Jack Stehn

Data Engineer

San Francisco, California • jack@ste.hn • www.linkedin.com/in/jack-stehn • github.com/jstehn • ste.hn

PROFESSIONAL SUMMARY

Data Engineer with a strong foundation in architecting scalable, production-grade data platforms. I bring a software engineering mindset to orchestrating end-to-end ELT/ETL pipelines and modern data warehousing. I focus on championing CI/CD, testing, and modular design to build systems that are maintainable, observable, and impactful.

EXPERIENCE

Data Scientist (Lead: Data Engineering, MLOps) - Ed Pioneers Fellow   Sep 2024 — Oct 2025

Caliber Public Schools

- **Modern Data Stack Architecture:** Architected and built a scalable platform on GCP. Orchestrated ELT pipelines using Dagster, dbt, and dlt to ingest data from disparate SIS (School Information Systems) and HR platforms into BigQuery.
- **Engineering Maturity (ROI):** Engineered a comprehensive People Team data pipeline, reducing manual consistency checks from months of collective annual work to seconds. Built a custom automated data quality framework to catch inconsistencies between systems.
- **API Integration & Interoperability:** Solved complex data interoperability challenges by designing integrations between Schoolmint and PowerSchool. Reverse-engineered undocumented APIs to create a unified data model for cross-functional analysis.
- **Solo Data Lead:** Owned the full data lifecycle as the sole data professional, partnering directly with C-suite and department heads to move the org from manual spreadsheets to automated warehousing.
- **Data Democratization:** Updated the organization's data security policy and designed data literacy training modules, empowering school leaders to access real-time metrics without technical bottlenecks.

Data Scientist (Data Engineering, MLOps)   Aug 2021 — Feb 2023

SetSail

- **Pipeline Architecture (AWS):** Led a critical overhaul of the AWS data infrastructure (S3, Athena, EMR). Implemented 'SQL Push-down' strategies and asynchronous DAGs, reducing data processing latency by 75% and scaling to multi-terabyte datasets.
- **Scalable Data Modeling:** Architected scalable Star Schema data models and optimized ETL/ELT processes, ensuring data readiness for ML and LLM integration.
- **Engineering Best Practices:** Championed the adoption of CI/CD pipelines (GitHub Actions), unit testing (pytest), and Agile methodologies within the data science team.
- **Business Impact:** Contributed to infrastructure enhancements that achieved 33% faster ramp times, 16% higher revenue, and 15x ROI for customers.
- **Cross-Functional Leadership:** Collaborated with Engineering, Product, and Support teams to spec out large-scale infrastructure restructuring.

Full Stack Engineer   May 2019 — Aug 2019

Los Medanos College

- **Full Stack Development:** Independently designed and developed a web application (Python, JavaScript, PHP, PostgreSQL) to guide student program selection.
- **Data Acquisition:** Built a database of 500+ university programs via web scraping (Beautiful Soup) and rigorous data cleaning.
- **Stakeholder Consulting:** Consulted with college stakeholders (district, counselors) to define user needs.

EDUCATION

University of California, Berkeley   Aug 2019 — May 2021

Bachelor of Arts in Data Science (Domain Emphasis: Quantitative Social Science)

GPA: 4.00/4.00

Highest Distinction (Summa cum laude). Outstanding Data Science Undergraduate Award (Top of Class).