



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)**

- Conçu et déployé une pipeline RAG (Retrieval-Augmented Generation) intégrant une recherche hybride (vectorielle + lexicale) pour interroger automatiquement des manuels techniques et plans d'armoires électriques.
- Développé deux microservices d'extraction documentaire (FastAPI + Docker) permettant la conversion automatique de fichiers PDF, Word, images et tableaux en Markdown structuré avec gestion du HTML et encodage des images en Base64.
- Implémenté une architecture modulaire multi-couches (Domain, Usecase, Delivery, Repository) en Go et Python pour assurer la scalabilité et la maintenabilité du système.
- Comparé et intégré plusieurs modèles d'embedding (text-embedding-3-large, mxbai-embed-large) et bases vectorielles (Elasticsearch, Pgvector, Azure AI Search), avec évaluation des performances (précision, latence).
- Mis en place une recherche hybride avec reranking, combinant similarité sémantique et pondération lexicale (BM25) pour améliorer la précision des réponses.
- Sélectionné et intégré des modèles de génération textuelle et visuelle (GPT-4o, Mistral 7B, LLaVA) pour produire des réponses enrichies et des descriptions d'images techniques (armoires, composants)..

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

- Conçu et déployé une pipeline de données complète pour centraliser et fiabiliser les données métiers issues de plusieurs sources internes et externes.
- Mis en place un data warehouse sur AWS (Redshift / S3) afin d'assurer la persistance, la scalabilité et l'accessibilité des données pour les analystes.
- Développé des workflows ETL/ELT automatisés avec Apache Airflow, gérant l'extraction, la transformation et le chargement quotidien des données structurées et semi-structurées.
- Implémenté un système de streaming temps réel avec Apache Kafka, permettant la collecte et la synchronisation continue de données issues de capteurs et d'applications internes.
- Intégré des bases SQL et NoSQL (PostgreSQL, MongoDB) dans la pipeline, avec gestion des schémas et indexation

Hard Skills

Data Engineering

- ETL / ELT Pipelines (Apache Airflow, AWS Glue) • Data Warehousing (AWS Redshift) • Data Lake (AWS S3) • Streaming temps réel (Apache Kafka) • Intégration SQL & NoSQL (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) • SQL avancé • 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 • Normalisation • Indexation • Data Validation

Visualization

- Matplotlib • Seaborn

Skills

Education

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

CREATIVE PORTFOLIO



[cliquez sur ce lien](#)

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