Suyog Mainali

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

PROFESSIONAL SUMMARY

Data Engineer (MSBA '26) with 5+ years of experience building reliable ETL and streaming pipelines, Spark workloads, and cloud-native data platforms on AWS. Experienced in PySpark, Kafka, Airflow, and dbt, with a strong focus on data quality, scalability, and feature-ready datasets that power analytics and machine learning products.

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
Bachelor of Technology, Computer Science Engineering

PROFESSIONAL EXPERIENCE

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

- Architected the company's flagship real-time cost prediction engine, processing **10 TB+ daily** across claims and provider data to deliver **sub-second** pre-visit out-of-pocket estimates to members.
- Took ownership of unstable core data pipelines and rebuilt PySpark and SQL workflows on AWS Glue, cutting recurring processing failures by **60%+** and stabilizing downstream analytics and ML systems.
- Built executive dashboards aggregating real-time feature usage and member behavior, eliminating **2-day manual reporting cycles** and supporting client pitches, renewals, and growth decisions.
- Enabled production AI models by designing feature pipelines from claims and clinical data, reducing ML data preparation time by **~40%** for risk scoring and care gap detection.
- Owned analytics reliability across Snowflake, Redshift, Tableau, and CI/CD workflows, maintaining **99.8% uptime** under HIPAA and SOC 2 requirements.

Om Hospital, Kathmandu, Nepal   April 2018 - July 2020
Data Engineer

- Led migration from paper records and legacy systems to DHIS2 for a hospital handling **250K+ annual visits**, preserving multi-year historical patient data.
- Built the hospital's first centralized data warehouse integrating visits, diagnoses, labs, and operations, enabling **organization-wide clinical reporting** for the first time.
- Implemented a digital visit recording and telemedicine system in Python, supporting engagement with **10K+ patients per month** and enabling longitudinal patient histories across departments.
- Trained **50+ doctors, nurses, and admin staff** through workshops and documentation, driving adoption and consistent use of new digital workflows.

CORE SKILLS & COMPETENCIES

- **Languages:** Python, SQL, PySpark, Bash
- **Data Engineering:** ETL and ELT pipelines, Apache Spark (RDD, DataFrame), Airflow, dbt, Kafka, Kinesis, AWS Lambda, Step Functions, batch and streaming systems
- **Cloud & Warehousing:** AWS (S3, Glue, EMR, Redshift), Azure (ADLS Gen2, Data Factory), Snowflake; dimensional data modeling and lakehouse architectures
- **Databases & Storage:** Redshift, Snowflake, PostgreSQL, MySQL, MongoDB, DynamoDB, Elasticsearch; Parquet, ORC, Avro, Iceberg and Delta formats
- **ML & ML Data Pipelines:** Feature engineering pipelines, training and inference datasets, embeddings and vector search (FAISS), MLflow; scikit-learn, TensorFlow, PyTorch
- **APIs & Analytics:** REST APIs, FastAPI, Flask, Plotly Dash, Tableau, Power BI, Streamlit
- **Microsoft & Azure Stack:** Azure Data Factory, ADLS Gen2, Power BI, Microsoft Fabric

PROJECTS

- **Amazon Reviews ETL Pipeline (SerpApi + Playwright)** – Processed 500K+ product listings and reviews and built a query API used for analytics dashboards and ranking tests.
- **Siftline – Grounded Document QA** – Built a private Q&A system using FAISS + TF-IDF retrieval and Transformer summarization to search 10,000+ internal documents for policy, compliance, and research use cases.

Work Authorization: F-1 student on STEM OPT (3 years of US work eligibility; requires sponsorship afterward).