

Fabio Caironi

Founder, Product Manager & Data Scientist | Open to roles in SF, Austin, NYC | +1-628-502-3773 | fabio130497@gmail.com |
linkedin.com/in/fabiocaironi | github.com/fabcairo

Summary

Technical Product Leader and Founder with a background in Mathematics and Data Science. Specialized in building developer-first AI infrastructure and serverless platforms from inception to market adoption. Expert at applying analytical rigor to translate complex system constraints, such as GPU cold starts and latency, into clean product primitives and intuitive developer tools. Deeply analytical, leveraging production and user data to resolve system bottlenecks and identify high-impact growth opportunities. Proven ability to scale technical operations by automating workflows and fostering cross-functional collaboration to synthesize insights into prioritized, high-velocity roadmaps.

Work Experience

ByteNite

San Francisco, CA

Founder & CEO (Acting Head of Product)

Jul '21 - Present

- Founded and led a cloud infrastructure startup in San Francisco—raised \$1.2M, joined notable accelerators (incl. Techstars), and partnered with AI startups to design and ship a serverless compute platform for AI inference, web scraping and video workloads.
- Operated as a technical founder, driving the evolution of a serverless compute platform from a raw concept to a live product used for AI inference; owned the full stack strategy (API, SDK, GUI) and took features from idea to production in days to meet immediate user needs.
- AI Solutions Architecture: Directly supported developer customers with AI inference optimizations and LLM deployments—diagnosed GPU cold starts and instrumented model-quality checks, delivering code solutions that enabled 5+ production deployments within 3 months and slashed setup time by 3×.
- Developer Experience (DevEx): Designed developer primitives (API, SDK, CLI) that translated complex container orchestration into JSON-based templates and intuitive guides, reducing developer onboarding time by 60% and enabling rapid deployment of 4 AI application categories.
- Technical Market Research: Conducted 100+ deep-dive developer interviews and engaged at 20+ tech conferences to document infrastructure friction and emerging cloud-native trends. Maintained a comprehensive dossier on competitive landscape, including neo-cloud positioning and AI application segmentation.
- Hiring & Team Management: Screened 100+ candidates to hire top-tier talent. Established core collaboration practices, including Agile sprint routines and internal playbooks that 2x'd team productivity within 3 months from onboarding.
- Infrastructure Management: Managed GKE-based infrastructure and production CI/CD pipelines, automating change management to ensure seamless deployments across heterogeneous cloud environments.
- High-Throughput Orchestration: Architected a novel system for balancing compute supply/demand with resource-aware optimizations and efficient job scheduling algorithms.
- Product Analytics: Leveraged production data to unveil hidden system inefficiencies and user patterns through advanced SQL queries and Metabase dashboards, resulting in engineering fixes and a 50% increase in computing throughput.
- FinOps: Engineered comprehensive cost-forecast models that reduced cloud compute expenses by 30% through synthesis of multi-vendor pricing data and standardized provisioning rules.
- GTM Strategy: Launched multi-channel marketing campaigns (email, SEO, social) that generated 5,000+ monthly user engagements. Orchestrated end-to-end GTM initiatives by synchronizing engineering releases with sales outreach, driving adoption through technical deal structuring and repeatable playbooks.
- Growth Assets: Created all public-facing assets (CMS website, developer documentation, landing pages, and onboarding flows) that directly supported Product-Led Growth and content marketing.

Education & Awards

Techstars San Francisco Accelerator

Dec '24

Certificate | Entrepreneurship & Product Management

Selected among <1% of applicants for Techstars San Francisco, Fall 2024 Cohort; \$120K funding.

Key Highlights: Mentor-led product sprints; Investor pitch workshops; Customer discovery methodologies.

Università degli Studi di Milano

Jan '22

Master of Science (MSc) | Data Science and Economics (Coursework completed)

Relevant Coursework: Machine Learning; Econometrics; Coding for Data Science and Data Management.

Capstone Project: Decomposing and Clustering the EU Business Economy

GPA: 30/30

Università degli Studi di Milano

Oct '19

Bachelor of Science (BSc) | Mathematics

Relevant Coursework: Linear Algebra; Probability Theory; Algorithms.

Honors: Scholarship Recipient

GPA: 108/110

Skills

Product Strategy: End-to-End Product Lifecycle Management · API & SDK Design · Developer Experience (DevEx) & Tooling · Data-Driven Product Analysis · Algorithm Design · Product-Market Fit & Use Case Analysis · Technical Writing & Documentation · Template Design · Technical Trade-Off Prioritization · Roadmap Development

AI & Cloud Infrastructure: Google Cloud Platform (Compute Engine, GKE, Cloud SQL Storage) · AWS (Solutions Architect Certified) · Kubernetes · Containerization (Docker) · Open-Source LLMs · Distributed Systems · GPU Taxonomy · Cloud Economics · AI Inference Optimization · Serverless & Container-Native Compute · Cloudflare (R2, CDN, DNS)

GTM & Business Development: Product-Led Growth (PLG) & Sales-Led Growth (SLG) Strategy · Customer Discovery & PoCs · Multi-Channel Marketing · Market Analysis & Positioning · Technical Deal Structuring · Post-Sales Support & Troubleshooting · Customer Relationship Management (CRM) · Venture Fundraising (\$1M+)

Technical Leadership: Cross-Functional Collaboration · Stakeholder Management · Translation for Non-Technical Stakeholders · Partner Management

Programming & Tools: Python · SQL · n8n (Workflow Automation) · Javascript (Basic) · C/C++ (Basic) · Metabase (Data Modeling & Dashboards) · Figma · Jupyter Notebooks · Webflow (CMS) · Adobe Creative Suite (Illustrator, Photoshop, Premiere Pro, After Effects)

Publications & Patents

Method and system for high-throughput distributed computing of computational jobs | US & EU Patent (Granted)

Nov '22

Patent Number: US12061607B2

Developed a system for dynamic partitioning and orchestration of jobs across heterogeneous grids, optimizing for capacity and availability.

Additional Information

Interests | 15 years as a professional classical pianist; passionate about the intersection of logic and creative expression.

Languages | English (Fluent); Italian (Native)

US Work Authorization | Possesses E-2 visa; O-1 eligible (requires sponsorship for new employer); available immediately.