

Jayas Piya Data Engineer

✉ jayaspiya52@gmail.com ☎ +977-9865005541 [in /jayaspiya](#) [/jayaspiya](#)

<https://jayaspiya.com.np/>

Profile

Results-driven Data Engineer with 3+ years of experience building scalable healthcare data platforms using **Databricks, PySpark, and Snowflake**. Adept at designing robust **ETL pipelines**, optimizing OLTP/OLAP systems, and ensuring data quality and compliance with healthcare standards (**HL7, FHIR**). Passionate about data modeling, SQL performance tuning, and collaborating across product, engineering, and analytics teams to deliver end-to-end data solutions.

Work Experience

Abacus Insights Nepal

Jun 2025 – Present

Software Engineer II

- Continued role after BOT transfer from TechKraft Inc.
- Building cloud-native ETL workflows using Databricks and AWS to ingest and process healthcare data at scale.
- Partnered with business analysts to define metrics and KPIs for reporting.
- Supporting platform scalability and data governance in compliance with HIPAA.

Techkraft Inc

Software Engineer II

Jan 2024 – May 2025

- Led ingestion of HL7 and FHIR data into enterprise data warehouse.
- Optimized complex mapping queries for performance improvement and efficient data retrieval.
- Collaborating with engineering leads to design modular, testable pipeline components.
- Mentored junior engineers and contributed to technical documentation and team knowledge base.

Software Engineer

Oct 2022 – Dec 2023

- Designed scalable PySpark pipelines for patient record processing using Databricks.
- Built stored procedures and UDFs for reusable healthcare logic.
- Created monitoring scripts and data validation rules to ensure end-to-end data integrity.
- Wrote modular query to transform staging data into analytics-ready tables.

Data Trainee

Jul 2022 – Sep 2022

- Designed PL/SQL scripts for data validation and business rule enforcement.
- Built interactive dashboards and automation scripts with Python and Google Apps Script.
- Explored Tableau for visual storytelling of internal operations metrics.

Skill Sets

Languages & Tools: SQL, PL/SQL, Python, PySpark, Git, Javascript

Cloud & Platforms: AWS, Databricks, Snowflake, dbt

Data Engineering: ETL, Data Modeling (OLTP/OLAP), Data Warehousing

Analytics & Visualization: Tableau, Streamlit

Healthcare Standards: HL7, FHIR

Education

BSc(Hons) Computing

Softwarica College of IT and E-commerce
Data Analytics & Engineering

2019 – 2022

Projects

Healthcare FHIR Data Ingestion Framework [🔗](#)

Built a robust ETL pipelines to process FHIR-based patient encounter, procedure, and condition resources using PySpark. Implemented feature engineering pipelines to derive patient-level features like prior admissions, chronic condition flags, and encounter history trends. Enabled downstream analytics for patient outcomes and operational efficiency.

30-day Patient Readmissions Predictive Model [🔗](#)

Developed a predictive model to classify 30-day readmissions using XGBoost, Random Forest, and logistic regression. Engineered custom features from ICD-coded conditions and discharge dispositions to boost predictive power. Performed hyperparameter tuning and cross-validation to optimize model performance.

Information Retrieval System: Scholar-Like Academic Search Engine [🔗](#)

Developed a fully functional academic search engine focused on research outputs from the Research Centre for Intelligent Healthcare (RCIH), mimicking Google Scholar in design and experience. Focused on extracting metadata such as title, authors, and publication year, and supported keyword-based querying with relevancy ranking and subject classification functionalities

Retail Intelligence System [🔗](#)

Academic Project

Integrated Tableau dashboards to provide real-time business intelligence and leveraged data warehousing techniques to enhance data accessibility and analytical performance; developed backend infrastructure using Oracle Linux 6.7 and Oracle Database 11g to ensure high availability, concurrency control, and robust security.

Certificates

- dbt Fundamentals - dbt Labs [🔗](#)
- Google Data Analytics - Coursera
- Big Data Fundamentals with PySpark - Datacamp
- Oracle Database 12c: Advance SQL - LinkedIn
- Oracle Database 19c: PL/SQL - LinkedIn
- Software Engineering Principles Python - Datacamp