

Matthew Osgood

Data Engineer

Greater Hartford, CT • (860) 604-1322 • Matthew_osgood@outlook.com • www.mosgood.me

SUMMARY

Data Engineer with 3+ years building production ETL/ELT pipelines and large-scale data-processing systems in financial services. Specialized in PySpark, Microsoft Fabric, and Delta Lake architectures across bronze/silver/gold lakehouse layers, with deep experience cleansing, deduplicating, and transforming raw source data into reportable, well-modeled (SCD Type 2) datasets. Strong in performance optimization — partitioning, clustering, and indexing — and in partnering directly with business stakeholders to translate their needs into reliable data products. Full-stack range from SQL Server internals to Next.js/TypeScript applications.

EXPERIENCE

Data Engineer — Planet Home Lending

July 2025 – Present

- Build and maintain production PySpark pipelines on Microsoft Fabric that ingest, cleanse, deduplicate, and transform mortgage and loan-servicing data across bronze, silver, and gold lakehouse layers.
- Model raw source data into Slowly Changing Dimension (Type 2) datasets for accurate point-in-time, reportable history; designed reusable SCD2 notebooks for Encompass loan data with custom column sanitization for numeric and special-character source fields.
- Optimize pipeline and query performance through partitioning, clustering, and indexing — eliminated 59 Spark cold starts in a high-cost Bronze refresh, materially reducing CU consumption on an F64 capacity.
- Build Gold-layer warehouse ETL using COPY INTO and MERGE patterns to incrementally load curated, analytics-ready tables for downstream reporting.
- Author SQL Server stored procedures (e.g., procLFNLoanIndicator) for dual Mercury and SFTP fixed-width file output, plus a PHL Retail transaction feed pipeline.
- Partner directly with business and reporting stakeholders to translate requirements into engineered datasets; host open-source LLMs (Llama 3) with RAG, embeddings, and vector search to power internal automation.

Data Engineer / Technical Business Analyst — Planet Home Lending

July 2023 – July 2025

- Acted as the technical bridge between business teams and data systems — gathered reporting and operational requirements, then engineered Python and SQL ETL pipelines and REST APIs to deliver exactly the data teams needed.
- Maintained continuous communication with business stakeholders to clarify data definitions, validate outputs, and turn ambiguous requests into well-scoped, deliverable specifications.
- Designed scalable backend systems to process large loan transfers reliably with strong data-integrity guarantees.
- Automated compliance checks, letter generation, and onboarding tasks using JAMS Scheduler and integrated APIs.
- Drove automation initiatives that reduced operational costs by approximately \$150K per month.

PROJECTS

Sell Local — Full-Stack Engineering Project *selllocal.app*

- Architected and built a production multi-tenant web platform end-to-end: TypeScript/Next.js frontend, Supabase/PostgreSQL backend, Stripe Connect payments, and AWS SES transactional email, deployed on Vercel.
- Implemented storefronts, pickup ordering, a customer notifications hub, bio-link pages, and digital product sales, with secure authentication and role-based access.
- Provide direct technical support to platform users — diagnosing payment-onboarding, DNS/domain, and integration issues and translating non-technical reports into resolved fixes.

Molloy Foundation Tournament Platform — Full-Stack Developer

- Developed a Next.js + Supabase + Stripe tournament registration platform with real-time updates and secure authentication for an annual charity basketball event.

SKILLS

- **Languages:** Python, SQL, T-SQL, TypeScript, JavaScript
- **Data Engineering:** PySpark, Delta Lake, Microsoft Fabric, ETL/ELT, Lakehouse (Bronze/Silver/Gold), SCD Type 2, MERGE / COPY INTO, partitioning & clustering, query optimization, SQL Server, PostgreSQL
- **Backend & APIs:** REST APIs, Django REST, Next.js API routes, Stripe integration, JAMS Scheduler
- **Cloud & DevOps:** AWS (SES), Supabase, Vercel, GitHub, CI/CD, Linux / WSL, Docker
- **AI / LLMs:** RAG, embeddings, vector search, prompt engineering, fine-tuning, Llama 3
- **Other:** Systems design, performance optimization, workflow automation, stakeholder communication

EDUCATION

B.A. Economics & Data Analytics — University of Connecticut

August 2018 – May 2022

Coursework: Machine Learning, Data Visualization, Statistical Modeling (Python), Business Information Systems (RDBMS).

CERTIFICATES & CONTINUING EDUCATION

- DeepLearning.AI — RAG, Prompt Engineering, Fine-Tuning LLMs, Hugging Face, Open-Source LLMs, PyTorch.
- Coursera — Python Automation, API Development.