

# 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 2+ years of expertise in automating data pipelines, optimizing ETL processes, and deploying scalable cloud solutions. Skilled in machine learning, big data technologies, and cloud platforms like GCP, AWS, and Snowflake. Focused on improving system performance, integrating AI-driven analytics, and ensuring data accuracy for business insights. Experienced in developing secure, efficient workflows that enhance decision-making and operational efficiency across diverse industries.

### Experience

---

#### Cognizant Technologies | *Data Engineer*

June 2023 – Present

- Integrated **10TB+** of data across AWS, Teradata, and Snowflake, reducing query execution time by **40%**.
- Build Data Pipelines to migrate **10 PB+** Data from On-Prem to Cloud using Control-M, AWS Glue.
- Created **10+** AI-driven automation tools & ML models with innovative blocker-solving skills, improving workflows & boosting productivity by 80%.
- Collaborated with **6+ cross-functional teams**, optimizing data pipelines and ensuring seamless integration with **Cloud Storage and Control-M**, minimizing processing time by **30%**.
- Acted as **Point of Contact (POC)** for a project building pipeline for Data generation from On-Prem to Cloud S3 and S3 to Snowflake.
- Conducted **100+ data validation tests**, identifying inconsistencies and improving **data quality by 95%** for business intelligence and reporting.
- Revised ETL pipelines, improving data ingestion performance by **30%** across Cloud Storage and Control-M, ensuring effective **teamwork** and **collaboration** with cross-functional teams.
- Led training sessions for 20+ analysts, developers, and managers, fostering collaboration, mentorship, and effective delivery skills.

#### Cognizant Technologies | *Big Data Engineering Intern*

January 2023 – May 2023

- Engineered over three data analysis models for hospital administration systems, decreasing patient data retrieval time by 40%.
- Optimized **10+ ETL workflows**, reducing the runtime by **30%** and reducing the data anomalies and integrity to less than **5%**.
- Designed a hospital database & ETL pipeline, achieving **80%** efficiency in data processing.
- Automated **50+ tasks** using Python & SQL, saving **20+ hours/week**.
- Mentored by **5+** senior data engineers, gaining hands-on expertise in big data, ML, and cloud technologies, demonstrating a willingness to **learn** and **open-mindedness**.

### Projects

---

#### LFB Network for Robust Medical Segmentation:

- Developed a Lightweight Feature Boosting (LFB) Network, improving segmentation accuracy by **15%**.
- Identified and rectified a critical image corruption issue affecting **5%** of incoming scans, personally correcting **500+** medical images to ensure accurate diagnoses and prevent potential misdiagnoses.
- Experimented machine learning model performance, reducing inference time by 30%, which significantly improved real-time diagnosis support and system responsiveness for medical applications.

#### Ecommerce Batch Processing Data Pipeline ([Verify](#)):

- **Designed and built a multi-layered (Bronze, Silver, Analytic) data architecture** utilizing Delta Lake on GCS, enforcing data quality checks (e.g., 10+ DLT expectations) which **reduced downstream data errors by 15%** for BI reporting.
- **Orchestrated end-to-end data workflows with Apache Airflow**, automating data ingestion from GCS, complex transformations on Databricks, and direct loading of **partitioned Parquet files (average 500MB/day)** into Big Query, achieving **99.5% pipeline uptime**.
- **Implemented robust Change Data Capture (CDC) and Slowly Changing Dimension (SCD) Type 2 logic** for customer and product dimensions using Apache Spark, enabling precise historical analysis and ensuring **100% data traceability** for critical business insights.

### Education

---

#### GMRIIT College | B. Tech in Electronics and communication Engineering

June 2018 – March 2022

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

## Skills

---

- **Programming Languages:** Python, SQL, Java (intermediate)
- **Big Data Technologies:** Spark Ecosystem, PySpark, Hadoop, Hive
- **Cloud Platforms:** Google Cloud Platform (GCP, IAM, Pub/sub, Vertex AI, Big Bucket, Big query, Data Proc), AWS (Glue, EMR, S3, EC2, Sage maker, Athena), Snowflake (Snowpipe, Snow stream, IAM), Azure, Data Bricks, Terraform.
- **Data Engineering & ETL:** Data Warehousing, Data Modelling, ETL Pipelines, Control-M, Airflow, Informatica, DBT
- **Databases & Storage:** MySQL, PostgreSQL, Big Query, Cloud Storage, NoSQL (MongoDB, Cassandra)
- **Data Analysis & Visualization:** Power BI, Excel (PowerPivot, Macros),
- **Machine Learning & AI:** Neural Networks, Generative AI, Deep Learning, Scikit-learn, Keras, TensorFlow, NumPy
- **DevOps & Scripting:** Big Bucket, GitHub, Shell Scripting, PowerShell, Docker, Jenkins
- **API Integration & Workflow Automation:** REST API, Postman, CI/CD, Python Automation

## Certifications

---

- Snowpro Core Certification – Snowflake
- Azure Data Science & Machine learning – Microsoft
- AI Foundations Certification – Oracle Cloud
- IBM Watson & AI Badges – IBM
- Generative AI Certification – Cognizant.
- Networking Fundamentals – Cisco
- Project Management Professional – Google (Coursera)
- Data Bricks Gen AI – Databricks
- Data Bricks Professional Data Engineering – Databricks
- MDM SAAS – Informatica