

Vinay Kiran Raju

Data Engineer

raj.vinay2408@gmail.com | +91 9491857211 | [LinkedIn](#) | [Portfolio](#) | [Blog](#)

PROFILE SUMMARY

Results-driven Data Engineer with 4+ years of experience designing, building, and optimizing scalable **data pipelines**, ETL/ELT solutions, and real-time analytics platforms. Skilled in **SQL Server**, **Python**, **Azure Data Factory (ADF)**, **Synapse**, and **Databricks**, with proven success in integrating diverse data sources, ensuring data quality, and improving processing efficiency by up to 40%. Adept at working with structured, semi-structured, and unstructured data across relational, NoSQL, and data lake environments. Strong collaborator with cross-functional teams, aligning data engineering solutions with business and architectural requirements

TECHNICAL SKILLS

- Programming:** Python (Pandas, NumPy, Scikit-Learn, Matplotlib, Plotly), SQL, PySpark, Scala.
- Data Engineering:** ETL/ELT Development, Data Modeling, ERD & Data Flow Diagram Creation, Orchestration (Apache Airflow, ADF)
- Cloud & Streaming:** Microsoft Azure (ADF, Synapse, Data Lake Storage), AWS (S3, Glue, Redshift, Lambda, Kinesis), Databricks, Hadoop, Apache Kafka
- Databases:** MySQL, Snowflake, Redshift, Hive
- Visualization & Tools:** Tableau, Matplotlib, Seaborn, Git, Microsoft Excel (VBA, Macros)
- Best Practices:** CI/CD Deployment, Data Quality & Validation, Data Privacy, Process Automation

WORK EXPERIENCE

Cognizant

New Jersey, USA

Data Engineer

Apr 2024 - Present

- Developed and optimized **ETL/ELT pipelines** integrating 10+ data sources into **Azure Data Lake Storage** and **Amazon S3**, reducing data latency by 20%.
- Designed and maintained **complex SQL Server queries** and Python scripts for data transformation, validation, and quality checks, ensuring 99% accuracy across pipelines.
- Leveraged **Azure Synapse** and Databricks to accelerate large-scale data processing, improving query performance by 15% for analytics workloads.
- Built **ERDs** and **data flow diagrams** to document and align database structures with architectural requirements.
- Orchestrated and monitored data workflows using **Apache Airflow** and **Azure Data Factory**, implementing error-handling and retry logic to ensure 24/7 reliability.

Natsoft Corporation

New Jersey, USA

Data Engineer

Mar 2023 - Apr 2024

- Engineered **real-time payment analytics pipelines** processing over 250TB of transactional data using AWS Lambda, S3-based data lakes, and Redshift.
- Implemented schema evolution and compliance-driven data transformations with PySpark, reducing anomalies by 30%.
- Optimized ETL loads with AWS Glue and SQL tuning, improving Redshift query performance by 40%.
- Developed **data validation frameworks** with Airflow DAGs, cutting payment analytics reporting time by 30% through automated Tableau dashboards.

Hyderabad, India

Data Analyst

Sep 2020 - Jan 2022

- Processed and analyzed 5M+ FX trade records using MySQL and Python to support **trade analytics** and risk assessment.
- Created **real-time Tableau dashboards** for FX trade data, reducing reporting time by 40% for 50+ stakeholders.
- Built machine learning models for trade risk prediction, enhancing decision-making efficiency by 25%.

PERSONAL PORTFOLIO PROJECTS

Zillow End-to-End Automated ETL Pipeline

Jan 2024 - Feb 2024

- Developed an end-to-end ETL pipeline for **Zillow**, extracting real estate data from their API, transforming it on an **EMR** cluster using **PySpark**, and loading it into S3.
- Extracted Automated the workflow with **Apache Airflow** and integrated **Tableau** for data visualization, showcasing expertise in data transformation and analysis.

Loan Default Prediction

Jan 2023 - Mar 2023

- Developed a Python-driven **Decision Tree ML model** for loan default prediction with **95.21% accuracy** and **92.17% AUC** on 35,747 entries.
- Optimized Improved data quality by handling missing values, encoding categorical features, and eliminating duplicates; visualized key insights using **Tableau**, enhancing loan risk assessment.

EDUCATION

Pace University, Seidenberg School of Computer Science and Information Systems

New York, USA

MS in Data Science | Concentration: Data Analysis, Visualization & Machine Learning

Date: May 2023

Gandhi Institute of Technology and Management

Andhra Pradesh, India

Bachelor of Technology in Electronics and Communication Engineering

Date: Apr 2021

CERTIFICATES

- **Kore.ai Advanced Chatbot developer** – Kore.ai, Mar 2024
- **Snowflake BUILD 2023 LLM Bootcamp** – Snowflake, Dec 2023
- **Azure Data Fundamentals** – Microsoft, Apr 2023
- **Tableau Desktop Specialist** – Tableau, Mar 2023
- **Google Data Analytics Professional Certificate** – Coursera, Nov 2022