

MANIKANTA POCHAMALLA

manikantapochamalla85@gmail.com | +1 (984) 329-1618 | United States
LinkedIn: [linkedin.com/in/manikanta-pochamalla](https://www.linkedin.com/in/manikanta-pochamalla) | GitHub: github.com/Mani88888 | Portfolio:
mani88888.github.io/manikanta-portfolio/

PROFESSIONAL SUMMARY

Results-driven **Data Engineer** with 4+ years of experience building, optimizing, and scaling data ecosystems in cloud environments. Proficient in **Python, SQL, and AWS** with extensive hands-on experience in **Spark, Databricks, Airflow, and Kafka**. Specializes in developing **ETL/ELT pipelines, data lake architectures, and ML-ready datasets** that enable real-time analytics and predictive modeling. Strong background in **data reliability, performance tuning, and automation** using Terraform, CI/CD, and cloud orchestration frameworks.

EDUCATION

M.S. in Computer Science – Campbellsville University | GPA: 3.7 | Graduated Aug 2024

PROFESSIONAL EXPERIENCE

Gatik – Data Engineer | Feb 2023 – Present

- Architected and deployed scalable data ingestion pipelines for autonomous vehicle sensor data using Python, AWS S3, and Kafka—processing 10TB+ weekly and enabling real-time analytics for vehicle performance and safety insights.
- Optimized SQL queries and Snowflake models, reducing query execution time by 40% and lowering compute costs by 25% through partitioning and indexing strategies.
- Implemented Spark jobs on AWS EMR and Databricks for large-scale sensor data transformations supporting ML model training and data science workflows.
- Automated ETL orchestration using Airflow and Step Functions, ensuring 99.9% reliability with advanced retry, alerting, and dependency management.
- Developed Python-based data validation framework, improving data consistency by 30% and reducing manual QA by 50%.
- Containerized ETL components with Docker and deployed via AWS ECS and EKS using Terraform, improving deployment speed and environment reproducibility by 60%.
- Collaborated cross-functionally with ML, analytics, and DevOps teams to standardize data access and accelerate model development lifecycles.

Deliverr – Data Engineer | Oct 2021 – Aug 2022

- Designed and managed ETL workflows in Python, Airflow, and Databricks, integrating multi-source logistics data into AWS and BigQuery environments.
- Engineered advanced SQL transformations and materialized views to enhance reporting capabilities, cutting dashboard latency by 35%.
- Processed 5TB+ daily data using Spark and optimized job performance, reducing pipeline runtime by 40%.
- Developed Terraform-managed AWS data infrastructure, improving scalability, cost-efficiency, and security compliance.
- Partnered with data science teams to deliver ML-ready datasets for route optimization and predictive demand modeling, reducing delivery delays by 15%.

PROJECTS

- Real-Time Data Streaming with Kafka and AWS Kinesis – Designed an end-to-end pipeline processing 50K+ events/min, delivering real-time analytics for autonomous vehicle telemetry.
- ETL Automation Framework – Created a Python- and SQL-based validation tool reducing manual QA time by 60% and improving pipeline accuracy.
- Data Quality Dashboard – Built with Python, SQL, and Looker to track 500+ quality KPIs, improving anomaly detection response time by 35%.

TECHNICAL SKILLS

Languages: Python, SQL, PySpark | **Cloud Platforms:** AWS (S3, EMR, Glue, Lambda, Step Functions, EventBridge, ECS, EKS), GCP (BigQuery) | **Data Engineering:** ETL/ELT Pipelines, Data Lakes, Warehousing (Snowflake, BigQuery) | **Big Data:** Spark,

Databricks, Kafka | **Orchestration:** Airflow, DBT | **Infrastructure:** Terraform, Docker, CI/CD | **Concepts:** Data Modeling, DataOps, Data Governance, ML Data Preparation

CERTIFICATIONS

- AWS Certified Data Engineer – Associate
- Databricks Certified Data Engineer Associate (in progress)
- SQL (Advanced) Certification – HackerRank (in progress)