PORTFÓLIO DE LIDERANÇA TÉCNICA & ARQUITETURA · 2024–2026

ALEXANDER MARRA MONCKS

Head de Engenharia · Engineering Manager · Arquiteto de Soluções & IA

Casos selecionados de liderança técnica, plataformas SaaS multi-tenant, orquestração multiagente de IA, growth engineering server-side, infraestrutura self-hosted e engenharia de sistemas de baixo nível.

35+ anos	35 → 136	4 países	15+ projetos
Experiência em TI	**Escala de equipe**	**Clientes atendidos**	**Em produção (2024–26)**
De dev C# sênior a Delivery Manager, CPO e fundador	US$ 1,5 mi → US$ 18 mi na Stefanini (Fortune 500)	Brasil, Estados Unidos, Suécia e Nigéria	Liderando 4 colaboradores + 9 agentes de IA

Liderança mão na massa com visão de negócio

Com mais de 35 anos em TI e MBA Executivo pela FGV, conduzo a MMSoft Soluções Digitais (desde 2011) transformando problemas de negócio complexos em plataformas em produção — do discovery, precificação e contrato à arquitetura, liderança do time e operação. Em minha trajetória corporativa, escalei operações enterprise na Stefanini (J&J, Honda, Embraer, CSN, Via Varejo), liderei times distribuídos na Agile Engine (Argentina, Ucrânia, Califórnia) e atuei como CPO na Falepaco (conta Coca-Cola).

- **Governança de entrega:** Estruturo equipes humanas e fluxos com agentes de IA sob gates formais (Dev → Code Review → QA → aprovação), isolamento de ambientes e proteção de credenciais.
- **Arquitetura compartilhada:** Plataformas como BIPo, CUFA, Mailer e Deskcomm reutilizam um mesmo control plane, infraestrutura self-hosted e observabilidade compartilhada.
- **Visão comercial e P&L:** Definição de oferta comercial (SaaS e licença VAR), pré-vendas e contratos no Brasil e no exterior.

Trajetória executiva e técnica

Período	Empresa & Papel	Escopo e Resultados Principais
2011 – Atual	**MMSoft Soluções Digitais — Fundador & Líder de Engenharia**	Time de 4 pessoas + 9 agentes de IA; clientes em 4 países; incl. Technical Lead na Agile Engine (2021–2022, times AR/UA/EUA).
2023 – 2025	**Falepaco — Chief Product Officer (CPO)**	Estratégia de produto e direção tecnológica; conta Coca-Cola (Instituto Coca-Cola Brasil); times distribuídos incl. Nigéria.
3 meses	**World Wide Technology — Dev. de Integrações Boomi**	Engajamento de três meses em integrações Boomi para gestão de ativos (mês de início e término não informado).
2016 – 2021	**Stefanini — Dev. C# Sênior → Delivery Manager**	Time 35 → 136 pessoas; receita US$ 1,5 mi → US$ 18 mi; contas J&J, Honda, Embraer, CSN, Via Varejo (BR e Argentina).
2010 – 2011	**Diebold — Desenvolvedor .NET Sênior**	Captação de dados preditivos e interface de caixas eletrônicos para o Bank of America.

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

CASO 01 · PLATAFORMA CARRO-CHEFE · 2024–2026 • Já no portfólio (atualizado)

PULSE / VAR / BIPo — Control Plane Multi-Tenant & Inteligência Territorial

Plataforma de inteligência política e geoespacial orientada por IA com comercialização SaaS e licença white-label (VAR).

Papel	Mercados	Modelo
Fundador, PO & Arquiteto Líder	**Brasil e Estados Unidos**	**Planos SaaS + Licença VAR**

Contexto: Campanhas, mandatos e agências precisavam de inteligência territorial (pesquisa de opinião, análise social e atuação georreferenciada) em modelo replicável por revendas.

Liderança & Papel: Defini a estratégia de produto (planos Light, Small e Full, com os módulos Engage e Reputação Social), a documentação de P&D na Lei do Bem e conduzi o time técnico.

Arquitetura: Control plane único (BIPo) automatizando VMs/containers via n8n + Terraform Runner em Proxmox bare-metal com Traefik, MongoDB 2dsphere (incl. TIGER/Line EUA) e escolha de modelo de LLM com custo medido por operação.

Resultados: Clientes pagantes no Brasil (diretos e via VAR), primeiro contrato internacional nos EUA (câmara de comércio, 13 meses), redução de custo infra via bare-metal e caso apresentado no AI Brasil Experience.

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

---

CASO 02 · ENGENHARIA COM IA & MCP · 2024–2026 • Atualizado (operação de 9 agentes é nova)

Orquestração Multiagente (9 Agentes), Servidores MCP & LMTagScript

Governança formal de engenharia assistida por IA somada a DSL proprietária de guardrails e conectores MCP em produção.

Papel	Escala	Foco
Arquiteto de IA & Líder de Engenharia	**9 agentes + 10+ projetos**	**Qualidade, custo e segurança**

Contexto: Escalar a produtividade de engenharia sem sacrificar revisão de código, QA, controle de orçamento de tokens e proteção absoluta de segredos em produção.

Liderança & Papel: Estruturei uma operação com 9 agentes especializados (Dev, Code Review, QA, Designer, Comercial, Marketing, Jurídico, Contábil e orquestração central), definindo hierarquia de gates e aprovação humana.

Arquitetura: Execução delegada a Claude Code CLI (Opus) via tmux, servidor MCP nativo para n8n Data Tables, monorepo MCP JSON-RPC/STDIO (@lmtagscript/mcp) e DSL LMTagScript em EBNF com LOOPGUARD.

Resultados: LMTagScript implantado em 10+ projetos de clientes com redução de 50% no tempo de desenvolvimento de soluções de IA; rastreabilidade completa de PRs, QA e zero exposição de credenciais.

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

CASO 03 · SAAS B2C/B2B EM PRODUÇÃO · 2025–2026 ● Já no portfólio (stack técnica atualizada)

Pro Hub Clubs — SaaS Esportivo com Visão Computacional e WhatsApp OTP

Plataforma de gestão de clubes, federações e quatro formatos de campeonatos de futebol virtual.

Papel	Tração	Stack
Tech Lead & Arquiteto	**~890 assinantes (~6 meses)**	**TypeScript / Prisma / PWA**

Contexto: Comunidades e ligas competitivas demandavam gestão de elencos, contratos, súmulas, calendários e comunicação automatizada em ambiente escalável.

Liderança & Papel: Conduzi a arquitetura técnica, a revisão automatizada de PRs da equipe e a definição do modelo freemium com planos pagos.

Arquitetura: Aplicação full-stack TypeScript (React 18 + Vite + Tailwind/shadcn + Express 5 + Prisma 7 + PostgreSQL) em Docker Swarm/Traefik, leitura de súmulas por visão IA, autenticação OTP via WhatsApp (WAHA) e modo PWA.

Resultados: ~890 assinantes em cerca de 6 meses no modelo freemium, suíte E2E e unitária (Playwright/Vitest) e evolução contínua em produção com suíte automatizada de testes.

Stack principal: TypeScript, React 18, Vite, Tailwind CSS, shadcn/ui, Express 5, Prisma 7, PostgreSQL, IA Vision, WAHA, Resend, PWA, Vitest, Playwright

---

CASO 04 · TELECOM & CRM · 2026 ● Atualizado (inclui Deskcomm & WAHA centralizado)

Barigui Telecom (Atendimento IA) & Deskcomm CRM / Gateway WAHA

Automação de atendimento omnichannel com LLMs locais em infraestrutura do cliente e unificação de WhatsApp corporativo.

Papel	Ambiente	Canais
Arquiteto & Líder de Entrega	**Proxmox do cliente / MMSoft**	**WhatsApp e Voz**

Contexto: Uma operadora de internet e TV precisava automatizar o atendimento no WhatsApp e por voz, integrado aos sistemas de gestão do provedor, sem ambiente de sandbox; internamente, a MMSoft precisava consolidar o CRM e os canais de WhatsApp.

Liderança & Papel: Conduzi o kickoff e o plano de entrega em duas ondas (piloto de WhatsApp e, em seguida, voz), além da transferência de conhecimento com o time do cliente.

Arquitetura: Agentes LLM com análise de sentimento e base RAG rodando em modelos locais no Proxmox do cliente; na MMSoft, migração do Deskcomm (Supabase/PostgreSQL) para VM dedicada com TLS e desenho do gateway WAHA centralizado.

Resultados: Piloto de WhatsApp estruturado em produção, com agentes locais e isolamento de acesso administrativo.

Stack principal: WhatsApp Business API / WAHA, n8n, LLMs locais, RAG vetorial, Supabase, PostgreSQL, Proxmox VE, Traefik, MCP

CASO 05 · AGENTE DE VOZ & WHATSAPP · NIGÉRIA · 2024–2025 • Já no portfólio

Tracy AI Agent — Concierge Conversacional para Secrets Palace

Atendimento autônomo em voz e WhatsApp para reservas VIP em operação noturna internacional.

Papel	Mercado	Integração
Arquiteto de Solução & Dev	**Nigéria (Inglês)**	**8 ferramentas MCP**

Contexto: Casa noturna de alto padrão recebia pedidos de reservas VIP durante a madrugada exigindo validação estrita de horário, cálculo de sinal e pagamento em tempo real.

Liderança & Papel: Projetei toda a arquitetura conversacional, contratos de integração, guardrails de produção e documentação completa de entrega para o cliente na Nigéria.

Arquitetura: Fluxo de 10 etapas orquestrado no n8n com síntese de voz ElevenLabs e 8 ferramentas MCP, com as regras de negócio (horário de funcionamento e cálculo de sinal) aplicadas no código das ferramentas, e não apenas no prompt.

Resultados: Sistema entregue pronto para produção com relatório técnico completo; o padrão arquitetural (LLM + MCP determinístico + n8n) tornou-se referência reutilizada nos projetos seguintes.

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

CASO 06 · SAAS SOCIAL & ENTERPRISE · 2023–2025 • Já no portfólio

CUFA Rio de Janeiro (1.000h) & Instituto Coca-Cola Brasil (Falepaco)

Gestão multi-tenant de projetos sociais patrocinados e liderança de produto (CPO) em ecossistema educacional enterprise.

Papel	Escopo CUFA	ERP / CRM
Arquiteto / CPO (Falepaco)	**1.000 horas + revenda**	**Conta Azul, Salesforce, Bitrix24**

Contexto: Na CUFA-RJ, projetos patrocinados exigiam três cotações por item, aprovação contábil/financeira e conciliação com o Conta Azul; na Falepaco, a conta Coca-Cola precisava integrar jornada educacional de alunos.

Liderança & Papel: Na CUFA, liderei o escopo de 1.000 horas e a entrega em fases, reaproveitando o control plane do BIPo. Na Falepaco (2023–2025), atuei como CPO conduzindo times distribuídos.

Arquitetura: Plataforma CUFA reaproveitando o control plane multi-tenant do BIPo com integração bidirecional Conta Azul; ecossistema Coca-Cola conectando Salesforce, Bitrix24 e WhatsApp via n8n.

Resultados: Escopo, plano financeiro e sustentação pós go-live validados na CUFA; entregas enterprise consolidadas na conta Coca-Cola com coordenação internacional.

Stack principal: SaaS multi-tenant, n8n, API Conta Azul, Salesforce, Bitrix24, WhatsApp, PostgreSQL, Docker

CASO 07 · CLIENTES INTERNACIONAIS · SUÉCIA & EUA · 2024–2026 • Já no portfólio (atualizado com consultoria 2026)

Sports Rekrytering (Suécia), GBA Facilities & MetaGami (EUA)

Integrações financeiras OAuth2, pipelines de prospecção B2B com IA e implantações reguladas nos Estados Unidos.

Mercados	Compliance	Entrega
Suécia e Estados Unidos	**SEC / SAML / RBAC / Auditoria**	**Middleware, IA e QA**

Contexto: Clientes fora do Brasil exigiam aderência a ERPs locais (Fortnox na Suécia), automação confiável de prospecção B2B e requisitos rígidos de segurança para gestoras financeiras nos EUA.

Liderança & Papel: Conduzi o desenho técnico, a estabilização de pipelines, o briefing de segurança para prospectos nos EUA e um diagnóstico técnico em código, com estimativa comercial em inglês, para o cliente norte-americano (MetaGami).

Arquitetura: Middleware OAuth2 multi-tenant da API Fortnox com CI/CD GitHub Actions; pipeline Apify + GPT-4o-mini + Pipedrive com controle de timeout/batch; arquitetura self-hosted com criptografia, SSO SAML, RBAC e inferência local.

Resultados: Plataforma sueca bilíngue (inglês e sueco) em operação, pipeline comercial estabilizado e diagnóstico técnico entregue ao cliente nos EUA.

Stack principal: Fortnox API, OAuth2, n8n, Apify, GPT-4o-mini, Pipedrive, GitHub Actions, Docker, SSO SAML, RBAC, criptografia

---

CASO 08 · GROWTH ENGINEERING & ADS API · 2025–2026 • Novo (ausente nos documentos atuais)

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

Arquitetura de conversão LGPD sem duplicação aliada à automação segura da Google Ads API com travas programáticas.

Papel	Compliance	Qualidade
Arquiteto de Growth & Dev	**Consent Mode v2 / LGPD**	**162 testes automatizados**

Contexto: Funis de infoprodutos digitais sofriam com bloqueio de cookies, dupla contagem de compras e necessidade de operar campanhas de Google Ads via API sem risco de alteração acidental de orçamento.

Liderança & Papel: Projetei a arquitetura de checkout dinâmico e mensuração desacoplada (Eliana Severino Receitas e TMBJ), além de especificar as travas de segurança do cliente de escrita da Google Ads API.

Arquitetura: Checkout dinâmico Asaas (preservando afiliados e cupons) com entrega via Deno/Supabase e Resend, disparo server-side via GA4 Measurement Protocol (source=mp) e Meta CAPI, Consent Mode v2, SEO/GEO (llms.txt) e cliente Python google-ads-writer com Service Account e 6 travas (criação apenas PAUSED, allowlist, dry-run).

Resultados: Validação E2E automatizada em sandbox Asaas via Playwright, zero duplicação de eventos de compra, 162 testes aprovados na suíte Ads e campanha Search lançada em modo especialista.

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

CASO 09 · DADOS EM ESCALA & MENSAGERIA · 2025–2026 • Novo (ausente nos documentos atuais)

MMSoft Mailer (v0.4.2) & Painel-CNPJ (~2,7 Milhões de Empresas)

Plataforma própria de e-mail transacional/cadências integrada a módulo analítico de segmentação B2B em larga escala.

Papel	Escala	Entregabilidade
Arquiteto & Líder de Produto	**~2,7 milhões de CNPJs**	**SPF / DKIM / DMARC**

Contexto: A operação comercial B2B demandava independência de plataformas caras para envio cadenciado com alta entregabilidade, além de cruzamento rápido sobre base cadastral empresarial de milhões de registros.

Liderança & Papel: Conduzi o roadmap evolutivo (v0.3.2 → v0.4.2), a revisão de segurança e RBAC e a governança de DNS de envio em subdomínio dedicado.

Arquitetura: Stack containerizada (API, Worker, Edge, Web, PostgreSQL, Redis) integrada à API/webhooks da Resend, 2FA TOTP, CLI administrativa, cadências D0/D3/D7/D14 com supressão automática por resposta, mais API de filtro em lote no Painel-CNPJ com perfil SDR e limites de memória.

Resultados: Prospecção e mensageria transacional centralizadas sob infraestrutura própria com autenticação completa de domínio (SPF/DKIM/DMARC/ImprovMX) e segmentação instantânea sobre ~2,7M CNPJs.

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

CASO 10 · WEB GAMES & RECUPERAÇÃO CRÍTICA · 2026 • Atualizado (Next.js 14 / 14 jogos)

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

Evolução de plataforma de jogos bilíngue para AdSense/SEO e resgate seguro de faturamento recorrente em produção.

Papel	Jogos	Foco
Arquiteto & Tech Lead	**14 jogos (PT/EN)**	**SEO, AdSense e billing**

Contexto: O Acerte a Mosca precisava adequar SEO, canonical/hreflang e a colocação de anúncios às políticas do Google AdSense, além de unificar os domínios PT/EN; o LotoJoy tinha a cobrança recorrente inoperante em produção e dependências desatualizadas.

Liderança & Papel: Orquestrei a auditoria técnica (Dev, Code Review e QA) do Acerte a Mosca, com prévia isolada para validação, e conduzi o plano de correção do LotoJoy com backup automático antes da publicação.

Arquitetura: Acerte a Mosca reestruturado em Next.js 14 (App Router), next-intl, Pixi.js 8, Prisma/PostgreSQL e servidor WebSocket unificado (Railway + Cloudflare), corrigindo canonical/hreflang e posicionamento de anúncios; LotoJoy com branch fix/asaas-cobranca-conciliacao e npm audit fix.

Resultados: Plataforma Acerte a Mosca com 14 jogos integrados (incluindo Kung Fu Castle) e trabalho de SEO/AdSense em andamento; no LotoJoy, cobrança recorrente reestabelecida e dependências atualizadas antes da publicação.

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

CASO 11 · HARDWARE FPGA, EMULADORES C++ & JOGOS 3D · AGO–OUT/2026 • Novo (ausente nos documentos atuais)

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

Engenharia de sistemas de baixo nível cobrindo síntese FPGA, emulação em C++, compilador/montador próprio e 4 jogos 3D.

Papel	Hardware	Qualidade
Engenheiro de Sistemas & Autor	**FPGA Gowin / V9968**	**160 testes + RTL check**

Contexto: Explorar os limites de arquitetura de 8 bits com o novo coprocessador gráfico 3D V9968/geo3d v2.0, unindo hardware reconfigurável, emuladores e software real rodando em ciclo exato.

Liderança & Papel: Liderei o projeto ponta a ponta em repositórios abertos no GitHub: do floorplanning físico em FPGA e Pull Requests em emuladores C++ até o ferramental de build e quatro títulos 3D jogáveis.

Arquitetura: Contribuições Verilog/RTL (V9968_Cartridge — posicionamento de 20 BSRAMs e fechamento de timing em FPGA Gowin); implementação do geo3d v2.0 em C++ no openMSX e blueMSX-plus; montador Z80/MegaROM próprio em Python 3.13 (msxAIbuilder) com WebMSX headless; jogos MSXDOOM, MSXWOLF, GeoFighter e Cube Maze.

Resultados: 70+ commits no V9968_Cartridge com zero negative slack, 160 testes passando no msxAIbuilder, equivalência byte-a-byte em 3.755 frames contra testbench Verilog no Cube Maze e 30 fps (420 vértices) no GeoFighter.

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

---

CASO 12 · INFRAESTRUTURA SELF-HOSTED · 2024–2026 • Já no portfólio (atualizado)

Plataforma Híbrida Proxmox VE + Docker Swarm + Headscale Mesh VPN

Arquitetura compartilhada de baixa latência e custo previsível sustentando dezenas de serviços produtivos e homologação.

Papel	Topologia	Foco
Arquiteto de Plataforma	**Proxmox + Swarm + VPN**	**Custo previsível e isolamento**

Contexto: Operar mais de uma dezena de produtos SaaS e ambientes isolados de desenvolvimento com margem saudável, controle de dados (LGPD/soberania) e separação estrita entre produção e testes.

Liderança & Papel: Conduzi auditorias de prontidão comercial, políticas de backup e snapshot, segmentação de acessos por VPN privada e padronização de publicações com plano de rollback.

Arquitetura: Cluster Proxmox VE com armazenamento ZFS e backups automatizados, Docker Swarm com Traefik e Portainer, ambiente de desenvolvimento em Debian e rede mesh privada Headscale/Tailscale com DNS interno, além de WebRTC (MiroTalk SFU) e entregabilidade via WHM/Cloudflare.

Resultados: Operação consolidada de múltiplos produtos próprios e de clientes, com custo de infraestrutura previsível e ambientes administrativos isolados da internet pública.

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

Histórico Enterprise, Docência & Próximos Passos

Entrega Enterprise & Gestão de Operações (Stefanini, Agile Engine, Diebold, WWT)

- **Stefanini (2016–2021):** De desenvolvedor C# sênior a Delivery Manager — expansão de 35 para 136 profissionais e de US$ 1,5 mi para US$ 18 mi em receita; contas J&J, Honda, Embraer, CSN, Via Varejo, Raízen, Ultragaz, Ale Combustíveis e Grupo Aldo.
- **Marcos auditados:** Honda Connected Car (14 engenheiros, 5.000+ veículos, US$ 800 mil, 99,8% uptime); J&J Big Data (54 fontes, 99,7% confiabilidade); Via Varejo (15.000+ usuários, 600+ lojas, +40% garantias); CSN (26 APIs, 85%+ testes, 3 meses antes do prazo).
- **Agile Engine (2021–2022):** Technical Lead coordenando time multidisciplinar distribuído entre Argentina, Ucrânia e Califórnia (EUA) em entregas C#.
- **Diebold (2010–2011) e World Wide Technology (engajamento de três meses):** Como Desenvolvedor .NET Sênior na Diebold, captação de dados preditivos e interface de ATMs para o Bank of America; na WWT, integrações Boomi em gestão de ativos.

Formação, Docência, Palestras & Comunidade

- **Formação & Certificações:** MBA Executivo em Administração (FGV, 2022–2024); Gestão de Negócios (FATEC-SP, 2016–2020); Análise e Desenv. de Sistemas (Unitins, 2008–2013); SAFe 5 Certified, DEPC, Empretec (Sebrae), Inovação na Era Digital (Mackenzie).
- **Docência & Palestras:** Professor convidado de IA e Automação na FAAP; instrutor executivo na lideres.ai (curso de 2 dias e workshop de 6h em agentes/n8n); palestrante no AI Brasil Experience; especialista na rede Guidepoint.
- **Comunidade:** Ensino voluntário de programação para comunidades carentes e informática para terceira idade; autor do conto “Tracy, a Voz que Não Tem Corpo”.

Vamos conversar sobre liderança técnica, arquitetura e IA

Aberto a posições remotas no Brasil e exterior como Head of Engineering, Engineering Manager, Technical Lead ou AI Solutions Architect. Posso apresentar em profundidade a arquitetura, os trade-offs e os resultados de cada caso.

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