

Bo-Jian (Eddie) Ho

Software Engineer

Sunnyvale, CA, USA, edwin.ho.bj@gmail.com, [linkedin.com/in/eddie-bojian-ho](https://www.linkedin.com/in/eddie-bojian-ho), +1-626-662-5830

With 2 years of professional experience building scalable data infrastructure and RAG systems. Skilled in designing data lakehouse architecture, distributed systems, ETL, and RAG pipelines, with expertise in Python, Java, and Machine Learning. Focused on system design, reliable data delivery, and cost-efficient infrastructure, with a goal to advance data-driven products and enable intelligent, real-time analytics.

Employment history

Member of Technical Staff, Mar 2026 – Present

Homekey Inc., Palo Alto

- Architected and implemented end-to-end *PostgreSQL* data model and storage infrastructure, migrating **9M+ California property records** from *MongoDB* with zero downtime and full data integrity validation
- Engineered ETL pipelines to ingest and process **900K+ geospatial records** from OpenStreetMap and municipal GIS databases, transforming vector data into optimized formats for BigQuery Geo queries
- Maintained and improved daily property acquisition pipelines ingesting **10K+ property listings** across Bay Area, Beverly Hills, and San Diego, orchestrated with *Dagster* for reliable and timely data delivery
- Developed production-grade semantic search API microservice leveraging *pgvector* for vector similarity search and semantic caching, **reducing query latency by 50%** and eliminating search timeouts under concurrent load
- Computed spatial index summaries for property tiles using *PostGIS*, accelerating neighborhood-level queries and enabling geospatial filtering for property search and analytics

Data Engineer, Aug 2025 – Mar 2026

Million Software Inc., San Francisco

- Developed production-grade REST APIs serving internal and customer-facing analytics at **20K+ daily requests**, achieving **sub-second response times** via *ClickHouse* materialized views, caching, and query optimization, with attention to data security best practices
- Reduced cloud costs by **\$2K/month** by migrating to managed services, utilizing ephemeral processing resources, optimizing *S3* storage tiering, right-sizing auto-scaling, and refactoring pipeline queries to cut redundant compute
- Engineered CDC pipelines using *Debezium*, *Apache Kafka*, and *ClickPipe* to sync **2TB+ historical data** and **10GB+ daily data** with zero downtime; reducing end-to-end data processing time by **70%**, and enhancing security engineering controls for data integrity
- Architected scalable data lakehouse infrastructure using *Apache Iceberg* on *AWS S3*, *PostgreSQL* (OLTP), and *ClickHouse* (OLAP), enabling high-throughput analytics and time-travel queries across **200TB+** of data, incorporating privacy infrastructure
- Implemented backend caching layer using *Redis* to reduce database load and improve response times for high-frequency service endpoints, ensuring secure data handling

Member of Technical Staff, Dec 2024 – Aug 2025

Chima Inc., San Francisco

- Developed end-to-end RAG system to enable semantic search for AI agents, architecting the full pipeline from query embedding generation using sentence transformers to vector similarity search with *Turbopuffer* and model-based reranking
- Optimized with embedding caching, batch processing, hybrid retrieval combining semantic and keyword search using *LangChain* and *LangGraph* to achieve consistent **sub-500ms** response time while handling concurrent user queries at scale
- Implemented incremental indexing pipeline using *Git* diff tracking and *S3 One-Zone* cache storage for selective re-embedding of modified code sections, **reducing re-indexing time by 50%** and achieving millisecond retrieval latency
- Established evaluation using LLM-as-a-judge methodology to validate semantic search performance against traditional grep-based search, demonstrating a **70% win rate** in relevance and accuracy across diverse query patterns
- Integrated semantic search as a *TypeScript* AI SDK agent tool, enabling agents to query knowledge bases, filter by metadata, memory, and chain searches dynamically for complex information retrieval tasks

Data Engineer, May 2024 – Nov 2024

Lesso America Inc., Los Angeles

- Built ETL pipeline monitoring and alerting system using *GCP Monitoring*, custom data quality tests, and *Airflow* sensors to detect volume anomalies, schema drift, and errors; cutting incident resolution time to **under 4 hours**
- Implemented automated invoice scanning feature using Mistral OCR, integrated directly into the internal finance system, eliminating manual data entry and enabling accounting clerks to process **30% more invoices** per day
- Established data governance policies including role-based access controls and column-level masking in *BigQuery* and *Data Catalog* to ensure compliance with data privacy requirements and secure handling of sensitive business information
- Created data model lineage visualizations enabling clear representation of complex data relationships and dependencies

Education

Bachelor of Science in Data Science, Sep 2021 – Mar 2024

University of California-San Diego, San Diego, CA

- GPA 3.8; Minor in Cognitive Science; Member of Data Science Student Society and Computer Science & Engineering Society

Skills

Python, SQL, Typescript, Java, React, Machine Learning, Distributed System, PySpark, AWS, GCP, BigQuery, Data Modeling, PostgreSQL, NoSQL, Data Structures, System Design, API Development, ML Pipelines, Airflow, Kafka, Iceberg, dbt

Links

GitHub: github.com, Personal Website: eddieho.xyz, Ami: ami.dev, Same: same.new, Langflow: langflow.org