

ROMAIN JOUHAMEAU

 Français - 34 ans - Paris  06 52 01 60 70  jouhameau.romain@gmail.com  LinkedIn  github.com/RomainEconomics  Site

Passionné par l'informatique et fort d'une solide formation en économie, j'ai su orienter ma carrière vers le Data Engineering et la Data Science, avec une spécialisation récente en LLM et systèmes RAG. Aujourd'hui, je mets à profit mes compétences pour automatiser l'extraction de données, créer des API robustes et optimiser les processus de déploiement grâce à des technologies comme Docker et Kubernetes. Bien que Python soit mon langage principal, je suis également ouvert à travailler avec d'autres langages tels que Typescript, Java, Rust, ou Go, afin de répondre aux besoins spécifiques des projets sur lesquels je travaille.

COMPÉTENCES TECHNIQUES

Conteneurisation Docker, Kubernetes, ArgoCD

Bases de Données PostgreSQL, Elastic Search, Redis, Weaviate

Machine Learning LLM, RAG, Weaviate, MLOps

Langages Python, Bash, Typescript, Go, Java

Tests Pytest, Docker, Gitlab CI/CD

Cloud AWS, Azure, OVH

EXPÉRIENCES RÉCENTES

Iceberg Data Lab

Nov 2022 – Aujourd'hui

Senior Data Scientist - Data Engineer

Automatisation de l'extraction de données ESG avec LLM et RAG

- Conception et déploiement d'une solution RAG (Retrieval-Augmented Generation) combinant LLM et la base de données vectorielle Weaviate pour l'extraction automatisée de données à partir de rapports semi-structurés (PDF).
- Mise en place d'un système d'évaluation pour monitorer et garantir la qualité des extractions.
- Développement d'une API REST avec FastAPI pour exposer les données structurées stockées dans PostgreSQL, facilitant l'intégration avec les systèmes existants.
- Responsabilité du déploiement (Docker/Kubernetes) et du monitoring en production, avec gestion des logs et métriques.
- Automatisation complète du processus d'extraction qui était auparavant manuel, réduisant les délais de traitement.

Migration vers Kubernetes

- J'ai mené le déploiement des services de mon équipe situés sur des VMs vers des clusters Kubernetes, notamment via ArgoCD pour une meilleure gestion du déploiement continu.
- Déploiement et gestion d'Apache Airflow sur Kubernetes, permettant l'automatisation de nombreux jobs (ETL, backups).

Crawler distribué pour l'extraction de rapports annuels

- Mise en place d'un crawler distribué pour l'extraction de rapports annuels (PDF principalement) des entreprises cotées.
- Cette application a permis de collecter efficacement des données à grande échelle.

Développement de packages Python internes

- Utilisation de UV pour uniformiser la création et le déploiement des package sur le Gitlab Package Registry interne.
- Ces packages ont permis d'uniformiser le code utilisé en production et de réduire le temps de mise en production.

CEPREMAP

Sep 2021 – Juin 2022

Assistant de Recherche

- Assistant de recherche auprès de Thomas Renault dans le cadre de la rédaction d'un opuscule, "Le Bonheur est sur Twitter : Un baromètre du moral des Français", pour l'Observatoire du Bien-être.
- Création d'un indicateur de bien-être avancé à partir de données Twitter (Django, MongoDB)

FORMATION

IPSSI

Oct 2022 – Oct 2023

M2 Dev, Data, IA

Matières Principales: Machine Learning, Deep Learning, PySpark, Hadoop, Web Scraping

Mémoire: L'évolution du traitement des données ESG avec le développement de l'IA et des Large Language Models

Université Paris 1 Panthéon-Sorbonne

Sep 2021 – Juin 2022

M2 Financial Economics

Matières Principales: Économétrie, Séries Temporelles, Asset Pricing

Paris School of Economics (PSE)

Sep 2020 – Juin 2021

M1 Analyse et Politique Economiques (APE)

Matières Principales: Économétrie, Séries Temporelles, Macro-économie, Micro-économie, R

Université Paris 1 Panthéon-Sorbonne

2017 - 2020

Licence Économie

Matières Principales: Mathématique, Statistiques, Macro-économie, Micro-économie

Certifications

- Azure: Azure Data Fundamentals

2021