

Clement Demonchy

cdw@cdemonchy.com

Senior data engineer and data architect freelance

clement-demonchy-55327778

cdemonchy.com

knil-sama

I have more than 10 years of experience working as a data engineer mainly in **Python** and **AWS**, assuming devops, architect and lead roles when needed.

I prefer working on missions where I can focus on workflows that bring value while making sure that the tradeoff (cost, reliability, performance) are understood and managed.

I appreciate when I can interact with the infra and product team directly.

I am most proficient at: **AWS**, **Python**, **PostgreSQL**, **Kubernetes** and **Databricks** I have experiences on **Azure**, **GCP**, **Snowflake** as well as **Scala**, **Neo4j** and **MongoDB** I like to dabble in **Rust** and **TLA**

TECHNICAL EXPERIENCE

Data Engineer

2026-03-19 — 2026-12-31

Leboncoin

Levallois-Perret

In a team responsible for the migration of all datasets to **Databricks**

- Identify and extract all sql codes from **Airflow** pipeline and notebook for analysis using Lakebridge
- Export billing informations from **Databricks** system tables to monitor cost
- Create dashboard and monitor in **Datadog** based on pipeline metrics

Data Engineer

2025-04-02 — 2025-10-07

Doctolib

Levallois-Perret

In a team responsible for exposition and anonymization of production tables

- Change Data Capture of relational database to **Iceberg** table using **Debezium**
- Create mirror tables with **Pyspark** job on **EMR** serverless
- Create and maintain dataset on google cloud using **Crossplane** and **Kopf**
- Install resources on **GKE** using **Argocd**

Data Engineer

2023-10-01 — 2025-03-31

Be:mo (at TotalEnergies)

Paris 2ème arrondissement

In a team responsible for working with electric vehicle supply equipment (EVSE) and exposing it with **REST API** for multiple BtB companies

- Migration of data workflow from **Azure** synapse to **AWS Databricks**
- Writing CI job using **Github Actions** and **Databricks** bundle
- Using unity catalog to provide external access to dataset for services like PowerBi
- Usage of DLT table for streaming dataset
- Setting test for spark on dev machines to improve feedback
- Managing **AWS** and **Databricks** resources using **Terraform** (tagging, creation, deletion)
- Define alerts based on **Lambda** metrics using **Cloudwatch**
- Expose metrics of data workflow with **PostgreSQL Aurora** and **Grafana**
- Documentation on workflow using Mermaid on **Github**
- Analyzing and reducing cost on **Azure** and **AWS**, for example by dividing cost by 5 for **AWS Lambda** exploiting caching between service and improving batching of SQS elements
- Optimize geolocalization operation on **PostgreSQL** and detect invalid coordinates

Lead data Engineer

2022-11-01 — 2023-07-01

Skillup

Paris 10ème arrondissement

I was responsible for a 4 person team of 3 Data engineers and 1 ML engineer at Skillup, an SaaS helping RH in the company to provide a better experience providing training for their employees and evaluating impact.

The team had a broad scope covering 4 principle subjects

- External data flux from customer with client import (user, booked training, ...) and export (training evaluation)
- Scrapping and cleaning training website for our catalog
- Provide internal metrics and dashboard so products team can gain insights
- Use latest AI technologies (LLM) to implement new feature for customer

My first task when starting here was to put in place a CI/CD using **Github Actions** to structure coding practice with automated testing and linting.

All the while implementing team's rituals (postmortem, retrospective, daily, ownership rotation) as well as mentoring and pairing to share best coding practice (**Ruff**, **Pydantic**, **Pytest**, ...).

This allowed us to refactor existing code more easily and made the team more resilient with people being able to work on other team member code.

The scrapping project required migration of **Airflow** orchestrated data pipelines that use Zyte for scrapping online training websites and extracting data using NLP models.

The frontend of Skillup being developed in Typescript we rely on GRPC **API** to interact with their system.
New services developed to allow complex importation of customer data were an Fastapi REST **API** to query **Neo4j**, **MongoDB** and **PostgreSQL** databases to reconcile existing data.
We then provided a website in intern developed in Dash for the product team in case a manual validation was required.
For the dashboarding we wanted to give a lot of autonomy for no tech savvy user and deployed both Metabase and Superset.
For the IA related project we needed to do some optimisation of OpenIA POC for generation of job's skills based on role title

Data Engineer

2020-03-01 — 2022-10-01

Talend

Suresnes

I was the first data engineer of a growing team of data scientists (5 to 7 people) that worked exclusively on state of the art machine learning for customer and internal needs.

We got a separate **AWS** account and a dedicated kubernetes namespace for our dev environment, as we were the first to do machine learning at Talend so a lot of new processes needed to be put in place.

In this context the first months were used to document and set guidelines for coding, testing, packaging and deployment using **Github Actions** and Jenkins.

We experimented with multiple platforms to deliver our solutions

- Data pipeline on **Databricks** with **Scala** and **Python** (**Pyspark** and **Koala**) to feed ML prediction models
- Deployment and optimisation of models to detect anomalies in customer's jobs in **Python** and **Flask**
- Extraction using Talend studio job(Java) to dump **PostgreSQL** database on **S3** bucket as parquet
- Extraction and computation on a **Snowflake** datawarehouse

We finally settled on Sagemaker for most of our projects and my role included finops responsibilities

- Managing **AWS** account resources using terraform (tagging, creation, deletion)
- Define with team's lead budget alerts and reports by slack/email with related dashboards to provide context
- Evaluation and optimisation regarding architecture of Sagemaker pipeline, model and endpoint on **AWS**

Other responsibilities included making people more aware of security with presentation, training on secureflag platform and setting audit tools in the CI with veracode.

Data Engineer

2017-08-01 — 2019-10-01

Jobteaser

Paris 9ème arrondissement

I joined Jobteaser as the second data engineer of a 5 person team that provided data analysis and machine learning for the whole company.

The company was moving a startup to a scaleup so a lot of projects were doing migration from the legacy system to the new system.

-Migrate Elastic Beanstalk REST **API** on **AWS** to **Kubernetes**.

-Rewrite/deploy job classification REST **API** in **Flask**.

-Write/deploy ETL and REST **API** for job recommendation project.

Main project beside that was deploying the backbone of our data stack with a **Kafka** that allowed us to stream our production database (MySQL) to our Postgresql using **Debezium** and change data capture patterns.

We then loaded it into our datawarehouse (**Redshift**) that give data analyst opportunity to do realtime dashboard of backend activity

Data Engineer

2016-09-01 — 2017-08-01

Ipsen

Issy Les Moulineaux

Alongside another data engineer we exploring multiple projects to resolve issue encountered by marketing and research teams sharing progress during spring (3 weeks) for Ipsen an pharmaceutical company

The two most advanced projects are listed below and all of them were hosted on **AWS EC2** using docker-compose.

-Team's responsible for handling clinical trials of new drugs needed insight on previous research done by other companies in the whole world, so we did data mining on official reports for specific diseases, cross referenced the sources then displayed it on an interactive map using Folium, we ended up pushing some fix to folium.

-Marketing wanted to get feedback regarding new medicines on social media so we targeted tweets related to either the disease or drug, developed an algorithm to identify key opinion leaders (either medical professional or public personality) then used nlp tools to do sentiment analysis.

Data Engineer

2015-09-01 — 2016-08-30

Engie

La Défense

At Engie I worked on 2 projects with them

PAP:

With another data engineer and an app team in order to provide a solution to traveling salesmen that need to plan their road, get customer information and sign contracts on their tablets.

We had a huge focus on geolocation and we had to put in place
-Mongodb to store customer's data (60 millions users)
-REST API written in Java for the application team
-Fix customer addresses using google maps data and apache Lucene
This project was successfully deployed and is still in production.
GDF 360:

To help desk support of Engie to provide better service to customers we implemented a POC in 6 months that displayed all known previous interactions between a customer and the company (call, bills, contract, ...) using a graph database (**Neo4j**) on a temporal axis.

EDUCATION

Master's degree in Computer Science, University Of Cergy-Pontoise (UCP)

2013-09-01 - 2015-06-30