

Gennaro Paglia

Senior Backend & Distributed Systems Engineer | Cloud-Native Architect | AI-Integrated Systems

Terni, Italy | +39 389 063 9413 | gennaro.paglia@gmail.com

Website: gennaropaglia.me | LinkedIn: linkedin.com/in/gennaropaglia | GitHub: github.com/pagliagen

PROFESSIONAL SUMMARY

Senior Backend and Distributed Systems Engineer with 15+ years of experience owning the entire product lifecycle, from architectural design to production-grade cloud deployment on Amazon Web Services (AWS). Deep expertise in microservices architecture, serverless infrastructure, event-driven systems, and the integration of cost-optimized, production-ready Artificial Intelligence (AI) workflows using Large Language Models (LLM), LangGraph, Retrieval-Augmented Generation (RAG), and semantic search.

SKILLS

Languages and Frameworks: Node.js (v22), TypeScript, JavaScript, Python, FastAPI, Flask, Express.js

Cloud and Infrastructure: Amazon Web Services (AWS), AWS Lambda, ECS Fargate, DynamoDB, SQS, API Gateway, CloudWatch, Infrastructure as Code (IaC), AWS CDK, Serverless Framework, Docker, Docker Compose, cross-account IAM architecture, serverless design patterns

Distributed Systems: Microservices architecture, event-driven architecture, Pub/Sub messaging, FIFO queues, distributed job orchestration, concurrency control, race condition mitigation, real-time state synchronization, WebSockets, asynchronous processing, retry logic, dead-letter queues

AI and Machine Learning Integration: LLM workflow orchestration, LangGraph, LangChain, Claude API (Anthropic), Retrieval-Augmented Generation (RAG), semantic search, vector databases (Qdrant), embeddings pipelines, AI cost optimization, prompt engineering

Databases: MongoDB, DynamoDB, Redis, Qdrant, MySQL, Oracle

API Design and Testing: REST API design, OpenAPI specification, Swagger documentation, API validation, unit testing, integration testing, Mocha, Chai, NYC code coverage

PROFESSIONAL EXPERIENCE

Senior Backend Engineer / Architect

Exelab | Remote, Italy | January 2020 – Present

- Owned the architectural evolution and end-to-end deployment of enterprise backend systems, driving technical decisions from initial design to production releases.
- Architected and shipped the Exelab Triage Alarm Analysis System, an AI-driven platform for automated production error triage across distributed AWS environments.
- Designed a 5-phase LangGraph workflow orchestrating Claude-based LLM reasoning, reducing AI inference costs by 35% (approximately 7,500 tokens saved per job) and cutting latency from 5-10 seconds to under 100 milliseconds via template-based generation.
- Engineered distributed FIFO job orchestration using DynamoDB atomic operations and AWS Lambda, eliminating concurrency issues in real-time state synchronization.
- Managed cross-account IAM architecture to securely aggregate and analyze logs and alarms across 10+ distinct AWS environments.

- Drove structural refactoring across legacy Node.js and TypeScript microservices, reducing technical debt and improving production stability.

Key Project: Remax Intelligent Broker Assistant (Event-Driven AI and RAG Pipeline)

Production-grade conversational AI system that ingests, analyzes, and enriches real estate workflows.

- Designed and deployed a real-time backend in Python and FastAPI processing broker notifications via WhatsApp webhooks.
- Built a Retrieval-Augmented Generation (RAG) architecture grounding LLM responses in internal documentation and contextual knowledge.
- Integrated bidirectional synchronization with the HubSpot CRM API to validate broker data, track interactions, and enrich lead profiles.
- Handled asynchronous processing and webhook reliability under high-traffic spikes using AWS event-driven components (Lambda, SQS, DynamoDB).

Key Project: Euroansa HubSpot Integration API (Serverless Event-Driven CRM Integration)

Serverless integration platform for bidirectional synchronization between enterprise systems and HubSpot CRM.

- Designed an AWS event-driven architecture based on Lambda, SQS FIFO queues, and DynamoDB.
- Implemented a REST API with OpenAPI validation and Swagger documentation.
- Built reliable asynchronous processing with dead-letter queues, object mapping, and cross-platform ID tracking.
- Integrated the full HubSpot lifecycle (CRUD operations, associations, webhooks) and achieved production-grade test coverage with Mocha, Chai, and NYC.

Senior Full Stack Developer

Kaleido | Italy | 2015 – 2020

- Developed custom web applications for enterprise clients.
- Designed and maintained backend services and system integrations.
- Improved development workflows and structured release processes.
- Translated business requirements into scalable backend solutions.

Analyst Programmer / Senior Analyst

Top Network | Italy | 2006 – 2015

- Backend development and enterprise system integrations for major clients including Telecom Italia, WIND, Tiscali, ATAC and ISTAT.
- CRM platform development (Siebel) and legacy system modernization.
- Long-term system maintenance and incremental improvements.

OPEN-SOURCE PROJECT

Full-Stack Distributed Platform | github.com/pagliagen/tenpennynovels

Event-driven, AI-integrated reference architecture showcasing production patterns.

- Designed a 5-service event-driven microservices architecture using Redis Pub/Sub.
- Implemented dual-storage semantic search (MongoDB + Qdrant vector database) with sub-100ms approximate nearest neighbor (ANN) queries.
- Built an asynchronous embeddings pipeline with retry logic and controlled concurrency.
- Implemented real-time multi-instance state synchronization via Socket.IO and Redis Adapter.
- Integrated the Claude API with persona consistency enforcement and anti-repetition logic.

- Orchestrated 9 Dockerized services (Node.js backends, Flask ML service, vector database) with Docker Compose.

Stack: Node.js 22, TypeScript (700+ source files), Express 5, MongoDB, Redis, Qdrant, Docker Compose

EDUCATION

Master of Science (MSc) in Aerospace Engineering

Università degli Studi di Napoli Federico II

LANGUAGES

- Italian: Native
- English: Professional working proficiency

Autorizzo il trattamento dei miei dati personali contenuti nel CV ai sensi del D.Lgs. 30/06/2003, n. 196 e dell'art. 13 GDPR (Regolamento UE 2016/679).