



Walid Karray

Architecte Solutions hands-on freelance · AWS · GenAI

✉ walid.karray@gmail.com 📍 Paris, France · Full remote possible linkedin.com/in/karray
 github.com/tux86 📄 medium.com/@walid.karray

PROFIL

Architecte solutions **hands-on** avec une **culture produit** : je conçois les architectures **et je les construis**, de l'idée à la production, et je les juge aux résultats utilisateurs et à la valeur métier. En **16+ ans**, je l'ai fait des deux côtés de la table : en conseil pour des grands groupes, et comme CTO co-fondateur d'une startup portée à **1M+ d'utilisateurs** (billetterie officielle de la **CAN 2023**). Aujourd'hui je livre de **l'IA générative qui exécute de vrais processus métier** (systèmes multi-agents, RAG et orchestration LLM en production à l'échelle de l'entreprise), sur des fondations solides en **serverless AWS**, architecture event-driven et IaC. Certifié AWS Solutions Architect. Disponible pour des missions freelance.

COMPÉTENCES CLÉS

Cloud & AWS	Serverless, Lambda, API Gateway, EventBridge, SQS, AppSync / GraphQL, Cognito, Aurora PostgreSQL, architecture event-driven
IA générative	AWS Bedrock, orchestration multi-agents, RAG, orchestration LLM, OpenRouter, Text-to-SQL, AWS Textract, vector search (MongoDB Atlas)
Langages & runtimes	TypeScript, Node.js, Bun, Python
IaC & DevOps	Terraform, CI/CD (GitHub Actions), Docker, LXC, Proxmox VE / PBS, security scanning (Checkmarx, Semgrep)

EXPÉRIENCE

- TotalEnergies** Architecte Solutions hands-on

Nov. 2023 – aujourd'hui · Île-de-France · Full remote

 - **DAIA** : pipeline GenAI (**modèles GPT**, plateforme LLM interne, intégré à **SAP Datasphere**) qui contrôle chaque demande d'achat au regard de la politique achats et approuve automatiquement les demandes conformes, supprimant la revue manuelle, soit **~2 M€/an d'économies**.
 - **Synaps** : assistant GenAI d'entreprise avec **orchestration multi-agents**, dont un agent autonome pilotant **SAP sans intervention manuelle**, et **l'interrogation en langage naturel de 5 bases de données** via RAG.
 - Les deux produits tournent sur la **plateforme serverless multi-tenant que j'ai conçue** (Lambda, API Gateway, EventBridge, SQS) : backend TypeScript, moteur d'exécution Python, **50+ step handlers** réutilisés à l'échelle du groupe.
 - Création du pipeline CI/CD des applications **SAP Fiori/UI5** : build et déploiement automatisés, intégrés à **ServiceNow**.
 - Tests E2E couvrant multi-tenancy, RBAC et partage de ressources ; CI/CD (GitHub Actions) avec security scanning bloquant (**Checkmarx, Semgrep**).
- SNCF Connect & Tech** Ingénieur Cloud · AWS Serverless

Sept. 2022 – Nov. 2023 · Paris · Hybride

 - Participation au développement du système de diffusion audiovisuelle nouvelle génération des gares, en serverless AWS temps réel : déployé dans **toutes les gares SNCF sans interruption de service**, touchant **21M+ de voyageurs**.
 - Conception de la stratégie de tests (unitaires, intégration, automatisés, charge) et des critères d'acceptation imposés avec l'équipe QA : le socle du déploiement sans coupure.
- Maxeon Solar Technologies** Architecte Solutions AWS

Juil. 2022 – Sept. 2022 · Full remote

 - Mission d'architecture pour un fabricant mondial de panneaux solaires : **migration AWS serverless-first** de la plateforme servant **5 marchés internationaux**, réduisant les coûts d'infrastructure et débloquent le passage à l'échelle.
 - Unification d'**Okta SSO, Salesforce CRM, AlphaESS** et des données énergie **IoT Net2Grid** en une plateforme unique, aux flux de données fluides entre systèmes.

- Plateforme SaaS de billetterie sportive en marque blanche (B2B2C), construite de zéro comme CTO co-fondateur en méthode **Lean Startup** : **1M+ d'utilisateurs** et **10M+ de billets vendus** en trois ans.
- Architecture de la stack serverless AWS, conçue pour auto-scaler d'une organisation cliente à l'autre.
- Billetterie officielle de la **CAN 2023** : la plateforme a tenu **350K connexions en pic journalier** et 25K billets vendus par jour.
- Levée de fonds seed, équipe d'ingénierie passée de **2 à 20**, et **40+ organisations clientes** (Betclic Group, AC Ajaccio, CAF).

- Participation au développement du CMS éditorial SaaS **adopté par 4 rédactions internationales** (France 24, RFI, MCD, Observers) publiant en **22 langues**.
- Développement backend d'**InfoMigrants** (FMM × Deutsche Welle × ANSA, co-financé par l'UE) : information vérifiée à destination des migrants, **6 langues**, mobile-first, des millions de lecteurs.

- Livraison de **l'API REST et du socle microservices** des plateformes numériques, avec des PoC (SonarQube, MongoDB) qui ont amorcé le passage aux microservices.

- Réalisation de l'API backend du service grand public « Visitez Paris en métro ».
- Extraction et transformation en Python des logs du réseau ferré, produisant des statistiques à destination du grand public.

- Développement du backend et des APIs REST (framework CodeIgniter), et rédaction de la documentation technique, pour la refonte de Hub'Scan, plateforme SaaS d'audit et de validation de tags analytics et marketing à grande échelle.

- Développement d'applications web d'entreprise (Java JEE, PHP) pour des clients des secteurs public, télécoms et bancaire, chez un partenaire officiel RedHat.

FORMATION

CERTIFICATIONS

LANGUES