

MADHU GHIMIRE

Data Engineer

Kathmandu, Nepal | +977 9864575872 | madhusudan2058@gmail.com | github.com/mgmr99 | linkedin.com/in/madhughimire

PROFESSIONAL SUMMARY

Data Engineer with 2+ years of experience building and scaling ETL/ELT pipelines for U.S. healthcare data. Led the migration of legacy MySQL batch processing to a distributed Apache Spark and Hadoop architecture, cutting processing time by over 90%. Skilled in Python, SQL, Pentaho Data Integration, AWS, Docker, and Elasticsearch, with hands-on experience owning new client implementations end-to-end, from data layout analysis to production release.

PROFESSIONAL EXPERIENCE

Data Engineer

Aug 2024 – Present

Zakipoint Health

Kathmandu, Nepal (Remote – U.S. Healthcare)

- Led migration of legacy in-house MySQL data processing to a distributed Apache Spark and Hadoop architecture, including cluster configuration and maintenance, reducing client processing time by over 90% (from 2–3 days to 4–5 hours).
- Built and maintained Bash automation scripts to index processed data from Hadoop to Amazon S3 and subsequently into Elasticsearch, streamlining the data indexing and search pipeline.
- Own new client implementations end-to-end: analyze incoming data layouts, design and build processing logic, write scrub scripts, and release to UAT/production under tight timelines.
- Design and build Python-based ETL pipelines that ingest data from REST APIs, outsourced databases, and files delivered via FileZilla/SFTP and Amazon S3.
- Developed API client automation using Docker, Linux, and Python, reducing recurring manual operational work across client integrations.
- Handle daily ad-hoc data requests and production issues, ensuring accurate and timely delivery for healthcare clients.
- Automated Pentaho job log monitoring with real-time Slack/email alerts for 5xx errors, reducing manual troubleshooting effort.

Data Engineer Intern

Apr 2024 – Jul 2024

Zakipoint Health

Kathmandu, Nepal

- Designed and optimized data pipelines for U.S. healthcare datasets using Pentaho Data Integration.
- Strengthened database management skills across SQL, Spark, Hadoop and PostgreSQL and used Elasticsearch to improve data indexing and search performance.

Backend Developer Intern

Dec 2023 – Mar 2024

Technorio Inc.

Kathmandu, Nepal

- Contributed to backend development and supported foundational software engineering practices in a production codebase.

PROJECTS

- Event-Driven Serverless ETL Pipeline on AWS — Built an end-to-end serverless pipeline ingesting JSON into S3, transforming to Parquet via AWS Lambda (Python), cataloging with AWS Glue, and enabling analytics through Amazon Athena.
- PAN Validation Pipeline (PostgreSQL) Built a data cleaning and regex-based PAN format validation pipeline with result categorization and automated summary reporting.

SKILLS

Languages & Databases: Python, SQL, MySQL, PostgreSQL

Big Data & ETL: Apache Spark, Hadoop, Pentaho Data Integration (Kettle), ETL Pipeline Design

Cloud & DevOps: AWS (S3, EC2), Docker, Linux, Bash Scripting, Git/GitHub

Search & Monitoring: Elasticsearch, Slack/Email Alerting, Log Monitoring Automation

EDUCATION

Bachelor of Science in Computer Science and Information Technology

Apr 2020 – Aug 2024

Orchid International College, Kathmandu