www.osmanrami.fr me@osmanrami.fr 06.02.73.49.22

RAMI OSMAN

Skills

Cloud / OPS
AWS • Multi-AZ Architecture • SageMaker • GCP Cloud Run • Cloud Armor VPC • Networking • Firebase • Cloud Analytics • Terraform (IaC) • Ansible • Kubernetes • Docker • SOPS • GitLab CI/CD (Enterprise) • GitHub Actions • TeamCity • Nexus • SonarQube • DevOps Process Automation • Global DevOps Delivery • HDS-compliant • Data Environments Monitoring • Observability (Grafana)

Architectures Applicatives
Hexagonale • Micro services

Languages / Framework
Java • Spring Boot • TypeScript • React / Next JS • Dart • Flutter

Messaging / QUEUE
Kafka • AWS SNS / SQS • GCP PubSub

AI / LLM
Local LLM • RAG • Agentic AI
• n8n • Tensorflow
• Fine tuning • ComfyUI
• Vespa AI
• Databricks

Database
Relational • PostgreSQL
In Memory • Redis
Documents • MongoDB / Firestore

Formation

EISTI (CY TECH)
Engineering Degree, specialization in Artificial Intelligence

Lycée St Exupéry - CPGE
Preparatory Classes MPSI-MP

Harbour Space - Barcelone
Cyber Security program / Linux System

Certifications

AWS Partner: Generative AI Essentials

Codility Silver Award

SkillValue Java 11

PROFILE

Solutions Architect & Tech Lead (7+ years) specializing in Cloud, distributed architectures, and applied AI, with experience on large-scale, mission-critical platforms for major enterprises (SNCF, APHP, BNP, Volkswagen). Strong expertise in designing scalable and secure architectures, as well as leading technical teams and complex projects.

EXPERIENCES

Tech lead / Architect Solutions at **Inventiv IT** — 2025 - Present - Paris

- ***Inserm*** *: Software Factory*

  Designed and implemented a company-wide software factory for 12 teams. Standardized CI/CD pipelines, improved code quality, and reduced delivery times. Technologies: GitLab Enterprise, Nexus, SonarQube, CI/CD, HDS-compliant data, Kubernetes, Ansible.

  Result: Automated CI/CD pipelines for 40+ projects with HDS compliance Kubernetes & Ansible industrialization reduced manual interventions by 50%

- ***Volkswagen Group*** *: GENWIN Web Platform & SA Platform*

  Architecture and deployment of a highly available web platform on multi-AZ AWS cloud infrastructure using Infrastructure as Code. Designed scalable and maintainable application architectures (Clean Architecture). Guided development teams on best practices (TDD, quality, CI/CD). Actively involved in functional scoping, architecture documentation (DAT), and sprint follow-up using Agile methodologies, in close collaboration with business teams.

  Designed an AWS Multi-AZ HA architecture ensuring 99.9% availability

- ***SNCF*** *: Data Market Place / Mobile Application MyCot*

  Designed a cloud-native data platform enabling large-scale data exposure and consumption. Implemented microservices-oriented architectures using managed services. Integrated observability and monitoring solutions.

  Deployed Spring Boot microservices on GCP used by 90,000 users

- ***RATP*** *: Platform Antilope Migration (Struts Java SE 8 to Spring Java 21)*

  Technical leadership of a legacy platform migration toward a modern Spring Boot microservices architecture. Experimented with and integrated AI tools (agents) to accelerate technical tasks and improve team productivity.

System Analyst at **IBM France** — 2023 - 2024 - Paris

- ***MARS INC*** *: Project 1Nav*

  Industrialized DevOps processes in an international environment. Automated deployments and data workflows.

  Automated global DevOps processes reducing manual tasks by 45%

- ***APHP*** *: E-Reply System / AI Models for medial followup / Web platform*

  Designed an AI-augmented medical monitoring platform for automated analysis and processing of patient communications. Integrated AI models (LLMs, RAG) and vector search engines to improve response relevance and assist medical teams. Operated in a highly demanding environment in terms of security and data confidentiality.

  Implemented Vespa AI Vector Search reducing email response time by 70%

- ***BNP*** *: Financial Instruments Account (CIF)*

  Development and evolution of a critical financial system based on hexagonal architecture and microservices..

Full stack developer at **Diverta Inc.** — 2022 - 2022 - Tokyo

Languages

Français (Native)

Anglais (Fluent)

Arabe (Fluent)

- ***Kuroko*** *: BaaS Platform (Backend as a Service) / BFF (Backend for frontends)*

Full stack developer at **Capgemini France** — 2018 - 2022 - Paris

- ***STELLANTIS*** *: Stellantis – GCB Project (Bill of Materials Transformation between FSA/PSA)*

  Transformation and harmonization of Bills of Materials between FSA and PSA systems in a large-scale industrial merger context. Designed and developed robust backend services ensuring data reliability, consistency, and traceability. Strong focus on software quality, performance, and integration within a complex, multi-team information system.

- ***GARI*** *: Collaborative platform*

Freelance and personal projects - 2020 - 2025 - Paris

- ***Leventis Platform*** *: Sales Prospecting Platform*

  Designed and developed a sales prospecting platform used across multiple business applications.
  Full Stack development (Next.js, NestJS, Supabase) improving sales team productivity.
  Implemented shared APIs enabling multi-application integration.
  Industrialized CI/CD via GitHub Actions automating 100% of deployments

- ***Clairvoyance*** *: Web & Mobile Application*

  Developed a web and mobile platform connecting psychic experts and users Flutter + Firebase + GCP architecture supporting 30,000 active users Implemented real-time system improving user interaction fluidity Integrated Google Analytics to optimize engagement and retention Deployed scalable cloud infrastructure ensuring high platform availability.

- ***Magik AI*** *: Web & Mobile Application*

  Designed a web and mobile interactive media generation platform based on open-source AI models Deployed on RunPod GPU cluster for distributed inference Implemented GenAI pipelines (RAG, fine-tuning, ComfyUI) reducing media generation time Built automatic upscaling system improving visual quality of generated content Orchestrated AI agents and n8n workflows automating media production.