Suyog Mainali

Williamsburg, VA · +1 (757) 206-8339 · smainali01@wm.edu · linkedin.com/in/suyog41 · Portfolio · GitHub

PROFESSIONAL SUMMARY

Data Engineer (MSBA '26, William & Mary) specializing in consumer data platforms, ETL/ELT automation, and ML-ready datasets using Python, PySpark, SQL, Airflow, dbt, and AWS/Azure. Experienced in data lakes, data warehouses, and feature store design with Docker/Kubernetes and CI/CD. Skilled in building high-throughput data pipelines, enforcing data quality and lineage, and deploying analytics for AI/ML and BI applications.

EDUCATION

William & Mary, Raymond A. Mason School of Business, Williamsburg, VA, USA August 2025 - May 2026
Master of Science, Business Analytics (**Expected May 2026**)

Vellore Institute of Technology (VIT), Vellore, India August 2018 - May 2022
Bachelors of Technology, Computer Science Engineering

PROFESSIONAL EXPERIENCE

Zakipoint Health, Sanepa, Lalitpur, Nepal July 2020 - August 2025
Senior Database Engineer

- Designed data pipelines (ETL/ELT) using Python, PySpark, and SQL to integrate claims, eligibility, provider, and EHR datasets into AWS S3 data lakes (Parquet/JSON).
- Built and orchestrated Airflow DAGs and dbt transformations for incremental modeling and testing, enabling continuous data ingestion and transformation.
- Deployed dimensional (star/snowflake) models to Redshift/Snowflake, tuning distribution/sort keys and clustering for performance optimization.
- Developed data quality, monitoring, and SLA tracking frameworks for near-real-time BI updates with 99.8% uptime.
- Automated deployment using Git, Jenkins CI/CD, and Kubernetes, supporting scalable workloads for consumer health analytics and machine learning datasets.

Om Hospital, Kathmandu, Nepal April 2018 - July 2020
Database Manager

- Built data warehouse and data lake pipelines with Python, SQL, and Azure ADLS, enabling data consolidation across clinical and operational systems.
- Developed data marts for executive dashboards in Tableau, improving operational decision-making and reducing patient wait times by 25%.
- Standardized schemas, indexing, and partitioning; improved query performance by 30%.
- Implemented data governance documentation, data dictionaries, and validation frameworks.

CORE SKILLS & COMPETENCIES

- Languages/Query: Python, SQL, PySpark
- Data Engineering: ETL/ELT, Data Pipelines, Data Lakes, Data Warehousing
- Orchestration & Automation: Airflow, dbt, Jenkins, CI/CDModeling & Storage: Redshift, Snowflake, PostgreSQL, MySQL, Parquet/JSONStreaming & Messaging: Kafka
- Cloud Platforms: AWS (S3, EC2, IAM, Glue), Azure (ADLS)
- DevOps & Ops: Docker, Kubernetes, Monitoring, Logging, SLA ManagementMachine
- Learning Readiness: Feature Engineering, Data Quality Validation, Model Data Pipelines
- Visualization & Reporting: Tableau, REST APIs, Automated Reports

PROJECTS

- **The $100 Question (FRED API Dashboard)**: Ingested FRED APIs via ETL/ELT to raw, staged, modeled layers with dbt tests; scheduled in Airflow, published to Aiven PostgreSQL, deployed on Render with CI/CD.
- **Hybrid Encryption Demo (Streamlit)**: Implemented AES-256-GCM + RSA-OAEP with config-driven JSON I/O; exposed REST services, containerized with Docker, deployed on Render.
- **LLM-Powered Knowledge Retrieval App**:Built ingestion→chunking pipelines with FAISS + TF-IDF, metadata in PostgreSQL; REST endpoints and Render deployment for low-latency retrieval.
- **Amazon-Reviews-ETL-Pipeline**: Developed an **ETL pipeline** in Python to scrape, clean, and load data from 74K+ Amazon products into PostgreSQL, with automated error handling and resume-from-checkpoint capability.