NIEL GABIN

ASPIRING DATA ENGINEER – BUILDING SCALABLE, RELIABLE, AND INSIGHTFUL DATA SYSTEMS

Phone : +33 6 45 06 98 04
Location : Rouen, France
E-mail : gabin.niel.pro@gmail.com
Linkedin : https://www.linkedin.com/in/gabin-niel-749062233/
Diving license

Professional Summary :

Data Engineer passionate about designing, developing, and optimizing end-to-end data pipelines. Experienced with distributed systems, ETL orchestration, and cloud platforms. Strong focus on performance, data quality, and scalability. Excellent communicator with a collaborative mindset, ensuring smooth integration between data and business teams. Open to relocation.

Professional Experience

SKF S2M — Data Engineer (Stage de 6 mois)
(Saint-Marcel, France — Février à Septembre 2025)

- Designed and deployed a Retrieval-Augmented Generation (RAG) pipeline integrating hybrid search (vector-based + lexical) to automatically query technical manuals and electrical cabinet schematics.
- Developed two document extraction microservices (FastAPI + Docker) enabling the automatic conversion of PDF, Word, image, and table files into structured Markdown with HTML handling and Base64 image encoding.
- Implemented a modular multi-layer architecture (Domain, Usecase, Delivery, Repository) in Go and Python to ensure system scalability and maintainability.
- Compared and integrated several embedding models (text-embedding-3-large, mxbai-embed-large) and vector databases (Elasticsearch, Pgvector, Azure AI Search), including performance evaluation (accuracy, latency).
- Set up a hybrid search with reranking, combining semantic similarity and lexical weighting (BM25) to improve answer precision.
- Selected and integrated textual and visual generation models (GPT-4o, Mistral 7B, LLaVA) to produce enriched answers and technical image descriptions (cabinets, components).

Mary Kay Interior — Data Engineer (Stage de 6 mois)
(Bali, Indonésie— Septembre 2025 à Février 2026)

- Designed and deployed a comprehensive data pipeline to centralize and ensure the reliability of business data from multiple internal and external sources.
- Implemented a data warehouse on AWS (Redshift / S3) to guarantee data persistence, scalability, and accessibility for analysts.
- Developed automated ETL/ELT workflows with Apache Airflow, managing the daily extraction, transformation, and loading of structured and semi-structured data.
- Implemented a real-time streaming system using Apache Kafka, enabling continuous data collection and synchronization from sensors and internal applications.
- Integrated SQL and NoSQL databases (PostgreSQL, MongoDB) into the pipeline, with schema management and indexing.

Skills

Hard Skills

Data Engineering
- ETL / ELT Pipelines (Apache Airflow, AWS Glue) · Data Warehousing (AWS Redshift) · Data Lake (AWS S3) · Real-time Streaming (Apache Kafka) · SQL & NoSQL Integration (PostgreSQL, MongoDB)

Cloud & DevOps
- AWS (S3, Lambda, Redshift, IAM) · Docker · CI/CD (GitLab) · Linux · Bash

Programming
- Python (Pandas, NumPy, FastAPI, Pydantic, Requests, Logging) · Go (Golang) · Advanced SQL · Java (Spring Boot)

Machine Learning & NLP
- Embeddings (FAISS, BM25, OpenAI API) · Regression / Classification · Clustering · Deep Learning (TensorFlow, PyTorch)

Data Modeling & Quality
- Data Modeling · Data Quality · Normalization · Indexing · Data Validation

Visualization
- Matplotlib · Seaborn

Education
- ESIGELEC – Graduate School of Engineering
- Master of Science in Engineering – Data Engineering & Artificial Intelligence
  Saint-Étienne-du-Rouvray, France | September 2020 – September 2025

Languages : French (Native) - English (Fluent) - Spanish (Good-Level)

CREATIVE PORTFOLIO

or click here