DHARAMA TEJA SAMUDRALA

New York, NY 14221

+18573983456 | dharmatejas102@gmail.com | linkedin | github

SUMMARY

Data Engineer with 3+ years of experience designing and implementing data pipelines that support **product features, personalization, and analytics**. Skilled in **Python, SQL, Spark, modern data lake and warehouse platforms (Azure, Snowflake)** with proven ability to build **real-time and batch data processing systems**. Expertise in data modeling, quality assurance, and **cross-functional collaboration with product managers and engineering teams**.

TECHNICAL SKILLS

Languages & Analytics:	Python, SQL, Pandas, NumPy, scikit-learn
Data Engineering & Cloud:	Azure Data Lake, Snowflake, AWS (S3, Lambda, EMR, Glue), Kafka, Airflow, Spark, Docker, Git/GitLab CI/CD
Databases:	Azure SQL Database, MySQL, PostgreSQL, MongoDB
Frameworks & Methodologies:	Data modeling, ETL/ELT design, Agile development, A/B testing, Data privacy (GDPR, CCPA)
Visualization & BI:	Power BI, Tableau, Matplotlib

EXPERIENCE

Citi   NewYork,USA
Data Engineer   March 2025 – Present

- Design and implement **data pipelines supporting product features, personalization, and analytics** using modern frameworks on Azure.
- Build and maintain **real-time and batch data processing systems** leveraging Kafka, Spark, and Azure services to process over 1M events per day.
- Collaborate with **product managers, project engineers, and insights engineers to understand data needs** and translate them into technical solutions.
- Ensure **data quality, reliability, and performance** across all product-integrated data flows through monitoring and optimization.
- Contribute to design of **scalable data models** supporting personalized user experiences and A/B testing.

MetLife   India
AWS Data Engineer   Dec 2021 – Jul 2023

- Developed **PySpark ETL pipelines** in Azure to transform and enrich **1–5 GB** of transactional and log data daily from multiple sources.
- Designed and orchestrated **Airflow DAGs** for ETL workflows, automating **5–10 recurring jobs** with dependencies, retries, and alerting.
- Integrated data from APIs, CSVs, and SQL databases using **PySpark**; performed cleaning, validation, and deduplication across **100K+ records** per batch.
- Implemented **data quality checks and audit logs** in pipelines, flagging anomalies and ensuring **99% consistency** between raw and processed data.
- Managed **Snowflake and Azure Synapse** data warehouses; optimized query plans and compression to reduce compute costs by **20%**.

PROJECTS

Microservices Communication Framework   Link to Project

- Architected 3 gRPC microservices (User, Order, Payment) with Envoy proxies; implemented **mTLS, circuit breakers, and retry policies for resilient communication**.
- Built **observability pipeline** using Prometheus, Grafana, and Jaeger; automated deployment with Docker and Helm; benchmarked latency across 100K+ requests.

- Reduced P99 latency by 24% via **load balancing**; canary routing cut deployment risk by 40%; quantified service mesh overhead at 8% for adoption decisions.

EDUCATION

Rochester Institute Of Technology   Rochester, NY, USA
Master of Science in Computer Science   Aug 2023 – Aug 2025