

Juan Nicolas Bautista Hobin

Software engineer: data infrastructure and AI agent systems

nico@jnbh.dev
jnbh.dev
github.com/nhobin219
Scottsdale, AZ

EXPERIENCE

Independent Engineer, open-source data infrastructure and AI agent systems

2026.04 – present

Available for independent consulting

Python · Claude Agent SDK · TypeSafe SDK · Iceberg · Arrow

- **agent-office** open-sourcing 2026.11: a runtime for agent organizations. A gateway, planner, and router run long-lived agents over one streamcast log, so the whole system is replayable and queryable, and each agent holds its own narrowly scoped credentials federated from OpenBao.
- **streamcast** github.com/nhobin219/streamcast: a JSON WebSocket pubsub framework. Messages are schema-validated and appended to a litelink log before subscribers see them, so streams double as queryable tables; at-least-once by default, with opt-in exactly-once and offset-based resume.
- **litelink** github.com/nhobin219/litelink: an embedded Iceberg storage engine for append-only data. Appends are durable on commit and land as sorted Parquet in an Iceberg table on local disk or S3, with no server, daemon, or catalog service.
- **penca** github.com/penca-io/penca: an experimental version-controlled lakebase in Rust (Flight SQL, DataFusion) with zero-copy branching, time travel, and row-level versioning on one columnar copy. Its ideas carried into litelink and streamcast.

Citadel Quantitative Developer, Equity Quantitative Research

2024.01 – 2026.03

New York, NY

Data Platform

- Designed and built from scratch, largely single-handedly, a high-availability, zero-downtime Python data service that serves mission-critical risk data at about 500 requests/s under a partial on-prem requirement.
- Optimized production trading and research analytic query workflows across cloud-native and on-prem technologies.
- Served as a top interviewer and intern mentor for EQR, and onboarded new engineers to Data Platform.

Alpha Strategies

- Member of the central liquidity team providing execution services to all of Citadel's fundamental equities businesses, working with quantitative researchers to implement and test ideas for optimizing the system's trading behavior.

Calligram, Inc. Co-founder

2022.06 – 2023.06

New York, NY · stealth

Go · Python · AWS · Kubernetes

- Co-founded and co-led a generative-AI startup (product, engineering, fundraising). Earned Y Combinator interviews, pre-seed investment from SBXi/77.xyz, and design-partnership negotiations with Crunchbase.
- Led MVP development on AWS (EKS, Docker, Pulumi IaC); owned deployments and cloud costs.
- Built natural-language APIs and RAG frameworks on GPT-4, LanceDB, and Qdrant (Go/Gin, FastAPI, Slack Bolt), with semantic query caching to keep inference costs down.

Appian Corporation Software Engineer

2020.03 – 2022.10

Remote

q/kdb+ · Go · Java · Python

- Built features for Appian's patented distributed data store (ADS) in q/kdb+, Go, and Java to improve its scalability and flexibility.
- Improved query performance with read-optimized snapshots and background clean-up processes that cut disk usage.
- Extended the ADS query language (GraphQL-inspired) with computational functions, and engineered high-throughput ingestion paths for bulk writes.
- Built a Python performance-testing framework in GitLab CI to validate query speed and reliability; onboarded and mentored engineers and resolved production issues.

- Engineered a kdb+ real-time stack that consolidated 10+ data sources into 3 normalized real-time streams, powering the back end of Virtu Financial's Open Technology Platform.
- Designed and implemented the gateway architecture: the core query routing and execution logic.
- Optimized ETL pipelines from SQL Server into kdb+, and built the integration layers: JSON REST services, AMPS streaming interfaces, and legacy C++/Java interop.
- Led client onboarding and database consolidation for a major fintech client (ITG), and owned production operations for global deployments: releases, incident response, and ITRS real-time monitoring.

SKILLS

LLM & agent tooling	Claude Agent SDK, OpenAI SDK, TypeSafe SDK, LiteLLM; agent orchestration, agentic software factories, RAG, vector search, semantic caching
Machine learning	scikit-learn, statsmodels; PyTorch (working knowledge)
Formats & engines	Apache Arrow, Arrow Flight SQL, DataFusion, Apache Iceberg, Parquet, Lance
Databases	kdb+/q, DuckDB, PostgreSQL, SQLite, LanceDB, Qdrant; time-series, embedded, and vector stores
Systems	Database design, distributed and real-time systems, streaming (Kafka, AMPS), gRPC, microservices
Languages	Python, q, Go, SQL, Julia, Groovy; Rust, Java, and R (working knowledge)
Infrastructure	AWS, GCP, Docker, Kubernetes, Argo, Terraform, Pulumi (IaC), OpenTelemetry, GitLab CI, UNIX, Git

EDUCATION

Massachusetts Institute of Technology *B.Sc., Applied Mathematics; concentration in Economics*

2017.06

RESEARCH & PUBLICATIONS

MIT Innovation Diplomat *Guadalajara, Mexico*

2016.05 - 2016.09

Spent a summer in Guadalajara researching the city's innovation ecosystem. Interviewed entrepreneurs, government officials, and venture capitalists, and compiled a report on the findings.

Y. Yoshimura, S. Sobolevsky, **J. N. Bautista Hobin**, C. Ratti, and J. Blat. Urban association rules: uncovering linked trips for shopping behavior. *Environment and Planning B: Planning and Design*, 2016. doi:10.1177/0265813516676487

ADDITIONAL

Languages	English (native), Spanish (fluent); boxing, hiking, running, travel, volunteering, jazz
Work authorization	US citizen. Also authorized in the EU, UK, and Mexico.