

Josh Goldstein

AI-Native Infrastructure · Senior Solutions Architect

76 Roebling St, Apt 1B, Brooklyn, NY 11211 305-321-6144 Josh4goldstein@gmail.com

<https://linkedin.com/in/joshsgoldstein> <https://github.com/joshsgoldstein>

HIGHLIGHTS

Technical strategist and hands-on solutions leader with over a decade of customer-facing experience across AI, search, and enterprise data systems. Experienced in architecting both the infrastructure and user-facing layers of AI systems, from retrieval pipelines to interactive applications. Proven ability to design, sell, and deliver complex AI-native infrastructure and retrieval-augmented applications that bridge product, engineering, and go-to-market teams. Deep expertise in vector databases, Kubernetes architectures, and LLM/MLOps pipelines, with a passion for transforming how enterprises deploy and scale agentic and generative AI systems — turning complex technology into scalable business outcomes.

WORK HISTORY

Customer Excellence Solutions Engineer — Weaviate

Aug 2024 – Oct 2025 · Remote / Amsterdam, ND

- Defined and launched the Bring Your Own Cloud (BYOC) offering, enabling Fortune 50 enterprises to deploy and manage Weaviate securely within their own cloud environments.
- Architected complex agentic and semantic search solutions, integrating multimodal retrieval and generative pipelines to demonstrate real-world enterprise use cases.
- Mentored new Solutions Engineers and contributed to core processes, including the product feedback loop and standardized POC framework, to accelerate enterprise delivery and adoption.
- Partnered across the organization to align roadmap priorities, create technical knowledge content, and strengthen Weaviate's enterprise positioning and presence in the market.

Senior Solutions Architect — Seldon

Jan 2022 – Jul 2024 · Remote / London, UK

- Built and led the North America Solutions Engineering function, working directly with enterprise and startup customers to deploy and scale Seldon's ML and LLM serving platforms.
- Designed and implemented enterprise-grade inference systems and deployment patterns for ML Endpoints, improving reliability, scalability, and time to production.
- Part of the team that designed Seldon's LLM product offering, defining early agentic AI patterns, customer experience flows, and integration models used in generative AI deployments — enabling enterprise teams to start exploring and adopting generative AI.
- Created repeatable ways to run pilots, onboard customers, and support post-sales projects, improving delivery speed and adoption across accounts.
- Contributed directly to MLServer, a core part of the Seldon product, fixing production issues for customers and rebuilding trust that led to a six-figure renewal.

TECHNICAL SKILLS

AI Infrastructure Design & Orchestration, Agentic Application Design (RAG frameworks, context management), Vector Databases & Semantic Search (Weaviate, multimodal retrieval), LLM & Generative AI Deployment, MLOps (Seldon Core, MLflow, Azure ML), Kubernetes (EKS / AKS), AWS / Azure Cloud Architecture, Helm / Terraform / Docker, CI/CD Automation, Monitoring & Observability, Python Development, FastAPI / Streamlit Application Development, Data Modeling & Pipelines, Authentication & Authorization, SQL / NoSQL / Graph Databases, Complex Enterprise Architecture, Product Feedback Loops & Experimentation

SOFT SKILLS

Customer Engagement & Success, Cross-Functional Leadership, Product & GTM Strategy Alignment, Technical Evangelism, Team Mentorship & Enablement, Solution Design & Implementation, Requirements Gathering & Scoping, Strategic Storytelling & Communication, Public Speaking & Conference Presentations, Self-Motivated & Outcome-Driven

SPEAKING

Purpose-Built Data Agents: This Is the Way

MLOps World / Gen AI Summit - October 2025

The Value of Data Agentic Tooling: A Work in Progress

- Worked across the organization to represent customer needs in engineering discussions, close feedback loops, and strengthen Seldon's presence and adoption in the MLOps community.

Senior Solutions Architect — Keeeb Inc. Jan 2021 – Dec 2021 · Boston, MA

- Led technical and go-to-market execution for Keeeb's federated search platform, overseeing pre-sales, customer enablement, and solution design for enterprise accounts.
- Closed the company's first U.S. enterprise deal in the manufacturing sector by designing and delivering a proof of concept that integrated internal data sources into the platform's discovery layer.
- Contributed to product connectors that improved data accessibility, demonstrated agile engineering execution, and unlocked key features required to close the deal.
- Structured and led a joint engagement with a systems integrator to migrate a customer's legacy data system to the Microsoft ecosystem, aligning technical scope, pricing, and delivery to ensure success.

Business Innovation Manager — Lucidworks Inc. Mar 2019 – Dec 2019 · San Francisco, CA

- Partnered with Product and Engineering leadership to design and prototype new capabilities for Lucidworks Fusion, drawing on experience as one of the platform's first enterprise customers and an early subject-matter expert.
- {'Helped define and deliver key product initiatives': 'a product merchandiser for digital commerce and a low-code/no-code framework for building search applications, contributing to both the initial and next-generation versions.'}
- Led Lucidworks' first beta program for new products, coordinating with early adopters to capture feedback and inform roadmap priorities.
- Collaborated with Customer Success and field teams to turn customer insights into product requirements, driving retention and expansion within enterprise accounts.

Solution Architect — Lucidworks Inc. Jun 2017 – Mar 2019 · San Francisco, CA

- Designed and implemented enterprise search and knowledge solutions using Lucidworks Fusion and Apache Solr for major technology and oil and gas customers.
- Influenced product direction by relaying customer needs to Product teams and validating new Fusion capabilities with key accounts.
- Integrated third-party technologies and vendor tools into the Lucidworks stack to expand solution capabilities and support complex customer environments.
- Created product demos for sales and marketing that contributed to Lucidworks' inclusion in the Gartner Magic Quadrant for Insight Engines.
- Mentored new Solution Architects following the Twigkit acquisition, training them in search theory, Fusion, and Solr technologies.
- Developed internal enablement programs and workshops to strengthen technical expertise across customer-facing teams and introduce new features in the Lucidworks product suite.

Sales Engineer — Lucidworks Inc. Feb 2016 – Jun 2017 · San Francisco, CA

- Supported sales teams by designing and presenting search-based solutions to enterprise clients across various.

MLOps World / Gen AI Summit - August 2025

WIP: Data and the Value of Data Agentic Tooling

Toronto Machine Learning Summit - June 2025

Digital Experience for the Next Generation Knowledge Worker
Activate 2019

Common Mistakes Made Deploying a Search Application with Lucidworks Fusion
Activate 2018

Ethics in Tech
UC Berkeley EECS Honors Program

From Model to Microservice > MLOps at Scale
Southern Data Science Conference

Various Apache Solr Meetups

EDUCATION

Bachelor of Science: Information Technology

Rochester Institute of Technology, Rochester, NY

Aug 2007 – Dec 2011

Concentrations in Finance and Communications.

- Led proof-of-concept projects for multi-million-dollar opportunities, directly contributing to 15 consecutive quarters of sales growth.
- Acted as a technical liaison between customers and engineering, ensuring field feedback informed early product releases.
- Enhanced Lucidworks' training program by leveraging AWS to improve reliability, scalability, and instructional content.

Co-Founder / Technical Lead — Mulsanne Dashboards

Apr 2014 – Jan 2016 · Houston, TX

- Designed and led development of Mulsanne Dashboards, an executive-level news and analytics platform that delivered personalized, business-critical updates to help leaders make better decisions throughout the day.
- Architected and implemented the platform's end-to-end infrastructure using Apache Solr, ExpressJS, and MongoDB, supporting real-time data ingestion, indexing, and search.
- Built an NLP/ML data pipeline that aggregated and classified news using RSS feeds, web scraping, NLP techniques, and bespoke machine-learning models developed with an external partner.
- Created internal monitoring and alerting dashboards to ensure system reliability and data accuracy.
- Defined product design, user workflows, and go-to-market strategy in collaboration with the founding team.

SharePoint Consultant — ImagineIT Feb 2012 – Apr 2014 · Houston, TX

- Designed and implemented SharePoint-based knowledge management systems for global oil and gas and construction enterprises, enabling collaboration across globally distributed teams.
- Built taxonomy-driven metadata tagging and classification models using the Expert Systems framework—an early NLP engine for document categorization and content discovery.
- Configured and deployed SharePoint 2010 and FAST Search on-premises, integrating metadata tagging with enterprise search and developing custom taxonomies for business unit workflows.
- Developed custom .NET, PowerShell, JavaScript, and HTML/CSS solutions to extend platform functionality and automate deployment tasks.
- Collaborated with Knowledge Management program teams on information architecture strategy, delivered end-user training, and advised on knowledge management best practices.