

Work Experience

Flexport

Bellevue, WA

Software Engineer II – *Backend and Infrastructure Development*

2023 – Present

- **Kafka Streaming Platform Migration**
 - Engineered the migration of the Kubernetes native Kafka Streaming Platform to AWS Managed Kafka
 - Deployed Helm charts and Argo CD applications for the continuous deployment of Kafka resources
 - Integrated Prometheus and Grafana dashboards for real-time monitoring of Kafka and k8s resources
- **Real-time Data Pipeline Development and Management**
 - Built and maintained real-time data streaming pipelines in Java and Python, processing over 200 TB of data
 - Integrated IAM role-based authentication into a Go-based open-source DynamoDB to Kafka connector
 - Hardened Snowflake data warehouse security with Okta and network policy, achieving an 80% improvement
 - Optimized data orchestration and telemetry workflows in collaboration with data science teams
- **Development Experience Improvement**
 - Refactored the JS/TS codebase for GraphQL Gateway, resulting in a more efficient dev environment
 - Implemented PR checks, linters, and GitHub actions to enhance the reliability of CI/CD pipelines
 - Migrated monolithic application to microservice architecture using AWS, Docker, and CDK Terraform

Cornell University

Ithaca, NY

Research Assistant – *Full Stack Development*

2023

- Led the development of Cornell BioLab website to visualize large-scale genomics datasets via React & Node.js

HealthPartners

Bloomington, MN

Epic Clarity Business Analyst – *Data Integration and Data Visualization*

2017 – 2021

- Led the data integration project for 60 retail pharmacies using ETL tools and **SQL**, reducing report errors by 25%

Project Experience (<https://annsway.vercel.app/projects>)

Cornell University Data Pipeline Development

Spring 2023

- Built a data pipeline to map data from AWS Data Lake to Redshift, enabling seamless data warehouse integration
- Utilized Spark to process transactional and process data (e.g., user activity logs) and transform it for Redshift

NLP Project: Stock Performance Prediction

Spring 2022

- Scraped 300+ 10-K reports and historical stock prices using Selenium and BeautifulSoup API in Python
- Trained and tested SVM, Logistic, and Random Forest classifiers with NLP features using sklearn and spaCy

Full Stack Car Dealership Website Development

Fall 2021

- Developed backend services using Spring Boot and JDBC API for managing car listing data
- Designed and implemented a MySQL database hosted on AWS and interactive search/sales pages in React & CSS

Education

Cornell University

Ithaca, NY

Master's Degree in *Information Science* (GPA: 4.0/4.0)

2022 – 2023

University of Minnesota - Twin Cities

Minneapolis, MN

Bachelor's Degree in *Statistics* (Dean's List; Maroon Global Excellence Scholarship)

2014 – 2016