Phanindra Reddy Myakala

San Francisco, CA • myakala.p@northeastern.edu • +1 (617)238-4092 • LinkedIn • Open to Relocate

"Data Engineer with 5+ years of experience in architecting scalable Big Data platforms, ETL/ELT pipelines, and ML solutions across GCP, Azure, and AWS. Expert in streaming data, distributed computing, and data warehousing with a proven track record of driving efficiency and business impact."

SKILLS

Programming Languages: Python (Pandas, NumPy, TensorFlow, PyTorch, Scikit-learn), SQL, R, Scala (Beginner)

Databases & Storage: Snowflake, PostgreSQL, BigQuery, Azure Synapse, MongoDB, Neo4j, ADLS, S3

ETL & Visualization: Airflow, DBT, Databricks, Azure Data Factory, Cloud Dataflow, Power BI, Tableau, Looker

Big Data Technologies: Spark, Kafka, YARN, HDFS, Hive, Presto, Flink

Cloud Platforms: GCP (Vertex AI, BigQuery, Dataflow), Azure (Data Factory, Synapse), AWS (S3, Lambda)

Other Tools: Git, Docker, Jenkins, MLflow, FastAPI, ELK Stack, Apache Atlas

WORK EXPERIENCE

Senior Data Engineer | Dell Technologies, Franklin, MA

Feb 2024 – Jul 2024

- Architected end-to-end data pipelines using Airflow and Spark, automating supplier quote processing and reducing manual effort by 88%, saving 200+ hours monthly and establishing near real-time data availability
- Orchestrated a comprehensive data lake migration to Databricks Delta, implementing a medallion architecture that decreased query latency by 65% and improved data reliability across procurement systems
- Designed and implemented a robust ML pipeline infrastructure with Docker, MLflow, and Jenkins, enabling continuous deployment with 99.9% uptime and comprehensive model observability
- Created scalable BI dashboards in Power BI using optimized SQL/Spark queries, driving improved procurement decisions across $100M+ in supplier contracts and contributing to a 12% cost reduction

Data Engineer | ISFT IT Services, Hyderabad, IN

Jul 2021 – Jul 2022

- Designed and implemented resilient batch and real-time ETL pipelines using PySpark, Kafka, and Flink, processing 50TB+ of financial data daily with 99.8% reliability
- Developed metadata-driven data ingestion architecture with Great Expectations and automated schema enforcement, improving data quality scores by 95% across enterprise-scale financial systems
- Engineered a comprehensive data monitoring system using ELK Stack and Apache Atlas, reducing pipeline failures by 40% and enabling complete data lineage tracking for compliance requirements
- Built and optimized FastAPI-based data services layer, reducing API response times by 70% while serving clean datasets to downstream ML models and business intelligence platforms
- Led migration of legacy ETL processes to cloud-native solutions, achieving 60% reduction in compute costs while enhancing performance

Data Scientist | Northeastern University Research Lab, Boston, MA

Jul 2024 – Present

- Architected a unified data ingestion framework for research datasets using Airflow and Spark, decreasing data preparation time by 75% and enabling faster research iterations
- Designed industry-relevant deep learning modules using TensorFlow and PyTorch, building frameworks adopted by 100+ students and researchers
- Implemented a highly scalable real-time data processing system using Kafka and Flink for streaming analytics, reducing data latency by 85%
- Created production-grade ML model deployment pipelines using Docker, Kubernetes, and MLflow, achieving 99.5% model serving reliability

Teaching Assistant | Northeastern University, Boston, MA

Sep 2023 – Jan 2024

- Developed comprehensive data engineering tutorials using Databricks, Spark, and SQL, driving a 25% improvement in student project quality and technical proficiency
- Mentored 150+ students on machine learning and data engineering topics, enhancing algorithmic thinking and coding fluency

PROJECTS

Real-time Health Analytics Platform

TensorFlow, GCP, Airflow, Kafka, Databricks

- Engineered a real-time data pipeline with Kafka and Spark Streaming that processed 10,000+ events per second, enabling predictive health analytics with 15% improved early detection rates
- Orchestrated data ingestion, preprocessing, and model deployment using Airflow, Docker, and Vertex AI, reducing development-to-production time by 40%
- Implemented a comprehensive data quality framework with Great Expectations, reducing data errors by 95% and ensuring high-quality insights

Financial Fraud Detection System

XGBoost, Snowflake, MLflow, Flink, Tableau

- Built a high-performance real-time fraud detection system using Flink and XGBoost, achieving 0.90 AUPRC and reducing false positives by 30%
- Designed and implemented a medallion data architecture in Snowflake, enabling 10x faster analytics queries and comprehensive fraud pattern detection
- Operationalized model scoring pipelines and dashboards in Tableau, providing real-time fraud alerting capabilities that prevented an estimated $200K in potential losses

EDUCATION

MS in Data Science — Northeastern University, Boston, MA

BE in Electronics and Communication Engineering — NIT Hamirpur, India

CERTIFICATIONS

- Google Cloud Professional Data Engineer (In Progress)
- Microsoft Certified: Azure Data Engineer Associate