

FRED W. NIX JR.

Principal Solutions Engineer | Infrastructure & AI
Platforms

Atlanta, GA · U.S. Army veteran
frednix@gmail.com · (404) 953-8305
linkedin.com/in/frednix · github.com/nixfred ·
resume.nixfred.com

SUMMARY

Principal solutions engineer for AI infrastructure with a managed book of business: pre-sales through post-sales, demos, custom code, and the executive business case. Twenty-plus years from EMC and Dell to Pure Storage and BlueAlly.

PROFESSIONAL EXPERIENCE

Senior Solutions Architect · BlueAlly Technology Solutions, Atlanta, GA

Apr 2026 – Oct 2026

Pre-sales and post-sales solutions architect supporting six account executives across the Southeast (enterprise, healthcare, higher education, public sector, nonprofits). Owned discovery, architecture, sizing, bills of materials, cost and return models, proposals, proof-of-concept design, demos, and custom code on 24 engagements in five months. Closed ~\$150K in the first 90 days and built a seven-figure qualified pipeline. Customers anonymized; figures real.

Pre-sales and book of business

- **Moved a global travel company off VMware across three continents:** discovery, hosted proof of concept, licensing strategy, and a parallel three-site migration plan ahead of staggered Broadcom renewals; stretched storage over NVMe/TCP on all-flash arrays avoiding ~\$500K of Fibre Channel fabric.
- **Modernized a 190+ property hotel chain:** 480-core VMware entitlement review that closed as a signed \$95K order in the first 90 days, legacy storage retirement (34 TB), flash capacity planning at 91% utilization, backup and recovery discovery, five-year cost model for leadership; storage follow-on in flight at ~\$210K gross profit.
- **Built the data center bill of materials** (servers, storage, firewalls) for an immersion-cooled colocation operator: 250,000 sq ft, three halls, 18 MW with a 24.5 MW site cap, GPU hosting relationship behind it.
- **Designed and priced further infrastructure:** hyperconverged-to-disaggregated refresh for a media ministry; 215 TB, 130-camera video archive for a state water district; all-flash storage proposal for a major airport; voice trunking proposal for a state university; 48-host virtualization cost model; hypervisor migration evaluation for an aerospace contractor.

Post-sales, security, and identity

- **Moved wireless authentication** for a 100-site HIPAA healthcare provider (600+ daily users) from on-premises servers and 100 branch VPN tunnels to a cloud service with device certificates managed for them.
- **Ran the identity and access program** at a global agriculture company (150,000 employees, 77 countries, ~2,000 legacy apps): Microsoft security posture, secure access from anywhere, badge and login convergence, AI governance. Caught a managed-services decommission that would have breached contracted minimums at a higher-education account; preserved revenue and standing.

Custom code, demos, and AI projects

- **Private GenAI reference architecture:** 2U four-GPU node (4× 96 GB, 384 GB per node), private AI appliance stack, orchestration and governance layers, sized to ~14M peak monthly tokens; packaged as a fixed-price 60 to 90 day proof of concept so a CIO could buy without a capital budget cycle.
- **Custom demos and code:** customer-specific demos and prototypes in ~1.5 days with Claude Code and reusable AI workflows; an internal sales intelligence system in TypeScript (append-only event log, searchable memory, MCP server) turning voice memos, notes, and email into opportunity tracking.
- **AI use-case portfolios and workshops:** nine use cases for an accredited research institution (~\$12.1M modeled annual value, 17,800 reclaimed hours, appliance sized 109 to 343 TB, five-year lease above \$2M); six use cases for an academic medical center (\$2.4M modeled savings); contract intelligence agents for a healthcare payments company (\$290K to \$305K proposal); Copilot Studio vs agent platform evaluation for an office real estate trust (\$6.7M perceived value); signed AI readiness and Copilot Studio statements of work (~\$55K) at an engineering firm; statement of work owner for a \$62.5K manufacturer-funded AI program. Coached engineers on Claude and MCP; monthly tech talk (~50 attendees).

AI Solutions Architect, Storage & Compute Practice Manager · Business Technology Architects, Alpharetta, GA

Mar 2025 – Apr 2026

- Built a production-grade, air-gapped AI inference platform for channel partners on modular blade compute, 48 GB GPU clusters, all-flash storage, and 100GbE. One-command deployment took a partner from bare metal to running inference in under an hour instead of days of manual build, with 100% success across partner rollouts.
- Multi-model serving (LLaMA, Mistral, custom fine-tunes) with vLLM, quantization, tensor parallelism, and hot-swap switching; OpenAI-compatible API and streaming chat for retrieval-augmented apps; live GPU metrics via DCGM, Prometheus, Grafana. Partner enablement through hands-on labs and implementation guides.

Enterprise Sales Engineer · Pure

2021 – 2023

Storage

Fortune 100 accounts, Southeast. 250% of quota in 2021, named Global Rookie of the Year. 100% uptime for a customer through a hurricane disaster recovery event.

Partner, VP Sales · EMOTO Electric

2023 – 2025

Motorbikes

Entrepreneurial operating role; sales up 225% in year one, expanded the product line, scaled to regional market leader.

Earlier career: EMC Corporation, Global Presales Operations & Technical Enablement (2010 – 2015): operations for 5,500 presales engineers, trained 1,000+, grew vSpecialists from 30 to 150+. Before that, CTO, solutions consulting, presales, and professional services roles. Annual sales quota achieved every year of career.

Solutions Architect, Converged Systems · Dell Technologies

2019 – 2020

Technical pre-sales for software-defined storage, cloud architecture, and data center consolidation.

Global Presales Director · Reduxio

2015 – 2019

First American employee; led US market entry. 150%+ of quota annually, 90%+ conversion across 50+ proof-of-concept projects.

SELECTED AI ENGINEERING PROJECTS · resume.nixfred.com/projects**Agentic engineering platform**github.com/nixfred

Four-host Claude Code platform with MCP integrations and local-first model routing (a 12 GB local GPU before any cloud call). Persistent agent memory in TypeScript/Bun with SQLite full-text plus vector search; lifecycle hooks that enforce coding standards, secret scanning, and pre-push guards; written operating rules the agents follow; scheduled recovery of missed session extraction.

AI audit trail and governance

Every agent session is committed to git with its transcript, decisions, and rejected approaches, then indexed for search, so any answer an agent gave can be traced months later. Paired with least-privilege agent permissions and an evaluation workbench for critical failures, coverage gaps, and drift: the audit trail businesses will expect from agents as from any system of record.

AI software factory

Reusable AI-assisted delivery workflow (TypeScript/Bun, Python, Cloudflare Workers/Pages, GitOps): build spec in, live production site out, end to end. 10+ production sites and 30+ desktop plugins shipped; the same tooling produces customer demos in ~1.5 days.

AI Systems Workbenchtools.nixfred.com

Fourteen browser-based, on-device instruments for designing and evaluating AI applications: prompt lab, token and cost planner, retrieval lab, evaluation workbench, agent designer, permission planner, failure investigator, drift monitor, AI stack mapper, and more.

Open source contributor

30+ published desktop plugins for Linux (Omarchy, Arch-based) and 120+ public repositories: custom development for hardware control, media, messaging, and system monitoring.

SKILLS

Infrastructure: two-socket and four-GPU rack servers, modular blade compute, all-flash NVMe arrays, scale-out NAS, VMware vSphere/VCF, hyperconverged and disaggregated designs, private cloud automation, Proxmox, NVMe/TCP, 100GbE, SD-WAN, backup and disaster recovery, colocation and immersion cooling, virtual desktops, Linux (Arch, Ubuntu), Bash, systemd, Docker, K3s, Tailscale.

Inference & GPU: GPU sizing from 48 GB and 96 GB cards to 384 GB nodes and clusters; private AI appliance stacks; vLLM, Ollama, multi-model serving, quantization, tensor parallelism, air-gapped AI, DCGM, Prometheus, Grafana.

Security: zero trust, certificate-based wireless authentication, air-gapped deployment, HIPAA and PCI scoped design, identity and access management, least-privilege agent permissions, AI governance.

Solutions engineering: technical discovery, solution strategy, use-case portfolios and value modeling, proof-of-concept design and evaluation, demos and custom prototypes, executive presentations, competitive positioning, cost and return models, bills of materials, proposals and RFP responses, lease and pass-through structures, manufacturer funding programs, partner enablement.

Agents & developer platforms: AI agent workflows, Claude Code, MCP, TypeScript/Bun, Python, Cloudflare Workers/Pages, OpenAI-compatible APIs, retrieval-augmented generation, LangChain, LlamaIndex, Copilot Studio, Git/GitOps, open source Linux desktop plugin development.

Education: B.S. Business Administration, Information Systems, Auburn University.