

# Tianyu Zeng

fzeng6@gmail.com (+206)-512-4349 Santa Clara, CA

## Technical Skills

**Languages:** Java, Python, Go, JavaScript, SQL

**Frameworks & Libraries:** Express, Guice, Struts, FastAPI, Jest, Spring

**Databases:** PostgreSQL, GCP BigQuery

**Dev Tools:** Kafka, RabbitMQ, Spark, Git

## Professional Experience

### IXL Learning Software Engineer, Marketing Data Platform

San Mateo, CA **Aug 2021 - Aug 2023**

- Led the design and implementation of a server-side tagging system using **Java**, **Kafka**, and **Struts 2**, enhancing conversion tracking accuracy and culminating in a 16.4% boost in accurately recorded conversion events.
- Constructed and maintained data streaming pipelines utilizing **Java**, **Kafka**, **ProtoBuf**, and **PostgreSQL**, facilitating the processing and loading of 20+ millions subscription and user data daily into Salesforce Marketing Cloud
- Developed RESTful APIs leveraging **Java**, **Struts 2**, **PostgreSQL**, and **Guice** to streamline the process and automatically generate data extensions in SFMC for launching geographically targeted email campaigns for IXL Live, increasing event registration rates significantly from 32% to 91% for the first time
- Designed an end-to-end solution in **Java** for daily SFTP batch uploads to reverse undesired transactions, and implementing an automated workflow for data retrieval and validation for accurate spending reporting
- Enhanced influencer marketing data management by seamlessly integrating multiple APIs for direct data ingress into **GCP BigQuery**, followed by routine data cleansing and dashboard updates, reducing data processing time by 27%.
- Contributed to the software development lifecycle by designing and implementing features for a subscription-based learning site, delivering clean and scalable code in **Java**.

### Hive AI Software Engineer

San Francisco, CA **Feb 2020 - Aug 2021**

- Established a self-service analytics tool using **React**, **Node.js**, and **PostgreSQL**, allowing data analysts to correct annotated documents and obtain useful metrics, resulting in a 30% drop in manual auditing time.
- Enhanced the labeling accuracy of existing data pipelines by 12.3%, a significant feat achieved by devising scalable rendering and ingestion workers leveraging the power of **Node.js**, **RabbitMQ**, **AWS S3**, and **Apache Mesos**.
- Implemented version control and tracking mechanisms in **Node.js** and **PostgreSQL** for over 500k image datasets and corresponding annotation data models, ensuring reproducibility and traceability of data changes.

### CUNA Mutual Group Data Scientist (intern)

Madison, WI **Jun 2018 - Aug 2018**

- Developed a gradient boosting tree algorithm in **R lang**, optimizing it for predicting auto insurance claims' severity loss, which resulted in a 7% reduction in mean square loss.
- Conducted data preprocessing in **Spark**, engineered features for data quality, and used **R** to build dashboards for real-time consumer sentiment analysis for stakeholders.

## Education

### Carnegie Mellon University M.Sc. in Software Engineering

Mountain View, CA **Dec 2023 (expected)**

### University of Wisconsin - Madison B.Sc. in Computer Science

Madison, WI **Jan 2016 - Dec 2019**

## Projects Experience

### Step Trending Electronic Data Interface (STEDI)

**May 2023 - Jul 2023**

- Developed a real-time analytics application using **Spark Structured Streaming** and **Python**; Set up and managed Spark Clusters to facilitate data processing and consumption from Apache **Kafka**.
- Utilized DataFrames for data aggregation, implemented composite DataFrame sinking to Apache Kafka, and performed visual inspections to ensure data accuracy.

### Emergency Social Network

**Jan 2023 - Apr 2023**

- Implemented RESTful APIs using Node.js, Express, and Postgres for message exchange, enhancing frontend and backend interaction, and integrated CircleCI to ensure seamless testing and automatic production deployment
- Ensured the quality of the application through Jest for unit and integration testing, leading to 83% overall code coverage