Vamsi Krishna Reddy Siddamreddy

USA | (405)-269-3367 | siddamvamsi2025@gmail.com| linkedin.com/in/siddamvamsi/ | github.com/SiddamVamsi264| Open to Relocate

PROFESSIONAL SUMMARY

Data Engineer with 3+ years of experience designing scalable, production-grade data pipelines in cloud environments (AWS). Proficient in building and optimizing ETL workflows using **AWS Glue**, **Redshift**, **SQL**, and **Python**, with a strong focus on automation, monitoring, and data reliability. Skilled in processing **1TB+ datasets**, integrating third-party sources like Salesforce, and enabling analytics and machine learning workflows.

TECHNICAL EXPERIENCE

Data Engineer (GRA) | *Oklahoma State University* ***Aug '23 – Present***

- Designed and maintained scalable ETL pipelines using **Python** and **SQL** to support academic and HR data needs.
- Built automated data ingestion frameworks to process data from 200+ sources, reducing processing errors by 25%.
- Integrated datasets into Tableau via scheduled pipelines, ensuring real-time dashboard accuracy for institutional KPIs.
- Collaborated with IT and business teams to standardize data models and define governance rules.

ETL Analyst| *Tata Consultancy Services* | *Client: Centrica Hive Home* ***Sept '21 – July '23***

- Developed over 50 data pipelines using **AWS Glue**, transforming structured/unstructured data from **S3** to **Redshift**.
- Automated ingestion of 1TB+ weekly **Salesforce extracts** using **Jenkins**, cutting refresh delays by 35%.
- Wrote and optimized complex **SQL stored procedures** for reporting layers and fact/dimension models.
- Implemented custom monitoring scripts in **Bash** and **Python**, achieving 100% job uptime and error logging.
- Coordinated closely with BI teams to align ETL outputs with business logic and SLA requirements.

Data Engineer | *Cognizant Technology Solutions* | *Client: Nestlé* ***Nov '20 – Sept '21***

- Developed **Python-based ETL scripts** to process **2M+ customer records** for sales and marketing data marts**.**
- Built **data ingestion** and **aggregation pipelines**, ensuring accurate and timely data delivery.
- Supported job scheduling**,** error handling**,** and post-load validations in collaboration with **DevOps**.
- Maintained source-to-target mappings and ensured pipeline accuracy through QA coordination**.**

SKILLS

Programming & Querying	**:** Python (Pandas, NumPy), SQL, Bash
Cloud & ETL	**:** AWS Glue, Redshift, S3, Jenkins, Lambda
Statistical Methods	**:** A/B Testing, Clustering (K-means), Time-Series Analysis
Databases	**:** MySQL, SQL Server, MongoDB, Salesforce
Workflow Automation	**:** Jenkins, Git, Cron, Jira
Data Modelling	**:** Star Schema, 3NF, NoSQL
Certifications	**:** Tableau Desktop Specialist, AWS Cloud Practitioner

EDUCATION

MS in Management Information Systems | *Oklahoma State University* *05/25*

- Key Courses: Programming for Data Science, Data Warehousing, Descriptive & Predictive Analytics

PROJECTS

Traffic Accidents Data Analytics System | *Python, SQL Server, MongoDB* ***GitHub***

- Engineered end-to-end pipeline for **700K+** UK road accident records across **OLTP**, **OLAP**, and **NoSQL** layers.
- Automated data extraction, transformation, and loading with **Python** and **SQL stored procedures**.
- Created reporting-ready views and summary tables for consumption by visualization tools.

Accident Severity Prediction Pipeline | *KNIME, Python* ***GitHub***

- Integrated **95K+ records** from **4 CRSS relational tables** into a unified dataset (~38K records) using **modular KNIME workflows**, reducing noise and improving data quality.
- Engineered **30+ features** (e.g., vehicle age, overspeed indicators, collision flags) and applied **SMOTE** to balance class ratio (3:1), improving minority-class recall by **20%**.
- Automated the full **ETL pipeline** with reusable nodes for **data ingestion**, **transformation**, and **feature engineering**, achieving **100% workflow reproducibility**.