

DANIEL CASO QUINTANILLA

Senior Backend Engineer | Python & NestJS | AI/LLM Integration (RAG) | Microservices
Lima, Peru | +51 928728362 | daniel.caso.quintanilla@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

Backend Engineer with 5+ years designing scalable microservices and REST APIs in production. Expert in Python (FastAPI, Pydantic v2, SQLAlchemy, async/await, Celery) and TypeScript (NestJS), with additional experience in Java (Spring Boot), Go, and C# .NET. Strong background in AWS and GCP, Docker, Kubernetes, and CI/CD, plus real LLM integration (Anthropic Claude, OpenAI) with RAG over ChromaDB and Qdrant. Proficient in Clean and Hexagonal Architecture, Event-Driven systems (Kafka, RabbitMQ, Redis Streams), and observability (OpenTelemetry, Prometheus, Grafana).

PROFESSIONAL EXPERIENCE

Kwema

St. Louis, United States (Remote)

Backend Engineer, Integrations

Mar 2025 – Present

- Developed a C# .NET plugin for Genetec Security Center 5.13, enabling real-time panic alerts on security dashboards via an embedded HTTP webhook server.
- Architected a BLE indoor positioning system replacing REST polling with Juniper Mist webhook integration, reducing emergency alert latency from seconds to sub-second.
- Built a FastMCP server in Python exposing 8 specialized tools for SQL Server observability (queries, indexes, blocked sessions, fragmentation), integrated into AI-assisted workflows.
- Led a GCP analytics initiative with Looker Studio dashboards over BigQuery and Vertex AI evaluation for anomaly detection on alert telemetry.
- Accelerated delivery through disciplined use of Claude Code and GitHub Copilot, enforcing mandatory code reviews and test-coverage gates for all AI-generated code.

Llatam S.A.C.

Lima, Peru (Remote)

Backend Engineer

Nov 2024 – Jul 2025

- Developed REST APIs in Python (FastAPI, Pydantic v2, SQLAlchemy, async/await) over PostgreSQL, applying Clean Architecture for testable domain logic.
- Optimized high-traffic endpoints with cursor-based pagination, composite indexes, and query-plan analysis, reducing p95 response times during peak hours.
- Integrated Redis caching (read-through and write-behind) and built CI/CD pipelines with GitHub Actions running pytest, ruff, and mypy gates for validated deployments.
- Authored an Alembic migration strategy for zero-downtime schema evolution, including expand-and-contract patterns for high-traffic tables.

Fide PBC

Madrid, Spain (Remote)

Backend Engineer

Mar 2020 – Oct 2024

- Owned core backend services in Python (Django, FastAPI) and Java (Spring Boot) supporting daily operations across 4+ years and multiple major releases.
- Migrated services to AWS and Azure using Docker, Kubernetes, and CI/CD pipelines, reducing manual deployment effort and standardizing release procedures.
- Built ETL pipelines on Azure Databricks with PySpark, and implemented event-driven patterns with Kafka, Redis Streams, and Celery for resilient async processing.
- Mentored junior engineers through pair programming and code reviews; led documentation-first initiatives to reduce onboarding time.

TECHNICAL SKILLS

Languages & Frameworks: Python (FastAPI, Pydantic v2, SQLAlchemy, Alembic, Celery, Django, Flask), TypeScript (NestJS, Node.js, React), Java (Spring Boot), Go (Gin), C# .NET

AI & LLM Engineering: Claude Code (agents, sub-agents, hooks, skills, MCP servers), GitHub Copilot, Cursor; Anthropic and OpenAI APIs; RAG with ChromaDB, Qdrant, pgvector; FastMCP; embeddings (bge-m3); local inference (llama.cpp)

Cloud & DevOps: AWS (Lambda, S3, API Gateway, RDS, CloudWatch, SQS, SNS, ECS), GCP (BigQuery, Vertex AI, Cloud Run, Pub/Sub), Docker, Kubernetes, Terraform, GitHub Actions, ArgoCD

Observability & Reliability: OpenTelemetry, Prometheus, Grafana, structlog, Sentry, Loki, Jaeger, SLO/SLI design

Data & Messaging: PostgreSQL, MySQL, SQL Server, MongoDB, Redis (Streams, Pub/Sub, caching), Kafka, RabbitMQ, NATS, PySpark, CDC with Debezium

Architecture & Testing: Clean and Hexagonal Architecture, Microservices, CQRS, Event-Driven, Saga, DDD; pytest, Hypothesis, Locust, Playwright, TDD, contract and E2E testing

Also exploring: SwiftUI (mobile, currently applied at Kwema)

EDUCATION

Universidad Privada del Norte — B.S. in Computational Engineering

Mar 2025 – Present

SENATI — Software Development

Mar 2018 – Jan 2021

Languages: Spanish (Native) | English (Professional Working Proficiency)