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 Engineer, Cognizant – Chennai   June 2024 – present

- Successfully migrated legacy Informatica mappings to AWS Glue using PySpark, modernizing ETL workflows for improved scalability, maintainability, and performance.
- Enhanced pipeline execution efficiency by approximately 40%, resulting in significant cost savings.
- Designed and implemented Slowly Changing Dimensions (SCD) Type 1 and Type 2 logic to manage evolving dimensional data, ensuring historical accuracy and data integrity.
- Implemented advanced performance optimization techniques such as Broadcast joins, caching, persisting, and checkpointing.
- Worked with a hybrid architecture combining AWS Glue PySpark, Glue Python shell jobs, and transient EMR clusters, selecting the optimal compute environment based on data volume and transformation complexity, This approach ensured scalability, cost-efficiency, and performance optimization across diverse ETL workloads.

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

IICS version control tool   Python, CI/CD, AzureDevops

- Designed and implemented a CI/CD pipeline in Azure DevOps to automate version control and comparison of Informatica mappings.
- Engineered a solution to extract and parse mapping XML and JSON files, highlighting differences in an HTML report with intuitive color coding (red for missing, green for matched transformations).
- Enabled seamless input of mapping names and automated transformation-level comparison, improving traceability and reducing manual validation effort.

Hetrogenious Data base 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.