

Aditya Varma G

(313)752-6583 • adityaganapathiraju@gmail.com • [Portfolio](#) • [LinkedIn](#) • [Ready to Relocate](#)

SUMMARY: Data Engineer with 6+ years of experience building scalable data pipelines and cloud-native analytics platforms across AWS & GCP. Skilled in Python, SQL, Spark, and managing structured/unstructured data to enable reliable, ML-ready data delivery.

EXPERIENCE

Data Engineer | Aetna

Sep 2025 – Present

- Leading migration from Teradata to GCP (BigQuery, Cloud Data Fusion, GCS) to modernize enterprise data analytics ecosystem.
- Architected CI/CD-driven ingestion pipelines using Apache Airflow and Dataflow, ensuring automated and version-controlled data delivery for regulatory and business reporting.
- Designed semantic layers and curated data marts aligned with analytics consumption patterns, improving querying performance by ~30%.
- Implementing data quality monitoring using Great Expectations + Cloud Logging, reducing data inconsistencies and validation failures by 40%.
- Partnering with AIML engineering teams to standardize feature-ready data formats and ensure governance for sensitive healthcare data

AI Data Engineer | Proxzar.ai

Feb 2025 – Sep 2025

- Built and deployed a spaCy transformer-based NER system for unstructured financial text; led a team of 4 engineers to achieve 92%+ entity-level F1, generating enriched features for risk scoring and customer behavior models.
- Developed end-to-end retraining pipelines using MLflow and automated metadata tracking, improving ML iteration speed and reproducibility by 40% while ensuring clean data lineage for AIML datasets.
- Automated ingestion workflows using AWS Glue, Lambda, and S3, cutting manual runbooks by 60% and enabling continuous integration of both structured (transactional) and unstructured (notes, call transcripts) sources.

Data Intern | General Motors

May 2024 – Aug 2024

- Developed scalable ETL workflows using Azure Databricks + Spark, reducing campaign execution data latency by 40%.
- Automated KPI dashboards to track A/B experimentation performance, improving marketing spend ROI by 25%.
- Enhanced customer segmentation strategies through Spark ML pipelines and predictive modeling, aiding targeted promotional campaigns.

Technology Analyst | Infosys

May 2022 – Aug 2023

- Engineered SQL Server-based ETL pipelines to automate real-time ingestion and anomaly detection across 15M+ financial transactions weekly, improving fraud monitoring responsiveness by 30%.
- Enhanced fraud detection ML workflows (Random Forest, XGBoost) by optimizing feature retrieval and query performance, reducing false-positive alerts by 21%.
- Modernized batch workloads by integrating AWS S3 + Lambda for scalable data staging and event-driven processing, reducing system overhead by 25%.

System Engineer | Tata Consultancy Services

Nov 2018 – May 2022

- Modernized legacy trading risk applications by migrating core services to AWS EC2 + RDS, increasing uptime and disaster-recovery compliance by 40%.
- Containerized 20+ back-end components with Docker and streamlined deployments through CI automation, shortening release cycles by 75%.
- Built REST service integrations and data validation logic to support regulated financial workflows, improving operational accuracy and audit readiness.

PROJECTS

V2X Collision prediction with ML | [LINK](#)

- Built an end-to-end AI/ML pipeline for predicting collisions involving vulnerable road users using V2X(Vehicle to everything) communication data from simulations with Veins (OMNeT++/SUMO).
- Designed modular experiment flows with LSTM and Transformer models, integrating Weights & Biases (W&B) for experiment tracking, checkpointing, and reproducible evaluation.

[A lot more Projects in Portfolio...](#)

EDUCATION

- Master of Science in Data Science and Business Analytics - **Wayne State University** (GPA: 3.93/4.0) *Graduated Dec 2024*
- Masters in technology in Software Engineering - **Birla Institute of Technology and Science** *Graduated Dec 2023*
- Bachelor of Engineering in Mechanical Engineering - **Sri Sairam Engineering College** *Graduated May 2018*

TECHNICAL SKILLS

- **Languages:** Python, SQL
- **Data Engineering:** Spark, PySpark, Databricks, ETL/ELT, API Integrations, Streaming Pipelines
- **Cloud & Platforms:** GCP (BigQuery, Dataflow, GCS), AWS (Glue, Lambda, S3), Airflow
- **Databases:** PostgreSQL, Oracle, Snowflake (Basic)
- **Data Quality & Governance:** Great Expectations, Data Catalog
- **DevOps:** Git, Docker, CI/CD