

Dinesh Saud

Scarborough, ON, CA | +1 (249) 989-4930 | dinesh.4777saud@gmail.com | [linkedin.com/in/sauddinesh](https://www.linkedin.com/in/sauddinesh)

EXECUTIVE SUMMARY

AWS Certified Data Engineer with 3+ years of experience in ETL/ELT development, data warehousing, and data modeling. Skilled in Python, PySpark, SQL, and Snowflake, with hands-on expertise in handling structured, semi-structured, and unstructured datasets. Experienced in building cloud-based pipelines (AWS Glue, Airflow, Azure Data Factory, Databricks) and optimizing queries for performance. Strong background in data validation, profiling, and pipeline automation, with growing exposure to AI concepts, prompt engineering, and retrieval-augmented generation (RAG).

PROFESSIONAL EXPERIENCE

Cotiviti	Kathmandu, NP
Data Engineer	Dec 2021 – Jul 2023
<ul style="list-style-type: none">Built ETL workflows in Python and PySpark (AWS Glue + Airflow) to process 2B+ rows monthly, ensuring clean and reliable data for analytics.Designed and optimized Snowflake & PostgreSQL data models, reduced reporting latency by 30% through SQL tuning and indexing.Integrated structured (SQL), semi-structured (JSON), and unstructured datasets from MongoDB, Hadoop, and RDBMS into a unified warehouse.Partnered with analysts to conduct data profiling, validation, and mapping, improving data quality across systems.Automated quality checks and alerts with Python + AWS Lambda, strengthening compliance and reporting reliability.	

Herald College	Kathmandu, NP
Research Assistant – Data Science	Jan 2021 – Dec 2021
<ul style="list-style-type: none">Performed EDA and statistical analysis on 500GB+ datasets using Pandas, NumPy, and SciPy.Applied machine learning models (regression, clustering, classification) with scikit-learn to derive insights.Created visualizations in Power BI, Matplotlib, and Seaborn to communicate findings.Contributed to data mapping and modeling for student research projects.	

Vox Crow Private Limited	Kathmandu, NP
Software Engineer Intern	Jun 2020 – Oct 2021
<ul style="list-style-type: none">Developed SQL-driven CRM reports that improved customer data visibility for 5+ projects.Resolved 15+ data integration and transformation issues, ensuring reliable reporting outputs.Enhanced ETL testing and defect tracking within Agile sprint cycles.	

PROJECTS AND LEADERSHIP

- Test Kit Classification (Client Project):** Processed 800+ medical scans with PyTorch/OpenCV, training ResNet-50 CNN models for automated ROI extraction and disease classification.
- Azure Data Engineering for Healthcare RCM:** Built a secure Azure Data Factory/Data Lake Gen2 pipeline with Azure SQL and Key Vault, automating KPI reporting for 10K healthcare records.
- Spark ETL Pipelines (Databricks Lakehouse):** Developed Spark/Delta Lake ETL with partitioning, bucketing, and Spark SQL to process 5M+ records for analytics and reporting.

EDUCATION

Georgian College	Barrie, CA
<ul style="list-style-type: none">Postgraduate Diplomas in Big Data Analytics (Deans list)Post graduate Diploma in Artificial Intelligence (Deans list)	Sep 2024 - Apr 2025 Sep 2023 - Apr 2024
Herald College Kathmandu Affiliated to University of Wolverhampton, UK	Kathmandu, NP
<ul style="list-style-type: none">B.Sc. (Hons.) Computer Science (<i>Best Final Year Project 2020</i>)	Jan 2017 – Jan 2020

CORE SKILLS

- **ETL/ELT & Data Platforms:** Python, PySpark, SQL, Airflow, AWS Glue, Azure Data Factory, Databricks, SSIS, Snowflake, PostgreSQL, Hadoop/Hive, Netezza
- **Data Quality, ML & Analytics:** Data profiling/validation, anomaly detection, Spark, scikit-learn, PyTorch, Power BI, Tableau, Matplotlib, Seaborn
- **Cloud & Delivery:** AWS, Azure, GenAI, RAG, AI-assisted SDLC tools, Agile (JIRA, sprint workflows, cross-team collaboration)