

Camille Lebrun

Data & IA engineer, consultant chez Théo : je bâtis et fiabilise vos plateformes data

Toulouse, France | <https://camillelebrun.fr> | <https://www.linkedin.com/in/camille-lebrun-ae>

Langues : Français (natif), Anglais professionnel courant | 4+ ans d'expérience, 19 missions

PROFIL

Data & IA engineer chez Théo, à Toulouse. Je migre, structure et fiabilise des plateformes data (dbt, BigQuery, Terraform, Python) pour que les équipes puissent s'appuyer sur leurs chiffres, y compris quand l'IA arrive : migration de Trino vers BigQuery de 1 000+ modèles dbt (95 %+ automatisés), data contracts contrôlés en CI, RAG sur 179 sources.

COMPÉTENCES CLÉS

- Plateformes data modernes et multi-cloud** : Trino vers BigQuery sur S3NS (1 000+ modèles dbt), BigQuery vers DuckDB sur Scaleway (Plein Data, dont DuckDB WASM), Snowflake.
- Analytics engineering** : Standards dbt livrés à des équipes clientes, macros, tests automatisés, CI/CD ; modélisation Medallion (Bronze, Silver, Gold).
- Data contracts et gouvernance** : Data contracts générés depuis les métadonnées et contrôlés en CI, adoptés par plus de 5 data engineers ; package d'observabilité (5 piliers).
- Pipelines ELT et orchestration** : Ingestion SAP, GA4, API, Airbyte et Funnel vers BigQuery ; orchestration Dagster, Kestra, Airflow.
- IA générative et RAG** : RAG en ligne sur 179 sources, évalué sur 52 questions ; veille LLM à sortie JSON validée.
- Dataviz, self-service et sécurité** : Dashboards pour la direction et les équipes métiers (Looker Studio, Power BI, Metabase), accès par Row-Level Security.

EXPÉRIENCE PROFESSIONNELLE

Data Consultant (data engineering) | Théo, Toulouse | févr. 2026 – aujourd'hui

- Grand groupe industriel : migration de plus de 1 000 modèles dbt de Trino vers BigQuery (cloud de confiance), automatisée à plus de 95 % par un CLI déterministe, équivalence vérifiée sur les modèles comparés.
- Grand groupe industriel : outillage de data contracts généré depuis les métadonnées SAP, contrôlé en CI, adopté par plus de 5 data engineers.
- Grand groupe industriel : package dbt d'observabilité (freshness, volume, schema, lineage) compatible Trino et BigQuery.
- Projet interne RH : dashboard Power BI des visites de médecine du travail et entretiens professionnels, avec Row-Level Security.

Analytics Engineer Consultant | Hanalytics, Paris | janv. 2025 – déc. 2025

- 9 missions de conseil menées en autonomie, le plus souvent en solo, du recueil des besoins à la livraison, avec 5 h gagnées par semaine sur chaque mission, chez des clients du retail, de l'e-commerce, de la mode, de la santé, du média, de l'édition, de la distribution et du tourisme.
- Marque santé et cosmétique e-commerce : traitements manuels remplacés par des pipelines ELT automatisés (Funnel et Airbyte vers BigQuery, standards dbt).
- Distributeur industriel : plateforme data de bout en bout (SAP, Stitch, Airbyte et Funnel vers BigQuery, dbt, Kestra), référent technique de l'intégration.
- Agence média : infrastructure legacy remplacée par BigQuery, dbt et Kestra, données de performance média collectées automatiquement depuis les DSP, équipes clientes formées à dbt.
- Réseau multi-magasins : un dashboard par magasin (accès par lien unique, Row-Level Security) et un dashboard direction, avec alerting Slack automatique sur données absentes ou anomalies.

Analytics Engineer (alternance) | Lucky Cart, Paris | nov. 2023 – déc. 2024

- MarTech du média e-retail : audit et monitoring de la data quality, application Streamlit qui génère les reportings retailers : 5 h gagnées par semaine.

Analytics Engineer (stage) | My Job Glasses, Paris | juin 2023 – nov. 2023

- EdTech : data marts multi-secteurs sur Snowflake avec dbt, dashboards Metabase pour les clients.

Data Analyst (stage) | Click & Boat, Paris | mai 2022 – oct. 2022

- Datasets spécialisés pour les équipes marketing et presse, après recueil des besoins marketing et sales.

PROJETS PERSONNELS

- Plein Data : plateforme de prix de carburants sur Scaleway** : Migration de BigQuery vers une architecture serverless souveraine : 55 M de lignes en Parquet, dbt-duckdb, 11 Jobs orchestrés, tout en Terraform.
- Berlio : aide au choix d'alimentation pour chats atteints d'IRC** : RAG sur 179 sources vétérinaires, évalué sur 52 questions : comportement correct dans 90 % des cas sur les sujets couverts, 100 % sur les questions adversariales.
- StackRadar : veille Modern Data Stack avec digest quotidien** : Veille quotidienne sur 19 sources, changelogs analysés par LLM avec sortie JSON validée, déployée sur Cloud Run.

OUTILS ET TECHNOLOGIES

- Cloud** : Google Cloud Platform (BigQuery, Cloud Run, GCS, Dataform, Dataflow), Snowflake, Trino, Scaleway, S3NS
- Transformation** : dbt Core et Cloud, SQL avancé, Python (FastAPI, Pandas, Streamlit, sqlglot)
- Orchestration et DevOps** : Dagster, Kestra, Apache Airflow, Docker, Terraform, GitLab CI/CD, Git
- Gouvernance** : Open Data Contract Standard (ODCS), Data contracts, FinOps BigQuery, Alerting Slack API
- Ingestion** : Funnel, Airbyte, Stitch, API REST, Couchdrop, Catchr
- IA** : Mistral AI, Groq, Upstash Vector, Embeddings, RAG
- Dataviz** : Looker Studio, Power BI, Tableau, Metabase

FORMATION ET CERTIFICATIONS

2026 : PL-300 Microsoft Power BI Data Analyst Associate | 2025 : Partner Certification Program Level 1, Funnel | 2024 : Master Data et IA, HETIC | 2020 : MSc Structural and Functional Biochemistry, Université de Toulouse