

## Libin Guo

832-668-4022 | Austin, TX | [libinguo89@gmail.com](mailto:libinguo89@gmail.com) | [LinkedIn](#) | [GitHub](#)

*Senior Software Engineer (Data Intelligence) with 7+ years of experience in distributed systems for large-scale data pipelines and data platforms and eval framework for agentic AI system*

## EXPERIENCE

### Apple, inc

#### **Senior Software Engineer(Data Intelligence)**

**Austin, TX | June 2024 - Present**

- Scaled and optimized Microservice app to automate maintenance spark jobs across 20 PB iceberg datalake, improving query performance and driving a 20% YoY increase in internal lakehouse adoption
- Built evaluation framework for internal text2sql/text2code LLM application, combining industry benchmarks with LLM-as-judge metrics to improve accuracy by 30%, reduce onboarding time by 100%
- Delivered a RAG enhanced chatbot w/ Lakehouse&snowflake MCP servers for runbook search, Q&A, table creation and answer data insight questions

#### **Software Engineer(Data)**

**Austin, TX | Dec 2021 - May 2024**

- Built and optimized billion-scale data pipelines (Airflow + Spark on AWS EKS), reducing runtimes by 75% and saving \$1.7M annually by optimizing Spark micro-batch execution
- Drove adoption of Snowflake + DBT across the analytics org by establishing standards and best practices — boosting team productivity and strengthening stakeholder trust in data quality

#### **JPMorganChase - Software Engineer(Data)**

**Dallas, TX | Oct 2020 - Nov 2021**

- Enhanced Kafka streaming application to ingest data into Oracle in microservice architecture
- Designed and implemented end to end streaming data pipelines w/ AWS Glue jobs and Lambda functions for ETL processes from AWS Kinesis to DynamoDB

#### **Healthmine - Data Engineer**

**Dallas, TX | Sep 2018 - Oct 2020**

- Designed, developed ETL data pipelines to load millions rows of data into Oracle, and migrated legacy Talend ETL workflows to Airflow to improve performance and maintainability
- Built Enterprise Data Warehouse in Postgres with daily ingestion pipeline in airflow , Spark and EMR Cluster to deliver business critical report with SQL in Apache Superset

#### **Qlarant Integrity Solutions, LLC - Senior Data Analyst**

**Dallas, TX | Mar 2016 - Aug 2018**

- Analyzed 100+ data samples to identify millions Medicare overpayments using statistical data-driven approach and automated adhoc SAS templates to reduce queries runtime by 75%
- Led proactive data analyses using SQL, SAS, and Tableau, uncovering hundreds of healthcare fraud leads

## SKILLS

- AI: Langchain, LangGraph, Langfuse, Tensorflow, RAG, weaviate, Mlflow, Dspy
- Languages: Python, Java, Linux, SQL
- Data: Apache Spark, Iceberg, Airflow, Kafka, Streamlit, DynamoDB, Snowflake, DBT, Pandas, Numpy
- Cloud Infra: AWS, Kubernetes, Docker, CI/CD, Kubeflow, Helm Chart

## EDUCATION

**University of Texas at Dallas - M.S., Information Technology and Management**

**GPA 3.82 | Dec 2015**

**Shandong Agricultural University - B.S, Computer Science; Minor in English**

**GPA 3.25 | Jul 2011**