TECHNICAL LEADERSHIP & ARCHITECTURE PORTFOLIO · 2024–2026

ALEXANDER MARRA MONCKS

Head of Engineering · Engineering Manager · Technical Lead & AI Solutions Architect

Selected case studies spanning engineering leadership, multi-tenant SaaS platforms, multi-agent AI orchestration, server-side growth engineering, self-hosted infrastructure, and low-level systems engineering.

35+ years	35 → 136	4 countries	15+ projects
In IT & Engineering	**Team scaling**	**Clients served**	**Shipped / in prod (2024–26)**
From Senior C# Dev to Delivery Manager, CPO & Founder	$1.5M → $18M revenue at Stefanini (Fortune 500)	Brazil, United States, Sweden, and Nigeria	Leading 4 collaborators + 9 specialized AI agents

Hands-on leadership backed by enterprise delivery discipline

With 35+ years in IT and an Executive MBA from FGV, I lead MMSoft Soluções Digitais (since 2011), taking ambiguous business problems all the way to production platforms—covering discovery, architecture, pricing, contracts, team leadership, and operations. Previously, I scaled an enterprise unit at Stefanini (J&J, Honda, Embraer, CSN, Via Varejo), led distributed teams at Agile Engine (Argentina, Ukraine, California), and served as CPO at Falepaco (Coca-Cola account).

- **Engineering governance:** Run human teams alongside a nine-agent AI engineering workflow with strict gates (Dev → Code Review → QA → human sign-off), environment isolation, and credential security.
- **Shared platform architecture:** Products such as BIPo, CUFA, Mailer, and Deskcomm share a unified multi-tenant control plane, self-hosted infrastructure, and a shared observability stack.
- **Commercial & P&L ownership:** Commercial offer design (SaaS and VAR licensing), pre-sales, and contracts in Brazil and internationally.

Executive & technical timeline

Period	Company & Role	Scope & Key Outcomes
2011 – Present	**MMSoft Soluções Digitais — Founder & Engineering Lead**	4-person team + 9-agent AI workflow; clients across 4 countries; includes Technical Lead at Agile Engine (2021–2022, AR/UA/US teams).
2023 – 2025	**Falepaco — Chief Product Officer (CPO)**	Product strategy and tech direction; Coca-Cola account (Instituto Coca-Cola Brasil); distributed teams including Nigeria.
3 months	**World Wide Technology — Boomi Integration Developer**	Three-month engagement on Boomi integrations for asset management (calendar start and end dates not provided).
2016 – 2021	**Stefanini — Senior C# Dev → Delivery Manager**	Scaled unit from 35 to 136 people and $1.5M to $18M revenue; J&J, Honda, Embraer, CSN, Via Varejo (Brazil & Argentina).
2010 – 2011	**Diebold — Senior .NET Developer**	Predictive data capture and ATM software interfaces for Bank of America.

alex@marramoncks.com.br • +55 (12) 99738-7715 • github.com/alexmoncks • mmsoft.com.br • LinkedIn: linkedin.com/in/alexander-marra-moncks

CASE 01 · FLAGSHIP PLATFORM · 2024–2026 • In existing portfolio (refreshed)

PULSE / VAR / BIPo — Multi-Tenant Control Plane & Geospatial Intelligence

AI-driven political and territory intelligence SaaS commercialized directly and through white-label VAR partners.

Role	Markets	Model
Founder, PO & Lead Architect	**Brazil & United States**	**SaaS Plans + VAR License**

Context: Political campaigns and agencies needed territory intelligence (polling, sentiment tracking, geospatial targeting) packaged so resellers could operate it under their own brand.

Leadership & Role: Defined the product strategy (Light, Small and Full plans, plus Engage and Social Reputation modules), Lei do Bem R&D documentation, and directed the engineering team.

Architecture: Single control plane (BIPo) automating VMs/containers via n8n + Terraform Runner on Proxmox bare-metal with Traefik routing, MongoDB 2dsphere (incl. US Census TIGER/Line), and cost-aware LLM model selection.

Results: Paying direct and VAR clients in Brazil, first US international contract (13-month chamber of commerce engagement), bare-metal margin preservation, and keynote case at AI Brasil Experience.

Core stack: Terraform, Proxmox VE, Docker Swarm, Traefik, React, Node.js, PostgreSQL/Supabase, MongoDB 2dsphere, Redis, n8n, Claude API, OpenAI API, Prometheus/Grafana

---

CASE 02 · AI ENGINEERING & MCP · 2024–2026 • Updated (9-agent workflow is new)

Nine-Agent AI Orchestration, Custom MCP Servers & LMTagScript

Formal multi-agent engineering governance paired with a proprietary guardrails DSL and production MCP tooling.

Role	Scale	Focus
AI Architect & Engineering Lead	**9 agents + 10+ client projects**	**Quality, cost & security**

Context: Scale engineering velocity with AI while enforcing strict code review, QA validation, token-cost governance, and zero secret leakage across concurrent projects.

Leadership & Role: Designed a nine-agent operating model (Dev, Code Review, QA, Designer, Commercial, Marketing, Legal, Accounting, plus central orchestration) with explicit review gates and human approval.

Architecture: Claude Code CLI (Opus) execution via tmux, custom n8n Data Tables MCP server, TypeScript JSON-RPC/STDIO MCP monorepo (@lmtagscript/mcp), and EBNF-based LMTagScript DSL with LOOPGUARD.

Results: LMTagScript deployed across 10+ client engagements reducing AI delivery time by 50%; structured PR/QA traceability and isolated project execution.

Core stack: Claude Code CLI, Grok Bot, MCP (JSON-RPC/STDIO), LMTagScript (Python/TS), n8n-mcp, Supabase MCP, tmux

CASE 03 · PRODUCTION B2C/B2B SAAS · 2025–2026 • In existing portfolio (stack updated)

Pro Hub Clubs — Competitive Sports SaaS with Computer Vision & WhatsApp OTP

Full-stack league, federation, and club management platform supporting four tournament formats.

Role	Traction	Stack
Tech Lead & Architect	**~890 subscribers (~6 months)**	**TypeScript / Prisma / PWA**

Context: Competitive e-sports clubs and federations needed roster management, contracts, automated match-sheet ingestion, schedules, and player verification.

Leadership & Role: Led system architecture, automated code-review workflows for the engineering team, and the freemium-to-paid product design.

Architecture: Full-stack TypeScript (React 18 + Vite + Tailwind/shadcn + Express 5 + Prisma 7 + PostgreSQL) on Docker Swarm/Traefik, AI vision OCR for match screenshots, WhatsApp OTP (WAHA), and PWA support.

Results: Reached ~890 subscribers in roughly six months on a freemium model, backed by automated Playwright/Vitest suites and continuous production evolution.

Core stack: TypeScript, React 18, Vite, Tailwind CSS, shadcn/ui, Express 5, Prisma 7, PostgreSQL, Vision AI, WAHA, Resend, PWA, Vitest, Playwright

---

CASE 04 · TELECOM & CRM · 2026 • Updated (adds Deskcomm & centralized WAHA)

Barigui Telecom (AI Support) & Deskcomm CRM / Centralized WAHA

Local-LLM WhatsApp/voice customer support on client bare-metal alongside unified CRM and WhatsApp gateway.

Role	Hosting	Channels
Architect & Delivery Lead	**Client Proxmox / MMSoft**	**WhatsApp & Voice**

Context: An ISP/cable operator needed WhatsApp and voice automation integrated with its management systems and without a sandbox; internally, MMSoft needed to consolidate CRM and WhatsApp channels.

Leadership & Role: Led the kickoff and a two-wave delivery plan (WhatsApp pilot, then voice), plus knowledge transfer with the client team.

Architecture: Sentiment-aware LLM agents with RAG running locally on the client's Proxmox host; at MMSoft, migrated Deskcomm (Supabase/PostgreSQL) to a dedicated TLS VM and architected a centralized WAHA gateway.

Results: Structured the WhatsApp pilot into production, with local agents and isolated administrative access.

Core stack: WhatsApp Business API / WAHA, n8n, Local LLMs, Vector RAG, Supabase, PostgreSQL, Proxmox VE, Traefik, MCP

CASE 05 · VOICE & WHATSAPP AI AGENT · NIGERIA · 2024–2025 • In existing portfolio

Tracy AI Agent — Production Voice & WhatsApp Concierge for Secrets Palace

Autonomous overnight VIP table booking agent operating across voice calls and WhatsApp in Nigeria.

Role	Market	Tooling
Solution Architect & Dev	**Nigeria (English)**	**8 custom MCP tools**

Context: A high-end nightclub handled VIP reservations overnight across phone and messaging, requiring strict enforcement of operating hours and mandatory deposits.

Leadership & Role: Owned end-to-end solution design, guardrail engineering, payment/deposit logic, and complete production handover documentation for the international client.

Architecture: 10-step conversational flow in n8n powered by ElevenLabs voice synthesis and 8 MCP tools, with business rules (operating hours and deposit calculation) enforced in tool code rather than in the prompt alone.

Results: Delivered production-ready with full handover reporting; the deterministic MCP + n8n agent pattern became a reusable blueprint across subsequent client deployments.

Core stack: OpenAI GPT-4o, ElevenLabs, Model Context Protocol (MCP), n8n, Node.js, Supabase, WhatsApp Business API

CASE 06 · NONPROFIT SAAS & ENTERPRISE · 2023–2025 • In existing portfolio

CUFA Rio de Janeiro (1,000h) & Instituto Coca-Cola Brasil (Falepaco)

Multi-tenant grant/cost governance SaaS with ERP sync and CPO product leadership for Coca-Cola.

Role	CUFA Scope	Integrations
Lead Architect / CPO (Falepaco)	**1,000h + reseller channel**	**Conta Azul, Salesforce, Bitrix24**

Context: CUFA-RJ required three-quote procurement, accounting/financial approvals, and Conta Azul reconciliation per sponsored project; Falepaco's Coca-Cola account needed unified student CRM and messaging.

Leadership & Role: Scoped the 1,000-hour CUFA engagement and phased rollout, reusing the BIPo control plane; as Falepaco CPO (2023–2025), led distributed product teams.

Architecture: CUFA built atop BIPo's multi-tenant control plane with bidirectional Conta Azul ERP sync; Coca-Cola educational platform connecting Salesforce, Bitrix24, and WhatsApp through n8n.

Results: Validated commercial and post-go-live support model at CUFA and delivered enterprise educational integrations across distributed teams.

Core stack: Multi-tenant SaaS, n8n, Conta Azul API, Salesforce, Bitrix24, WhatsApp, PostgreSQL, Docker

CASE 07 · INTERNATIONAL CLIENTS · SWEDEN & USA · 2024–2026 • In existing portfolio (updated with 2026 US advisory)

Sports Rekrytering (Sweden), GBA Facilities & MetaGami (USA)

OAuth2 financial integrations, AI-scored B2B lead pipelines, and regulated US deployments.

Markets	Compliance	Delivery
Sweden & United States	**SEC / SAML / RBAC / Audit**	**Middleware, AI & QA**

Context: International engagements demanded native Swedish ERP integration (Fortnox), resilient B2B lead scoring, and strict security/audit controls for SEC-regulated US financial firms.

Leadership & Role: Led technical design, pipeline stabilization, US security briefings, and a code-level diagnostic with a commercial estimate in English for MetaGami.

Architecture: Multi-tenant Fortnox OAuth2 middleware with GitHub Actions CI/CD; Apify + GPT-4o-mini + Pipedrive workflow with batch/timeout recovery; self-hosted architecture with encryption, SSO SAML, RBAC, and local inference.

Results: Bilingual (English/Swedish) sports recruiting platform in production, stabilized outbound automation, and a technical diagnostic delivered to the US client.

Core stack: Fortnox API, OAuth2, n8n, Apify, GPT-4o-mini, Pipedrive, GitHub Actions, Docker, SSO SAML, RBAC, Encryption

---

CASE 08 · GROWTH ENGINEERING & ADS API · 2025–2026 • New (not in current résumé/portfolio)

Server-Side Conversion Tracking (GA4 MP / Meta CAPI) & Google Ads API Client

LGPD-compliant, zero-duplication revenue attribution paired with a guarded Python client for Google Ads.

Role	Privacy	Tests
Growth Architect & Lead Dev	**Consent Mode v2 / LGPD**	**162 automated tests**

Context: Digital product funnels suffered from cookie loss and duplicate purchase events, while campaign operations needed programmatic Google Ads tooling without risking unapproved budget changes.

Leadership & Role: Architected decoupled dynamic checkout and server-side attribution (Eliana Severino Receitas and TMBJ) and specified mandatory safety guardrails for Google Ads API writes.

Architecture: Dynamic Asaas checkout (preserving affiliates and coupons) with Deno/Supabase and Resend delivery, server-side GA4 Measurement Protocol (source=mp) and Meta CAPI, Consent Mode v2, SEO/GEO (llms.txt), and Python google-ads-writer using Service Account auth and 6 safety locks (PAUSED-only creation, allowlist, dry-run).

Results: End-to-end Playwright verification in Asaas sandbox, zero duplicate purchase conversions, 162 passing tests on the Ads client, and live Search campaign launched in Expert Mode.

Core stack: Python, Google Ads API, GA4 Measurement Protocol, Meta CAPI, Google Consent Mode v2, Deno, Supabase, Asaas API, Resend, Playwright

CASE 09 · DATA AT SCALE & MESSAGING · 2025–2026 • New (not in current résumé/portfolio)

MMSoft Mailer (v0.4.2) & Painel-CNPJ (~2.7M Company Records)

Proprietary transactional/cadence email platform paired with multi-million-row B2B lead segmentation.

Role	Data Scale	Deliverability
Architect & Product Lead	**~2.7M companies**	**SPF / DKIM / DMARC**

Context: Outbound B2B operations required independence from costly SaaS mailers alongside fast filtering over millions of official Brazilian corporate registry records.

Leadership & Role: Directed the product roadmap (v0.3.2 → v0.4.2), RBAC and security reviews, and subdomain DNS deliverability setup.

Architecture: Containerized stack (API, Worker, Edge, Web, PostgreSQL, Redis) wired to Resend API/webhooks, 2FA TOTP, admin CLI, D0/D3/D7/D14 cadences with auto-suppression on reply, plus batch CNPJ list filtering with SDR role isolation and memory limits.

Results: Self-hosted outbound and transactional messaging with full domain authentication (SPF/DKIM/DMARC/ImprovMX) and real-time segmentation over ~2.7M business records.

Core stack: Node.js, Python, PostgreSQL, Redis, Resend API, ImprovMX, Docker, WHM DNS (SPF/DKIM/DMARC), RBAC

CASE 10 · WEB GAMES & PRODUCTION RECOVERY · 2026 • Updated (Next.js 14 / 14 games)

Acerte a Mosca (14 WebSocket Games) & LotoJoy (Billing + DevSecOps)

Modernizing a bilingual HTML5 gaming platform for AdSense/SEO and safely rescuing production subscription billing.

Role	Catalog	Focus
Architect & Tech Lead	**14 games (PT/EN)**	**SEO, AdSense and billing**

Context: Acerte a Mosca needed SEO, canonical/hreflang and ad-placement work to align with Google AdSense policies, plus PT/EN domain consolidation; LotoJoy had broken recurring billing in production and outdated dependencies.

Leadership & Role: Orchestrated the technical audit (Dev, Code Review and QA) of Acerte a Mosca, with an isolated preview for validation, and led the LotoJoy fix plan with an automated database backup before release.

Architecture: Acerte a Mosca refactored on Next.js 14 (App Router), next-intl, Pixi.js 8, Prisma/PostgreSQL, and unified WebSocket server (Railway + Cloudflare), fixing canonical/hreflang and ad placement; LotoJoy billing reconciliation branch and npm audit fix.

Results: Acerte a Mosca platform with 14 integrated games (including Kung Fu Castle) and SEO/AdSense work in progress; on LotoJoy, recurring billing restored and dependencies updated ahead of release.

Core stack: Next.js 14, React 18, Pixi.js 8, WebSockets (ws), Prisma, PostgreSQL, Railway, Cloudflare, Google AdSense, Asaas

CASE 11 · FPGA HARDWARE, C++ EMULATORS & 3D GAMES · AUG–OCT 2026 • New (not in current résumé/portfolio)

MSX2 + V9968 / Geo3D Ecosystem (Verilog RTL, C++ & Z80 Assembly)

Full-stack low-level systems engineering spanning FPGA synthesis, C++ emulators, custom assembler tooling, and four 3D titles.

Role	Target	Verification
Systems Engineer & Author	**Gowin FPGA / V9968**	**160 tests + RTL match**

Context: Push 8-bit hardware boundaries with the V9968/geo3d v2.0 3D coprocessor by bridging FPGA RTL, cycle-accurate C++ emulation, compiler tooling, and playable 3D software.

Leadership & Role: Drove end-to-end open-source delivery across GitHub repositories—from FPGA physical BSRAM placement and upstream emulator PRs to custom build chains and four complete 3D/2.5D games.

Architecture: Verilog/RTL contributions (V9968_Cartridge — 20-BSRAM placement and timing closure on Gowin FPGAs); geo3d v2.0 C++ emulation in openMSX and blueMSX-plus; Python 3.13 Z80/MegaROM assembler (msxAIbuilder) with headless WebMSX; MSXDOOM, MSXWOLF, GeoFighter, and Cube Maze.

Results: 70+ commits in V9968_Cartridge achieving zero negative slack, 160 passing tests in msxAIbuilder, byte-identical 3,755-frame RTL comparison in Cube Maze, and 30 fps (420 vertices) in GeoFighter.

Core stack: Verilog / SystemVerilog (Gowin FPGA), C / C++ (openMSX, blueMSX-plus, MinGW-w64), Z80 Assembly, Python 3.13, Three.js, Pytest

---

CASE 12 · SELF-HOSTED INFRASTRUCTURE · 2024–2026 • In existing portfolio (updated)

Hybrid Proxmox VE + Docker Swarm + Headscale Mesh VPN Platform

Shared, margin-friendly infrastructure powering 10+ production apps and isolated staging environments.

Role	Topology	Focus
Platform Architect	**Proxmox + Swarm + VPN**	**Predictable cost and isolation**

Context: Operate over a dozen SaaS products and isolated dev environments with predictable fixed costs, data sovereignty, and strict separation between production and staging.

Leadership & Role: Led commercial-readiness audits, backup and snapshot governance, private VPN segmentation, and standardized releases with rollback plans.

Architecture: Proxmox VE cluster with ZFS storage and automated backups, Docker Swarm with Traefik and Portainer, a Debian development host, and a Headscale/Tailscale private mesh VPN with internal DNS, plus WebRTC (MiroTalk SFU) and WHM/Cloudflare deliverability.

Results: Reliable multi-product operations on predictable infrastructure cost, with administrative surfaces isolated from the public internet.

Core stack: Proxmox VE, ZFS, Docker Swarm, Traefik, Portainer, Headscale / Tailscale, Debian Linux, nftables, Cloudflare, WHM, MiroTalk SFU

Enterprise Track Record, Teaching & Next Steps

Enterprise Delivery & Operations Management (Stefanini, Agile Engine, Diebold, WWT)

- **Stefanini (2016–2021):** Progressed from Senior C# Developer to Delivery Manager—scaling headcount from 35 to 136 and revenue from $1.5M to $18M across J&J, Honda, Embraer, CSN, Via Varejo, Raizen, Ultragaz, Ale Combustíveis, and Grupo Aldo.
- **Audited enterprise outcomes:** Honda Connected Car IoT (14 engineers, 5,000+ vehicles, $800K budget, 99.8% uptime); J&J Big Data (54 sources, 99.7% reliability); Via Varejo (15,000+ users, 600+ stores, +40% warranties); CSN (26 APIs, 85%+ test coverage, 3 months early).
- **Agile Engine (2021–2022):** Technical Lead coordinating a distributed team across Argentina, Ukraine, and California (USA) on C# deliveries.
- **Diebold (2010–2011) and World Wide Technology (three-month engagement):** As Senior .NET Developer at Diebold, predictive data capture and ATM software for Bank of America; at WWT, Boomi asset-management integrations.

Education, Teaching, Speaking & Community

- **Education & Certifications:** Executive MBA in Administration (FGV, 2022–2024); Business Management (FATEC-SP, 2016–2020); Systems Analysis & Development (Unitins, 2008–2013); SAFe 5 Certified, DEPC, Empretec (Sebrae), Innovation in the Digital Age (Mackenzie).
- **Teaching & Speaking:** Guest Professor of AI and Automation at FAAP; executive AI trainer at lideres.ai (2-day course & 6-hour agents/n8n workshop); AI Brasil Experience speaker; Guidepoint expert advisor.
- **Community:** Volunteer instructor teaching programming in underserved communities and computer skills to older adults; author of the short story “Tracy, a Voz que Não Tem Corpo.”

Let’s discuss engineering leadership, architecture, and AI

Open to remote roles worldwide as Head of Engineering, Engineering Manager, Technical Lead, or AI Solutions Architect. Happy to walk through the architecture, trade-offs, and metrics behind any of these engagements.

alex@marramoncks.com.br • +55 (12) 99738-7715 • github.com/alexmoncks • mmsoft.com.br • LinkedIn: linkedin.com/in/alexander-marra-moncks