

Guru Balamurugan

📍 New York, NY ✉ gurub.3174@gmail.com in guru17 ❄ Snowflake Adv. Data Engineer

Experience

AI Engineer • Mitsubishi UFG Capital Analytics • NYC *Jul 2025 – Present*

- Architected Snowflake Medallion pipeline + REST API and secure data shares for 3 GP clients (\$500B+ AUA); enabled shift from quarterly to daily fund-performance reporting.
- Shipped AI-powered client analytics portal with semantic-layer-guardrailed agent for self-service dashboard builder; replaced 2-week analyst reporting cycle for 50+ users across clients, enabling 10% pricing-tier upgrade.
- Engineered dbt MDM hub-and-spoke models resolving 10k+ funds with 50+ SME-defined automated tests and anomaly detection; per-client PII isolation prevents cross-client data leakage and LLM hallucinations.
- Replaced brittle Python ingestion with Openflow event-driven CDC/SCD connectors with schema enforcement; cut new-client data onboarding by 30%, accelerating time-to-revenue on new GP mandates.
- Built AWS PrivateLink gateways isolating client data from public internet for BI tools; TPRM and Legal cleared.
- Automated production DB deployment via Azure DevOps CI/CD agent with Azure AD inherited RBAC; compressed release cycle from 4 weeks to 1 week, accelerating turnaround on client-requested data-model changes.

Data Science & ML Intern • Vertice Analytics • Durham, NC *May 2024 – Apr 2025*

- Shipped merchant-mapping app (JavaScript+Docker+AWS) for customer success team with LLM extraction + RAG + Human Review over 1T+ PostgreSQL; replaced brittle ILIKE SQL, 800% query speedup, 83% accuracy.
- Built 20 self-proposed merchant filters via mapping app + automated filter-request workflow to 20+ credit unions; saved both teams 2 weeks each, 3x transaction coverage.
- Developed client-requested (RF + LightGBM + LinReg ensemble) model on AWS EC2 + FastAPI classifying members into 5 tiers; 76% accuracy, 4k+ members across 5 credit unions targeted for offers.

Data Analyst Intern • Synchrony Financial • Stamford, CT *May 2022 – Aug 2022*

- Engineered Python + SQL transformations over 500M+ consumer transactions into Tableau dashboards with Director; surfaced credit-card approval-odds insights for Newegg marketing on product-bundle strategy.
- Delivered data insights via demo to 5 Newegg marketing executives, shifting product-bundle decisions from intuition to data-driven analytics; curated narrative for senior business stakeholders.

Projects

AI Document Extractor

- Shipped self-initiated POC parsing unstructured document (e.g. capital calls) into columnar database; chosen over \$2M/yr vendor proposal at >90% accuracy.
- Developed end-to-end extraction pipeline (parse → classify → LLM extract → RAG → human review) with Pydantic structured outputs, dbt lineage, and retry logic; writes 11 metrics across 3 PE document types to Snowflake.
- Engineered 100+ business-rule validation layer with MCP (CapIQ + Preqin API + internal DB) for data gaps and reviewer escalation; eval harness against golden records caught extraction inconsistencies; TPRM-cleared.
- Pitched MVP to President, CTO, and Directors with cost-and-accuracy analysis vs third-party vendors; picked over \$3M internal initiative, led adoption roadmap, and built CI/CD deployment with RBAC integration.

SEC ADV Market Research Application

- Built Python + dbt pipeline ingesting SEC ADV filings into Snowflake with LLM-normalized administrator names; president-requested universe tracker covering 50+ private-markets fund administrators across \$128T AUM.
- Delivered Power BI dashboards with Natural language AI agent + MCP-connected competitor lookups for marketing team; replaced manual 2-3 week research cycle with daily refresh, 10 active users.

Technologies

Python, SQL, LangGraph, A2A agents, MCP servers, RAG, Claude API, AWS Bedrock, Pydantic, Docling, Scikit-Learn, Snowflake (Cortex Analyst, DMF, Openflow), dbt, AWS, PrivateLink, Azure DevOps, Docker, Power BI, Streamlit, Git

Education

BS in Computer Science • University of North Carolina at Chapel Hill

May 2025