VINAY RAM GAZULA

Newark, NJ   908-552-1879   gazulavinayram@gmail.com   vinaygazula.dev   linkedin.com/in/vinayramg

PROFESSIONAL SUMMARY

Data Engineer with 3 years of experience in building and maintaining scalable data pipelines to enhance data quality and drive business growth. Proficient in ETL/ELT processes, data modeling, distributed computing frameworks like Trino, Apache Spark, Databricks and Snowflake. Effective at collaborating with data scientists, engineers and analysts in cross-functional teams. Strong research background with multiple publications focussing on Machine Learning and Explainable AI.

EXPERIENCE

New Jersey Institute of Technology — Newark, NJ
Research Assistant | NJIT Engineering Education Research — July 2024 — May 2025

- Analyzed institutional undergraduate student data with machine learning and developed regression models (Multiple Linear Regression, Random Forest, XGBoost, LightGBM) in R
- Performed ablation studies to determine the influence of academic, demographic and socioeconomic factors on students performance

Research Assistant | Data and Knowledge Engineering Lab — Sept 2023 — June 2024

- Integrated xAI tools (LIME, SHAP, Anchors, PDP, ALE) into "SolarFlareNet"—a transformer framework for space weather research
- Published research findings at FLAIRS 2024 and IEEE ICTAI 2024

Impetus — Bengaluru, India
Data Engineer — July 2020 — Aug 2023

- Leveraged Apache Parquet snappy compression and Amazon S3 life cycle policies reducing storage and I/O costs by 30%
- Designed data models for KPI generation and efficiently handled 10+ ad hoc requests, ensuring alignment with business needs
- Developed idempotent ETL pipelines that process terabytes of data using Apache Spark
- Utilized AWS MWAA to orchestrate complex data workflows with data quality checks, enhancing pipeline reliability and operational efficiency
- Leveraged AWS Glue Data Catalog and AWS Lake Formation to standardize metadata management and enforce data governance policies, reducing integration complexities and accelerating data discovery

TECHNICAL SKILLS

Languages	:	Python (PySpark, Polars, Pandas, Boto3), SQL, Scala (Apache Spark), Go, Rust, Bash, R, C
Databases	:	PostgreSQL, MySQL, Oracle (PL/SQL), MongoDB, SQLite, Microsoft SQL Server, Neo4j, DuckDB
Cloud	:	AWS (S3, Glue, Lambda, Athena, Redshift, Aurora, RDS, DynamoDB, Firehose, SageMaker Data Wrangler)
		Azure (Data Factory, Data Lake Storage, Synapse Analytics, Blob Storage)
		GCP (Cloud Storage, BigQuery, Dataflow, Dataproc, Bigtable)
Big Data	:	Trino, Databricks, Snowflake, Apache Spark, DBT, Apache Flink, Apache Kafka, Apache Hive
ETL/ELT Tools	:	Apache Airflow, Dagster, Pentaho Data Integration, Informatica
Visualization	:	Tableau, Looker, Power BI, Dash (Python), Apache Superset, Excel
Data Modeling	:	Normalization (3NF), OBT, Star Schema, Snowflake Schema, Data Vault
CI/CD	:	Git, GitHub, GitLab, GitHub Actions, Docker, Kubernetes, Terraform, Jenkins

PROJECTS

EDA on Global Soil Respiration Data | *MySQL, Tableau*   Github

- Analyzed the Global Soil Respiration Database (SRDBv5) by cleaning and querying data in MySQL and creating interactive Tableau dashboards for visual insights

WebScraping Using R | *R, rvest, ggplot*   Github

- Developed an R-based web scraping tool using rvest to collect and update articles from "Parasites & Vectors", optimizing resource usage and performing data cleaning and exploratory analysis with regex and ggplot

AlgoTrade API | *Python, yfinance, Pandas, Tensorflow, ks-api-client*   Github

- Developed a fully automated NSE stock trading bot in Python by integrating real-time and historical data with yFinance, training ML models (including LSTM) for stock price prediction, and executing live trades via the Kotak Securities API

RESEARCH PUBLICATIONS

1. Interpretable Deep Learning for Solar Flare Prediction — IEEE ICTAI 2024 — 2024
2. An Interpretable Transformer Model for Operational Flare Forecasting — FLAIRS 2024 — 2024

CERTIFICATIONS

- Data Engineering Bootcamp — DataExpert.io — 2025
- Google Data Analytics Professional Certificate — Coursera — 2023

EDUCATION

New Jersey Institute of Technology — Newark, NJ
*Master of Science in **Data Science*** — GPA: 3.89/4

SRM University AP — Amaravati, India
Bachelor of Technology — GPA: 8.51/10