MARIANO DÍAZ

Senior Full Stack Engineer | AWS & Generative AI

TypeScript · Node.js · React / Next.js · React Native · PostgreSQL · AWS · Terraform · LLMs · RAG · AI Agents

lth.nano@gmail.com | linkedin.com/in/diaznano | +54 9 381 412-6387

Tucumán, Argentina (GMT-3) · Open to remote work and relocation

PROFESSIONAL SUMMARY

Full stack software engineer with 7+ years of experience building web and mobile products and taking them to production on AWS. I work in a forward deployed way: together with each client I understand the process, design the solution and implement it integrated with their systems all the way to production. I build backends with Node.js (NestJS, Express) and Ruby on Rails, frontends with React and Next.js and React Native apps, deployed on EC2, Lambda, API Gateway, RDS PostgreSQL, S3 and CloudFront, with GitHub Actions CI/CD and Terraform. Since February 2024 I've been taking generative AI to production: agents with RAG, tool calling and MCP handling about 17,000 conversations per month across 35 clients, a multimodal OCR document pipeline and a multi-agent architecture. Today I split my time between development and technical leadership of a team of about 10 people; previously Team Lead on TaDa, an app with 10M+ downloads.

PROFESSIONAL EXPERIENCE

Xetro · Remote, Argentina / Spain

Software and AI consultancy where I have worked since 2019. Through Xetro, I was placed on staff augmentation assignments at the companies listed below.

Software & AI Projects Director

November 2024 – Present

50/50 role split between hands-on development and technical leadership of a team of about 10 people (engineering, design and paid media), serving a portfolio of 35 clients in healthcare, tourism, logistics, food service and retail across Argentina and Spain.

- **Forward deployed with clients:** map each client's process (healthcare, tourism, logistics, food service and retail), define the solution and implement it integrated with their CRMs, management system databases and payment providers, through to production.
- **AI agents in production:** conversational agents on the WhatsApp Business API with tool calling, conversational memory and confidence- and risk-based human handoff: about 17,000 conversations per month across 35 clients (500 to 1,000 per client, measured in Chatwoot and our in-house CRM), with about 80% closed by the agent without handoff to a human.
- **AI quality and cost:** select and combine models (Claude, OpenAI, Gemini) based on cost, latency and accuracy; volatile data (prices, stock, appointments) is always queried from the database, with abstention rules. Migrated context to RAG on Pinecone and pgvector: about 15% lower API (token) cost per conversation, measured month over month.
- **Document AI and multi-agent systems:** document extraction and validation pipeline with multimodal OCR (GPT-4o-mini) and an extractor agent (Claude Sonnet), about 100 documents per month; multi-agent architecture for a logistics platform; agents connected to internal tools and data through MCP and function calling; n8n automations.
- **AWS deployment:** NestJS, Express, Ruby on Rails and Next.js projects on EC2 behind Application Load Balancers, with Route 53 and SSL; RDS PostgreSQL with snapshots; images and documents in S3, served through CloudFront when needed.
- **Serverless, auth and mobile:** AWS Lambda functions in JavaScript (Node.js) and APIs with API Gateway; user sign-in with Amazon Cognito; React Native apps with AWS Amplify, push notifications and image uploads to S3.
- **CI/CD, infrastructure and observability:** GitHub Actions pipelines deploying to AWS; part of the infrastructure in Terraform, containerized with Docker; monitoring with CloudWatch (logs, metrics and alarms), Sentry and Grafana, plus automated daily audits of agent responses. Reduced costs by consolidating services onto fewer instances.
- **Product and integrations:** freight logistics platform on Rails 7 organized in 12 modules; integrations with Mercado Pago, Chatwoot and Meta Conversions API.

React Team Lead

March 2023 – February 2025 · Promoted to Director in November 2024

- Led a team of up to 10 people on an appointment scheduling web app for an international company: planning, estimation, code review and Scrum.

- Built with React, Redux and REST APIs.
- Shipped the first AI agents to production for clients in February 2024.

Technical Interviewer | Eureka Labs
August 2025 – June 2026 · Remote · Fixed-term contract, alongside Xetro

- Assessed about 30 candidates for JavaScript roles (Node.js, React, Next.js and React Native) using my own evaluation criteria and live coding exercises, and wrote hiring recommendations.

React Native Developer | Bankingly
June 2024 – November 2024 · Remote · Staff augmentation through Xetro

Digital banking platform for financial institutions.

- Built critical digital banking flows (login, transfers, payments and biometrics) with React Native, TypeScript, Redux and Jest.

React Native Developer | Cultura IT
November 2023 – June 2024 · Tucumán, Argentina · Staff augmentation through Xetro

GoodSeed Notebook: crop tracking app that replaced a manual paper-based process.

- Implemented an offline-first architecture with data sync and wrote native camera modules for Android and iOS.

React Native Team Lead / React Native Developer | Onetree
February 2021 – March 2023 · Remote · Staff augmentation through Xetro · Team Lead since February 2022

Product context: TaDa, a beverage delivery app with 10M+ downloads in 10+ countries.

- Joined as a developer and after one year moved up to lead one of the project's two 5-developer teams.
- Refactored the full checkout, implemented analytics with RudderStack and improved performance with the backend team (React Native, Redux, GraphQL).

Web & Mobile Developer | Diveria, Kubikware and Xetro
June 2019 – September 2021 · Remote · Staff augmentation through Xetro and Xetro projects

- Diveria: logistics web app for Avancargo (Next.js, NestJS). Kubikware: React Native apps Modamily and Breast Advocate. Xetro: Workast and Forbes8. All published on the App Store and Google Play.

EDUCATION & LANGUAGES

Information Systems Engineering | Universidad Tecnológica Nacional (UTN) · 2013 – 2019 · Advanced student: 1 course remaining
Systems Analyst (intermediate degree) | Universidad Tecnológica Nacional (UTN) · 2019
Academic exchange | Université de Technologie de Belfort-Montbéliard, France · 2017

Languages: Spanish (native) · English (B1, intermediate) · French (B1, DELF certified) · Catalan (basic, in progress)

CORE SKILLS

- **Generative AI:** LLMs (OpenAI, Claude, Gemini), AI agents, multi-agent systems, RAG, embeddings, Pinecone, pgvector, prompt engineering, tool calling, MCP, guardrails, human-in-the-loop, multimodal OCR, n8n
- **AWS:** EC2, Lambda, API Gateway, S3, CloudFront, RDS PostgreSQL, Application Load Balancer, Route 53, SSL, Cognito, Amplify, CloudWatch; in environments managed by other teams: VPC, DynamoDB, ElastiCache, EKS
- **DevOps & Observability:** Terraform, GitHub Actions, Docker, Git, Sentry, Grafana, RudderStack; also Railway, Supabase and Neon
- **Backend:** TypeScript, Node.js, NestJS, Express, Ruby on Rails, REST APIs, GraphQL, webhooks; Python for automation and scraping
- **Frontend & Mobile:** React, Next.js, Redux, Tailwind CSS, Hotwire, React Native (iOS and Android)
- **Databases:** PostgreSQL, MySQL, MongoDB, Redis, Supabase, Neon, query optimization
- **Integrations:** WhatsApp Business API, Chatwoot, CRMs, Mercado Pago, Meta Conversions API
- **Leadership:** technical leadership of 5 to 10-person teams, code review, sprint planning and estimation, Scrum, mentoring, technical interviewing, client management