

# Naveen Kumar Pulamarasetty

Visakhapatnam | naveenkumarp051@gmail.com | 9182490735 |

[linkedin](#)

## Summary

---

Data Engineer with 1+ year experience in SQL, Pyspark, AWS, Informatica. Skilled in building Data pipelines with AWS Glue and EMR, and implementing SCD types for historical tracking. Cloud integration and Automation.

## Skills

---

**Languages:** SQL, Python

**Analytics and Visualization:** Pandas, PySpark, Matplotlib, Qlik sense, Tableau.

**Tools and Technologies:** Informatica, AWS Glue, EMR, Lambda, Redshift, Apache AirFlow, GIT, Azure Devops.

## Experience

---

**Data Analyst**, Cognizant – Chennai June 2024 – present

- Worked on the analysis and preparation of Source-to-Target (S2T) transformation documents to support the migration of provider data from the legacy system (Cactus) to the new platform (TruProvider).
- Analyzed and compared database models of both applications to identify schema differences and define accurate data mappings.
- Developed SQL logic to extract, transform, and load (ETL) data for derived components, ensuring data consistency and preventing data loss during migration.
- Collaborated with business and technical teams to interpret transformation rules and implement business logic for complex data elements.
- Participated in testing and validation phases by verifying data outputs and resolving discrepancies to ensure alignment with business requirements.

**Data Analyst Intern**, Cognizant – Coimbatore Jan 2024 – May 2024

- Gained foundational knowledge in Data Warehousing and SQL.
- Designed and developed interactive dashboards using Qlik Sense, enabling data-driven decision-making through clear and informative visualizations.

## Education

---

**Raghu Institute of technology**, Visakhapatnam, A.P June 2020 – May 2023

- B.Tech in Mechanical Engineering - GPA: 8.9/10.0

## Projects

---

**SAS DataMapper** Python, SAS code, LLM

- Developed a Python-based tool that parses SAS code and generates a Source-to-Target (S2T) transformation document to streamline data lineage tracking.
- Integrated Gemini AI model to intelligently interpret SAS logic and automate mapping of source datasets to target structures.
- Enhanced data engineering workflows by reducing manual effort and improving accuracy in documentation for ETL processes.

**Hetrogenious DataBase connector** Python, Gemini LLM

- Developed a Python-based application to compare data across on-prem and cloud databases using AI-driven natural language queries and manual SQL inputs.
- Leveraged Gemini AI for query generation and Apache Spark for distributed processing, enabling fast and scalable data validation with detailed difference reports.