

Jeff LeFevre

Systems Software Engineer | Open Source Lead | Data Management

Software engineer and researcher with 15 years experience in databases and storage systems.

[Homepage](#) | [Google Scholar](#) | [LinkedIn](#) | [Email](#) | [Phone](#) | Tampa, FL 33611

EXPERIENCE

Malleo.ai (Ai Safety Startup) — Scalability Advisor | 2026 –

Architect the parallelization and deployment of Adversarial Evaluation Pipelines for Lapis, Malleo's automated cognitive red-teaming engine used for agentic LLM security.

Outlier.ai — Expert Reviewer/Top Contributor | 2025 –

Engineer complex prompts and technical code reviews for high-quality frontier model training data; develop and apply rigorous evaluation rubrics to audit model outputs for technical accuracy, safety, and alignment; annotate and rank multi-turn responses to provide ground-truth data for reward model optimization.

Personal Sabbatical — Relocation to Florida | 2023 – 2025

Family leave; self-study of vector indexing and LLM adversarial induction

Keebo.ai — Senior Software Engineer | 2023

Developed automated view selection and maintenance algorithms for Snowflake databases.

UC Santa Cruz — Open Source Fellow & Adjunct Professor | 2017 – 2023

Created and led the development of Skyhook, extending Ceph with programmable semantics for relational data management. **Result:** Project merged into the upstream **Apache Arrow** ecosystem (Arrow 7.0.0). Implemented zero-copy data sharing and "Arrow-native" storage, allowing filtering and projections to be offloaded to storage nodes. Mentored Google Summer of Code (GSoC) and IRIS-HEP fellows for five cycles as well as BS/MS/PhD students.

HP Vertica — Senior Software Engineer | 2014–2016

Architected and developed high-performance, fault-tolerant parallel connector for Vertica and Apache Spark (Scala) with pushdown computation, achieving "exactly-once" semantics without external coordinators (e.g., Zookeeper); engineered distributed training data pipelines for Vertica and DistributedR; optimized data movement for enterprise-scale ML workloads, ensuring low-latency integration between storage and compute.

Google / NEC Labs / Teradata — Software Engineer Intern | 2007 – 2013

Developed prototypes for storage platforms and automated physical design tuning for large-scale analytics.

SELECTED PUBLICATIONS

"Conversational Authority Capture as a Novel Attack Surface in Large Language Models" | (preprint) 2026.

"Mycelium: A Transformation-Embedded LSM-Tree" | (Under submission) 2026.

"Skyhook: Towards an Arrow-Native Storage System" | [CCGrid 2022](#).

"Building the Enterprise Fabric for Big Data with Vertica and Spark" | [SIGMOD 2016](#).

"MISO: Souping Up Big Data Query Processing with a Multistore System" | [SIGMOD 2016](#).

"Large-scale Predictive Analytics in Vertica: Fast Data Transfer, Distributed Model Creation, and In-database Prediction" | [SIGMOD 2015](#).

MAJOR PROJECTS

SkyhookDM: Lead Architect and core developer. Enabled server-side computation on programmable distributed object storage. *Merged into Apache Arrow 2022.* [Github](#).

Vertica-Spark Connector: Lead Architect and core developer. Enterprise-grade parallel connector for large-scale enterprise data. *Released 2016.* [Github](#).

TECHNICAL SKILLS

Languages: Python, C++, C, Java, SQL.

Distributed Systems & Storage: Ceph, RocksDB, Apache Arrow, Data Pipelines (ETL, Training, Evaluation), Parquet, Spark, Object Storage, S3.

AI & Vector Search: HNSW/Vector DBs, LLM Adversarial Testing, MLOps.

Cloud & Infrastructure: AWS, Docker, Kubernetes, Git, CI/CD.

PATENTS

[US11106672B2](#) (2021): Queries based on ranges of hash values.

[US10909119B2](#) (2021): Accessing electronic databases.

[US10503718B2](#) (2019): Parallel transfers of electronic data.

[US9569491B2](#) (2017): Multistore online tuning system.

[US9477708B2](#) (2016): System for multistore execution environments with storage constraints.

[US9183253B2](#) (2015): System for evolutionary analytics.

EDUCATION

Ph.D. in Computer Science & Engineering | UC Santa Cruz

M.S. in Computer Science & Engineering | UC San Diego

B.S. in Computer Science & Engineering | USF Tampa