

Mark Shervey

Data Engineer

San Francisco, CA

markshervey@gmail.com
linkedin.com/in/mark-shervey
415.755.0388

Summary

Data engineer with 22 years in data platforms and warehousing, plus a strong software development background. Delivered production data systems across healthcare, media, and finance; led enterprise-scale ETL and pipeline work built for performance and scale, working closely with clinical, product, and business teams.

Employment History

Care Indeed, 2025–2026

Senior Data Engineer, Consultant

At Care Indeed, a Bay Area home care and home health provider serving seniors and families across the region, I owned the organization's cloud data platform, led the migration off its legacy EHR, and built the HIPAA compliance foundation for how clinical, caregiver, and operational data is stored and used.

Python, SQL, AWS (Redshift, S3, Glue, Lambda, SAM, RDS, CloudWatch), HIPAA, SOC 2

HHS Technology Group, 2024–2025

Lead Data Engineer

I led initiatives to improve how client organizations use their Medicaid and Medicare data. I worked closely with clients and internal teams to design and implement new features in HIPAA-compliant environments, mentored junior developers on best practices in Python, Snowpark, and AWS, and helped drive code quality across the organization. I audited our internal codebases on an ongoing basis, surfacing HIPAA and SOC 2 compliance gaps and prioritizing their remediation.

Python, SQL, Snowflake, Domo, Airflow, AWS (EC2, RDS, S3, CloudWatch), Linux, HIPAA, SOC 2

Milu Health, 2022–2023

Data Engineer, Consultant

Milu Health, a startup helping employees and self-insured employers save money on healthcare by presenting high-value care options, needed to analyze the pricing data insurers were required to publish under the federal Transparency in Coverage rule, which took effect in 2022. I built and maintained the pipelines that ingested terabytes of that data from insurers each month and made it available for application integration and in-depth analysis. The result let employees compare costs across their employer's insurance options and make more informed decisions, sparing both employee and employer unnecessary expense without sacrificing quality of care.

Python, SQL, AWS (EC2, RDS Aurora PostgreSQL, S3, Athena), Linux

Icahn School of Medicine at Mount Sinai, Critical Intelligence Center, 2020–2022

Director of Engineering, Lead Data Engineer

I led engineering for Mount Sinai's Critical Intelligence Center from its founding during the first two waves of the pandemic, an effort that brought together clinicians, informaticians, data scientists, and technologists from across the health system to focus on COVID-19 clinical research. I was a working manager, leading a small team spanning systems architecture, data engineering, and clinical data science. As one of the center's leads, I worked with the IRB to keep our solutions compliant with internal standards and HIPAA, and I was accountable for patient safety across all of them. I also partnered with hospital leadership to make sure business requirements were met.

I scoped and oversaw the creation of a multi-modal data warehouse including all hospital system clinical, genetic/genomic, waveform, imaging, and pathology data to enable COVID-19 clinical research, sparing researchers the task of interfacing with each individual source system across the hospital. I led the launch of COVID360, a consolidated view of individual COVID-19 patients surfaced directly in Epic Hyperspace, deployed into production and actively used by clinicians during care.

Python, R, SQL, AWS (EC2, RDS, S3), Azure (Virtual Machine, Azure Storage, Data Factory, Data Lake), Epic Systems (Hyperspace, Clarity, Caboodle, App Orchard/FHIR), SQL Server, Oracle, MongoDB, MySQL, OMOP/OHDSI, HIPAA

HumanFirst (acquired by ICON in 2024), 2019–2020

CTO, Data Engineer

I led engineering for the Atlas platform, which enabled organizations to run remote clinical trials using connected devices such as Fitbit and Apple Watch. Atlas indexed those devices, detailing their validated endpoints and the peer-reviewed research citing them.

Python, Next.js, SQL, AWS (EC2, RDS, S3), PostgreSQL, Airtable

Icahn School of Medicine at Mount Sinai, Institute for Next Generation Healthcare, 2016–2019

Staff Engineer

I ran engineering for a lab of 30+ clinical informaticians, enabling them to perform and scale clinical research across the entire health system's data. I led the prototype build for Lab100, a new type of wellness clinic, and managed its vendor relationships. I also co-founded the Center for Biomedical Blockchain Research, a group exploring how blockchain could be used in the health system. Separately, I stood up the pipelines and databases behind multi-study wearable sensor collection, streaming participant data into research-ready storage. Across all of this work, I owned compliance, ensuring each system met HIPAA along with the hospital's data privacy and IRB requirements.

Python, R, MySQL, PostgreSQL, AWS GovCloud (EC2, RDS, S3), Epic Clarity, OMOP/OHDSI, blockchain, HIPAA

MTGx/Viaplay, 2012–2016

Data Engineer

I joined Viaplay, a leading Nordic competitor to Netflix, and built the company's content recommendation platform and its foundational viewership database. I later moved to the Business Intelligence unit, reporting directly to the CEO, and migrated the reporting platform behind programming, content-bidding, and pricing decisions from Python and MySQL to an enterprise Spark/Hadoop environment on AWS EMR.

Python, Spark, Hadoop, Pig, Hive, Impala, Airflow, AWS (EMR, EC2, S3), MySQL

Wells Fargo, Wholesale Services Group, 2004–2012

Data Warehouse Engineer

I worked across multiple groups within the Wholesale Services Group (WSG), culminating in a seat on the Wells Fargo/Wachovia merger integration team. I designed and built data marts and warehouses for several WSG business units, drawing fragmented customer data into views built for each group's use cases.

Python, .NET, Oracle, SQL Server

Education

Washington State University | B.A., Management Information Systems

Relevant Publications

Disease heritability inferred from familial relationships reported in medical records. *Cell* 173 (7), 1692–1704.e11

Modernizing and designing evaluation frameworks for connected sensor technologies. *npj Digital Medicine* 3 (1), 37

Longitudinal autonomic nervous system measures correlate with stress and ulcerative colitis disease activity and predict flare. *Inflammatory Bowel Diseases* 27 (10), 1576–1584

Privacy-preserving methods for feature engineering using blockchain: review, evaluation, and proof of concept. *Journal of Medical Internet Research* 21 (8), e13600

Sleep in the natural environment: a pilot study. *Sensors* 20 (5), 1378