Bo-Jian Ho

Data Engineer | AI Engineer

Sunnyvale, CA, USA, +1-626-662-5830, edwin.ho.bj@gmail.com

Professional summary

With 5 years of experience in data science and 1 year in full-stack development, I specialize in data modeling, cloud solutions, and tools like SQL, Clickhouse, and Airflow, as well as building web applications with React and Next.js. My work has improved operational efficiency and driven business growth.

Employment history

Data Engineer, Aug 2025 - Present

Million, San Francisco, CA

- Architected scalable data infrastructure using PlanetScale (OLTP), ClickHouse (OLAP), and AWS S3 data lake, supporting 1TB+ daily data with real-time Metabase dashboards for operations teams
- Engineered CDC pipelines with Debezium and streaming workflows via AWS MSK, migrating 100TB+ historical data with zero downtime and reducing processing time by 70%
- Optimized data storage with star schema and data tiering in ClickHouse and AWS S3, improving query performance by 45% and reducing storage costs by $5K monthly
- Built production-grade APIs handling 2B+ data point requests daily for operational analytics, enabling real-time cost optimization that identified $2K in monthly infrastructure savings
- Developed automated ETL workflows using dbt and Dagster, streamlining data flow and enabling faster business intelligence insights

AI Engineer, Member of Technical Staff, Dec 2024 - Aug 2025

Chima, San Francisco, CA

- Led R&D initiatives in developing intelligent integration agents, reducing manual API integration time by 60% through automated documentation research and configuration generation.
- Architected and implemented a multi-agent system with Langflow that automated integration setup across 5 platforms, handling authentication, endpoint configuration, and testing validation.
- Created specialized prompt engineering frameworks that improved agent accuracy by 40% and reduced hallucination rates in integration tasks by 35%.
- Established automated testing frameworks for integration agents, achieving 95% test coverage and reducing integration errors by 50%.

Data Scientist, May 2024 - Nov 2024

LESSO, Los Angeles, CA

- Trained machine learning algorithms and deep learning with PyTorch for claims forecasting models, K-means Clustering, XGBoost, and LSTM to predict demand and inventory management resulting in a 30% improvement in inventory turnover and a reduction in excess stock by 25% within 3 months.
- Automated the extraction, transformation, and load processes (ETL) using Python and SQL, leading to a 50% reduction in the time required to prepare data for analysis.
- Implemented a new E-R modeling system for efficient database structuring that streamlined data flow and reduced redundancy, resulting in a 25% increase in productivity for the data analytics team.
- Architected optimized data architecture strategies, such as Snowflake schema and virtual warehouse within Snowflake, improving data query performance by 40% and reducing storage costs by 25%.
- Developed a comprehensive data governance framework that included secure role-based access controls (RBAC) and data masking to ensure that sensitive data in Snowflake was protected.

Student Engineer

UC San Diego, San Diego, CA

- Optimized the performance of data warehouses by implementing indexing strategies and query optimizations that resulted in a 30% improvement in data processing speeds.
- Managed the lifecycle of big data solutions, including building and maintaining NoSQL databases such as Cosmos and Bigtable, which handled over 5 TB of data.
- Devised data validation protocols and error identification mechanisms to ensure a 99.9% accuracy rate in data reporting.
- Initiated and led the deployment of an automated data quality system that reduced manual error correction by 75%, significantly increasing data reliability for end-users.

Education

Bachelor of Science in Data Science, Sep 2021 - Mar 2024
University of California-San Diego, San Diego, CA
GPA 3.8. Member of Data Science Student Society

Courses

Certified Azure Data Engineer Associate
Microsoft Inc.

Advanced SQL for Data Scientists
DataCamp

Skills

SQL, Python, Typescript, React, Node.js, Spark, Kafka, DBT, Airflow, PostgreSQL, MySQL, MongoDB, BigQuery, Apache Airflow, Data Modeling, Snowflake, Communication, Teamwork, Time Management, Problem-Solving, Fast Learner, Continuous Learning.

Links

LinkedIn: linkedin.com, GitHub: github.com, Personal Website: eddieho.xyz, Ami: [ami.dev](ami.dev), Same: same.new, Langflow: langflow.org.