

Ngoc Dung Bui

Staff Data Developer | AI & Cloud Architect

Montreal, Quebec, Canada • buind173@gmail.com • (438) 337-6578

Staff-level Data & AI Engineer with 16+ years of software development experience, including **7+ years specialized in modern Cloud Data Architecture (Snowflake/AWS/Azure)**. Extensive background in enterprise system integration (SAP/Java), evolving into a focus on building scalable **GenAI** solutions and **MLOps** pipelines. Proven ability to lead technical initiatives, mentor teams, and bridge the gap between legacy enterprise data and modern AI-driven analytics.

WORK EXPERIENCE

Senior Data Engineer

IBM (Client: Air Canada), Montreal, Canada

March 2020–Present

- **Architected a hybrid Data Lakehouse & EDW ecosystem** to master increasing data complexity: leveraged **Databricks on Azure ADLS** (Medallion Architecture) for processing and **Snowflake** (Data Vault 2.0) as the central warehouse; streamlined **100TB+ of daily data**, improving accuracy by **35%** and reducing processing time by **23%**.
- **Engineered an internal GenAI product**: Built an AI-powered log analyzer using **Streamlit and LLMs**, which automated ETL troubleshooting and reduced log analysis time by **50%**.
- **Orchestrated 3000+ ELT jobs** across a hybrid ecosystem (Python, Talend, Databricks); led the critical migration of legacy logic from Teradata to Snowflake, utilizing clustering keys to **reduce query execution time by 40%**.
- **Real-time Ingestion & Governance**: Implemented streaming pipelines using **Azure Event Hubs (Kafka compatible)** and Databricks, while defining cross-functional security protocols to ensure strict data compliance.

Principal Data Engineer

Fintech StartUp, Montreal, Canada

08/2024 – 12/2024

- **Architected and delivered a cloud-native financial intelligence platform** on **AWS** in a high-velocity startup environment; took the product from concept to MVP in under 5 months.
- **Orchestrated end-to-end ELT pipelines** using **MageAI** and **dbt**; implemented modular SQL transformations and automated data quality testing to ensure reliable financial reporting.
- **Built real-time data ingestion systems** using **Apache Kafka** to stream high-frequency stock market data, ensuring sub-second latency for investment signal processing.
- **Integrated Generative AI (LLAMA-3)** with **RAG** workflows to analyze unstructured financial news, storing vector embeddings in **PostgreSQL (pgvector)** for semantic search.

Senior Fullstack Developer (DevOps & Integration)

The Boeing Company, Texas, USA

04/2018 – 02/2020

- **Designed and deployed a centralized API Management solution** on **AWS**, facilitating secure data exchange between global enterprise systems using **Java** and **ReactJS**.
- **Standardized data integration patterns** for SAP Hybris; developed RESTful APIs and implemented a log analytics solution using the **ELK Stack (Elasticsearch, Logstash, Kibana)** to monitor API health and data traffic.
- **Led DevOps transformation initiatives** by building robust **CI/CD pipelines** using **Jenkins** and **Ansible**; automated infrastructure provisioning and code deployments, reducing release cycles.
- **Directed integration efforts** between SAP/PI, API Gateway, and Salesforce; engineered XML schema definitions to ensure seamless customer data synchronization across diverse platforms.

Senior Application Developer

Panasonic Vietnam Co. | Hanoi, Vietnam

04/2013 – 10/2017

- **Spearheaded enterprise data integration strategies** using SAP ABAP; engineered robust data flows to ingest and synchronize diverse external sources into SAP systems.
- **Led complex troubleshooting and root cause analysis** for critical data failures; implemented performance tuning resolutions that ensured system stability for business operations.
- **Partnered with business analysts and stakeholders** to translate functional requirements into technical data specifications, ensuring RICEF objects aligned with business analytics needs.

Previous Experience: **SAP Application Developer** (Novellus), **Java Developer** (Pentalog)

EDUCATION

Master of Science - Artificial Intelligence & Machine Learning - John Moores University, Liverpool, UK

October 2025

CERTIFICATIONS

- **Snowflake Pro Core Certification**
- **AWS Certified Data Analytics**
- **Azure Data Engineer Associate**

TECHNICAL SKILLS

- **Cloud & Platforms:** AWS (S3, EC2, Lambda, AWS Glue), Snowflake, Databricks, Azure (ADF, Event Hubs, ADLS).
- **Data Engineering:** dbt, MageAI, Apache Kafka, Spark, Terraform (IaC), Docker, Kubernetes, CI/CD (Jenkins, GitLab).
- **AI & Machine Learning:** Generative AI, LLMs (LLAMA-3), RAG, Vector Databases (pgvector), Streamlit, MLOps.
- **Languages:** Python, SQL, Java, SAP ABAP.