



Marc Giannuzzi, Data Engineer

(+33) 6 95 64 64 54, marc.giannuzzi@gmail.com

LIENS

[GitHub](#)
[Linkedin](#)

EXPERIENCE

Sept. 2025 — Aujourd'hui

Waiyer — Auto entrepreneur

Paris, FR

Contexte : Waiyer est une entreprise spécialisée dans le conseil en data et en architecture cloud. Waiyer permet d'optimiser la gestion des données grâce à des solutions sur mesure, en utilisant les dernières technologies. Le but est de concevoir et à mettre en œuvre des solutions efficaces et pérennes pour aider les entreprises à tirer le meilleur parti de leurs données.

- Mise en place de l'architecture des données
- Orchestration de pipelines de données
- Création de dashboards analytiques
- Business Intelligence
- Software engineering : développement de scripts, APIs, outils d'automatisation
- Vision stratégique : conseils sur les choix technologiques, évolutions de l'architecture, roadmap technique
- Vision fonctionnelle : collaboration avec les équipes métier pour définir les besoins, les KPIs, les dashboards adaptés
- Veille technologique : suivi des évolutions dans le domaine de la data, identification des opportunités d'innovation pour les clients

Compétences : Kubernetes · GCP · GitLab CI/CD · Python · AWS S3 · SQL · MongoDB · PostgreSQL · Teradata · PowerBI · Elastic Stack (ELK)

Avr. 2025 — Sept. 2025

LCL [SFEIR] — Data Architect

Villejuif, FR

Contexte : LCL (Le Crédit Lyonnais) est une banque française majeure, filiale du groupe Crédit Agricole, qui propose des services bancaires aux particuliers, professionnels et entreprises. Optimisation de l'organisation interne de l'entité SNI et mise en place des tableaux de bord analytiques pour son CODIR. L'enjeu était de consolider les données issues de différentes sources, orchestrer et automatiser leur traitement via une architecture scalable et maintenable.

- Audit du SI existant : analyse des stacks en place (bases, scripts, outils), comparaison PostgreSQL, MongoDB, Teradata
- Architecture technique hybride (on-premise first, comparaison et propositions d'équivalents Cloud GCP)
- Orchestration Kubernetes de scripts Python dockerisés.
- Stockage via PostgreSQL (mode analytique) et S3
- Design des schémas de données (tables, colonnes, index).
- Sélection des bibliothèques Python (pandas, SQLAlchemy)
- Intégration des API Jira et autres progiciels dans les pipelines
- Dataviz : dashboards Power BI dynamiques pour le CODIR.
- Encadrement technique : collaboration avec 2 data engineers et 1 PO, revue technique et support à l'implémentation Bonnes pratiques, guides centralisés dans Confluence (architecture, code)

Compétences : Kubernetes · GCP · GitLab CI/CD · Python · AWS S3 · SQL · MongoDB · PostgreSQL · Teradata · PowerBI · Elastic Stack (ELK)

Contexte : SFEIR est une entreprise française, premier partenaire Google Cloud Platform GCP (Services 2023, Training 2024, EMEA 2025), spécialisée dans le développement de logiciels et l'accompagnement technologique des entreprises. Elle aide ses clients à créer des applications web, cloud, mobile ou data en utilisant des technologies modernes. En charge de **recrutements** de profils techniques:

- Data Engineers
- Data Scientists
- ML Ops
- Devops
- Fullstack

Compétences : GCP (Google Cloud Platform)

Contexte : Engie est le premier fournisseur de gaz naturel et d'électricité verte en France, ainsi qu'un groupe mondial dans le secteur de l'énergie.

Data Engineer faisant partie de l'équipe agile (5 personnes) développant les produit Mon Bonus Engie / Ecodefi+, ICE (Impact Carbone Elec) et Ecogestes Dynamiques évoluant dans le framework SAFe

- Ecodefi+ (Mon Bonus Engie) est un service gratuit proposant aux particuliers d'être récompensé sur leur sobriété énergétique lors de challenges (**300k clients**)
- ICE (Impact Carbone Elec) permet aux clients d'**analyser les pics de consommation carbone dans le futur et d'adapter leur consommation** en fonction (**1M clients**).
- Ecogestes Dynamiques contribue à expliquer les impacts des objets électriques dans les foyers (**1M clients**)

Missions

- Implémentation et développement des produits
- Gestion des données à grande échelle (bases des données, orchestration des flux...)
- Elaboration des KPIs techniques et fonctionnelles
- Monitoring et Alerting des solutions
- Création d'APIs REST (100k appels/jour), latence < 1sec
- Participation à l'architecture des micro-services AWS et des plateformes digitales
- Optimisations (mémoire, temps d'exécution, refactoring...)
- Optimisations des flux bases de données (structure, index...)
- Reporting automatisé via SNS
- Gitlab CICD
- Rédaction de roadmaps dans le cadre de la migration d'une technologie centrale utilisée depuis plusieurs années au sein du train SAFe
- Participations aux architectures technico-fonctionnelles
- Scaling d'algorithmes de Deep Learning (DeepAR, Temporal Fusion Transformer...)

Compétences : c3ai · JavaScript · Jenkins · AWS Lambda · Amazon Kinesis · Databricks · Elastic Stack (ELK) · Python (langage de programmation) · AWS S3 · Terraform · Amazon CloudWatch · Splunk · MLflow · Amazon DynamoDB · TIDB · Apache Airflow · Amazon Simple Queue Service (SQS) · Amazon Simple Notification Service (SNS)

Contexte : Visian est une entreprise de **conseil** spécialisée dans les projets utilisant la donnée à grande échelle. Ma présentation vidéo de l'entreprise et du métier: Youtube

- Recrutements de profils techniques (Data Ingénieurs, Data Scientists, Devops, Fullstack)
- **Direction du projet Pysquad**, développant en interne des projets digitaux
- Coordination de l'implémentation générique d'un **projet d'Intelligence Artificielle (IA)** générative pour différentes tâches d'inférence (textual, vision, multimodal etc...) avec l'utilisation des services ML/DL Azure et Azure AI Studio
- **Création d'une data plateforme** en vue d'aider les sales à récupérer les données de missions, prospects, clients, consultants etc

Compétences : Docker · Azure AI Studio · Django · Streamlit · PowerBI · FastAPI · Elastic Stack (ELK) · AWS RDS · Redis · AWS Code Artifacts

Contexte : Aqua Digital Rising est une **startup fintech** qui ambitionne de révolutionner le pari sportif en analysant toutes les statistiques des joueurs et des équipes.

Gestion de la donnée

- Gestion du stockage (création, structuration, nettoyage) de plusieurs bases de données (MongoDB, SQL server, HDF5) à partir de données récupérées par les fournisseurs Sport Radar et Opta
- Structuration des flux de données en Python à travers les différents composants de l'architecture et optimisation de la vitesse de lecture grâce aux index SQL Server (plusieurs centaines de Go)
- Création d'une API REST hautement configurable permettant à l'équipe front-end de récupérer les agrégations de la base de données et d'autres informations statiques
- Sauvegardes temporaires des résultats de requêtes de l'API grâce à Redis
- Mise à jour en direct des données statistiques avec l'utilisation de RabbitMQ
- Analyses et data visualisation
- Algorithmes de machine learning (clustering) et visualisations pour l'analyse des statistiques des joueurs et des équipes

Autres missions

- Configuration de divers services Azure (Azure Functions, Azure Web Services, Azure Artifacts, Azure Pipelines, Azure Keyvault) en lien avec la stratégie Cloud de l'entreprise
- Analyses quantitatives de crypto-monnaies avec InfluxDb pour alimenter la stratégie d'investissement
- Coordination de 2 ingénieurs (un data scientist et un développeur front-end)

Compétences : Microsoft SQL Server · Scikit-learn · Redis · JIRA · RabbitMQ · Bases de données · InfluxDB · Bash · MongoDB · Microsoft Azure

Contexte : Orange Business Services est une société de **conseil** dans la transformation digitale dans des projets de communication.

- Analyses et diagnostics des anomalies du réseau VoIP, notamment avec du machine learning (clustering)
- Création d'un graphe de réseau VoIP dynamique avec Kibana puis Neo4J pour une visualisation plus facile, intégré dans une page web
- Analyses des métriques Devops (utilisation CPU, stockage...) d'un cluster de serveurs Elastic
- Automatisation de la suppression des logs Elastic avec log4j

Note: VoIP = Voice over Internet Protocol est un réseau qui permet les appels téléphoniques via Internet

Compétences : HTML · Scikit-learn · Elastic Stack (ELK) · Feuilles de style en cascade (CSS) · Python (langage de programmation) · Neo4j · JavaScript · Bases de données · Bash

Contexte : Digiberries est une **agence de référencement web**, améliorant les stratégies SEO des clients.

- Développement d'un script permettant l'automatisation de création de rapports excel à partir de données Google Ads
- Création d'un outil d'analyse SEO et de création de rapport automatisé PDF avec Lighthouse et des packages NodeJS
- Automatisation d'envoi de mails

Compétences : Lighthouse · Python (langage de programmation) · Google Ads · Node.js

Contexte : Eudonet développe des logiciels CRM dans un contexte B2B.

- Mise en place de la suite Elastic (ElasticSearch, Kibana, Logstash + Filebeat) pour l'analyse de fichiers logs Windows IIS
- Création d'un outil de monitoring avec des dashboards dynamiques dans Kibana
- Création d'un algorithme Python pour la détection de baisse d'utilisation du CRM des clients. Utilisation de la librairie FBProphet
- Création d'une plateforme web (Javascript, Node.js) dynamique permettant une visualisation des résultats de l'algorithme précédent
- Visualisations PowerBI pour la gestion interne des projets des employés
- Création de scripts powershell pour la gestion d'applications IIS
- Requetes SQL pour la création de rapports sur l'utilisation des clients et la modification de paramètres
- Sprints agiles hebdomadaires

Compétences : Microsoft SQL Server · HTML · Windows PowerShell · Elastic Stack (ELK) · Feuilles de style en cascade (CSS) · Python (langage de programmation) · Microsoft Power BI · JavaScript · Bases de données · Bash

Contexte : Veesion est une scale-up française voulant **détecter automatiquement les vols** dans des supermarchés.

- Création d'une plateforme interne permettant de visualiser et labeliser (afin d'améliorer l'algorithme de détection de vols de l'entreprise) en direct des vidéos avec Django
- Classification de données permettant l'apprentissage du modèle de computer vision

Compétences : HTML · Feuilles de style en cascade (CSS) · Python (langage de programmation) · JavaScript · Amazon S3 · Django

PROJETS

2024	Intelligence Artificielle
	Création automatisée de contenus digitaux multimodaux
2024	Finance
	Algorithmes et visualisations d'aide à l'analyse financière (Python)
2021	Algorithmes de tri
	Plateforme de visualisation d'algorithmes de tri (React)
2020	NLP (Natural Language Processing)
	Agent conversationnel affichant les paroles d'une chanson (Chatbot Messenger)
2020	Computer vision
	Algorithme de computer vision transformant les gestes du doigt en équations via une webcam (Tensorflow, OpenCV)
2019	Deep Learning
	Algorithme de scoring santé de recettes cuisine (Tensorflow)

CERTIFICATIONS

2025	GCP - Associate Cloud Engineer
	https://www.credly.com/badges/69fe0262-094b-4cee-82a8-a70cd10ec091
2025	GCP - Professional Data Engineer
	https://www.credly.com/badges/888d00d8-5399-4f4c-a4ce-04d05901344a

EDUCATION

Sep. 2018 — Juin 2021	ESILV (École Supérieure d'Ingénieurs Léonard De Vinci)	Courbevoie, FR
	<ul style="list-style-type: none">• Master en Informatique, Big Data et Objets Connectés• Spécialisation en Data Science	
Sep. 2019 — Jan. 2020	UCR (University Of California Riverside)	Riverside, California USA
	<ul style="list-style-type: none">• <u>CS 235</u> Data Mining Techniques• <u>CS 171</u> Introduction to Machine Learning and Data Mining• CS 009 Python programming	

Sep. 2016 — Déc. 2018

Classe Préparatoire aux Grandes Écoles (CPGE) Fénelon Sainte-Marie Paris,
FR

- MPSI
- MP

LANGAGES

English: Fluent

Français: Natif