Mekhma Tamang

mekhmanju@gmail.com | mekhzz.github.io | Kathmandu, Nepal

Data Professional & Aspiring Analyst Self-motivated professional with experience in report design and building scalable ETL/ELT architectures. Proven track record in implementing Medallion Architecture and RAG systems to automate workflows and drive operational efficiency. A collaborative problem-solver eager to master new technologies and deliver high-performance data insights

Skills

- **Programming Languages:** Python (PySpark, Pandas), SQL (SparkSQL, PostgreSQL), Shell Scripting
- **Tools & Framework:** Databricks, dbt, LaTeX, Appsheet(no-code), Google Earth Engine, Docker, Git, Tableau, Looker studio, ETL/ELT, Medallion architecture, Linux, Reverse API Engineering, Selenium, BeautifulSoup, AWS RDS, AWS S3

Education

- **Bachelor's Degree in Computer Applications**
  Sept 2019 - Oct 2024
  Mahendra Morang Adarsha Multiple Campus — Tribhuvan University

Experience

- **Data Filtering Associate**
  Dec 2024 - present
  Veel, Kathmandu
  - ETL Automation: Architected custom scrapers (Meta, LinkedIn, CRM) and integrated them into a Medallion Architecture to automate the validation of 1M+ records monthly, reducing manual intervention by 40%.
  - Infrastructure Optimization: Developed lightweight, production-grade scrapers for the AI/ML team, increasing data delivery speeds by 50x and enhancing model training accuracy by 40% through refined data selection.
  - Generative AI (RAG) Development: Researched and implemented a Retrieval-Augmented Generation (RAG) system using the Vanna framework, enabling non-technical stakeholders to perform natural language queries against "Gold Layer" SQL databases.
  - Data Governance: Designed a multi-stage automated validation system that cleanses large-scale datasets against global AI integrity standards, ensuring a zero-deficiency data lifecycle.

- **Agriculture Census Enumerator**
  Apr 2022 - Jun 2022
  Census Bureau of Statistics, Nepal
  - Conducted a comprehensive survey of around 900 rural households and agricultural holdings to collect critical data for national policy planning based on sample-based census.

- **Population Census Enumerator**
  Oct 2021 - Nov 2021
  Census Bureau of Statistics, Nepal
  - Managed large-scale field data collection for National Agricultural and Population Censuses, ensuring 100% accuracy for 700+ households while maintaining strict data integrity and validation protocols.

Personal Projects

- **Automated Web Server Log Analysis (Ubuntu/Apache/Bash)**
  - Infrastructure Monitoring: Configured an Ubuntu Server with Apache2 and UFW to simulate a live production environment.
  - Log Analysis Pipeline: Engineered a Bash script using awk, grep, and sort to parse access logs, identifying top request URLs, unique IPs, and 404 error trends.
  - Automation: Scheduled daily 2:00 AM automated reporting via Cron to ensure continuous visibility into server health and traffic anomalies.
  - *https://github.com/MekhzZ/AutomatedWebServerLogAnalysis*

- **Payer-Scale Medallion Lakehouse (Databricks/PySpark)**
  - Constructed a Medallion Lakehouse to process longitudinal patient data, implementing HIPAA-compliant SHA-256 hashing for PII/PHI protection.
  - Engineered Clinical Metric Flags for high-risk chronic cohorts and calculated Total Cost of Care (TCOC) to drive healthcare business intelligence.
  - I have documented my learnings from scratch on US Healthcare analytics in below GitHub link.
  - *https://github.com/MekhzZ/PayerScaleMedallionLakehouse*

- **HealthSparq Provider Data Engine (Python/Security)**
  - Developed a high-throughput synchronization engine to bypass Imperva WAF and retrieve 3,865+ restricted healthcare records.
  - Engineered Specialty-Based Sharding to overcome API pagination limits and implemented a Stateless Token Handshake for stable authentication.
  - Learned about token authentication via api call, sharding to get limit locked FHIR data from provider directory.
  - *https://github.com/MekhzZ/HealthSparqProviderDataEngine*

- **Wikipedia Tabular Data Scraper (Streamlit/Pandas)**
  - Architected a live web-tool to automate the extraction of unstructured Wikipedia tables into structured, analysis-ready CSV/JSONL formats.
  - Integrated custom User-Agent rotation and request throttling to ensure compliance with web scraping policies.
  - learned to seemlessly integrate UI with custom scraper. Updating packages periodically helped me to learn how packages played vital role in deployment.
  - *https://github.com/MekhzZ/WikipediaTabularDataScraperModule*1

Certifications

Relational Database, *Applied Statistics with Python*, *Data Analysis with Python*,
Data Visualization using Tableau, *Building No-Code Apps with AppSheet: Foundations*,
LFS101: Introduction to Linux, *Microsoft Fabric*, *Databricks Fundamentals*,
Get Started with Databricks for Data Engineering, *DevOps*