

YASWANTHREDDY VARIKUTI

DATA ENGINEER|AI

☎ +917997792835 | ✉ yaswanthreddyvarikuti@gmail.com | [linkedin.com/in/varikuti-yaswanthreddy/](https://www.linkedin.com/in/varikuti-yaswanthreddy/) | Portfolio-link

Summary

Data Engineer with over 2 years of experience building cloud-native data infrastructure across Google Cloud Platform and AWS. Designed and deployed production-grade pipelines that processed over 10PB of data, reduced job runtime by 40%, and automated over 50 manual workflows. Proficient in implementing data transformation frameworks using Spark, Airflow, and dbt, and in integrating quality checks with Great Expectations. Collaborated with cross-functional teams to streamline ML pipelines and ensure robust, traceable data systems in production. Delivered hands-on training and support to 20+ stakeholders, driving operational efficiency and platform adoption.

Experience

Cognizant Technologies | *Data Engineer*

June 2023 – Present

- Built and deployed scalable pipelines to migrate 10PB+ data from on-prem to AWS using Glue and Control-M, reducing execution time by 40%.
- Designed automation scripts and machine learning tools that optimized operations, decreasing manual workload by 80%.
- Implemented data pipelines from S3 to Snowflake with CDC and SCD logic, enhancing data traceability and reporting by increasing the.
- Reduced downstream pipeline latency by 30% with improved orchestration and modular pipeline design.
- Conducted over 100 validation tests to ensure data accuracy, achieving 95% success in production.
- Acted as the point of contact for cloud pipeline deployments, ensuring alignment between teams and timely delivery with 100% accuracy and requirements.
- Led training sessions for 20+ analysts, developers, and managers, fostering collaboration, mentorship, and effective delivery skills.
- Collaborated with 6+ cross-functional teams, optimizing data pipelines and ensuring seamless integration with Cloud Storage and Control-M, minimizing processing time by 30%.
- Documented workflow processes and reusable components, reducing troubleshooting time by 25%.

Cognizant Technologies | *Big Data Engineering Intern*

January 2023 – May 2023

- Automated more than 50 ETL tasks using Python and SQL, saving 20+ hours of manual work each week.
- Designed and implemented an ETL pipeline and database system for healthcare analytics, improving patient record retrieval by 80%.
- Refined existing ETL jobs, cutting down runtime by 30% and reducing data anomalies significantly.
- Collaborated with 3+ senior engineers on production deployments, gaining experience in Spark and GCP-based infrastructure.

Key Projects

Ecommerce Batch Data Pipeline | GCP, Databricks, Airflow, Delta Lake, Big Query(Verify):

- Built a multi-layer data architecture with Delta Lake on GCS, applying 10+ DLT validations to reduce data issues by 15%.
- Automated ingestion and transformation via Airflow and Spark, maintaining 99.5% uptime.
- Created dimension tracking logic (SCD Type 2, CDC) for analytics accuracy, which decreases the storage costs and latency by 30%.

Medical Image Segmentation | LFB Network, TensorFlow:

- Developed a segmentation model that reduced inference time by 30% and improved accuracy by 15%.
- Addressed corrupted data samples by validating and correcting 500+ image records, enhancing model reliability.

Education

GMRIT College | B. Tech in Electronics and communication Engineering

June 2018 – March 2022

- GPA: 8.34| Minor: IoT and Data Science.

Skills

- **Languages:** Python, SQL, Java
- **Big Data & ETL:** Spark, PySpark, Hadoop, Hive, Airflow, Control-M, DBT, Informatica.
- **Cloud Platforms:** AWS (Glue, EMR, S3, EC2, SageMaker, Cloud Watch, Dynamo DB), GCP (Big Query, DataProc, Pub/Sub, Vertex AI), Snowflake, Azure, Databricks, Terraform.
- **Databases:** MySQL, PostgreSQL, NoSQL (MongoDB, Cassandra)
- **Data Modelling:** SCD, CDC, Start & Snowflake Schema, Dimensional Modelling.
- **ML/AI Tools:** TensorFlow, Keras, Scikit-learn, Generative AI, NLP, NumPy, Azure Machine Learning
- **DevOps & Infra:** Docker, Jenkins, Git, Terraform.
- **API Integration & Automation:** REST API, Postman, CI/CD, Python Automation, Shell Scripting.

Certifications

- Databricks Certified Data Engineer – Professional
- Snowpro Core – Snowflake
- Google Cloud Associate Engineer – GCP
- Azure Data Science & Machine Learning Associate – Microsoft
- AWS Cloud Practitioner – AWS
- Generative AI – Cognizant
- AI Foundations – Oracle Cloud