Vivek Verma

Full Stack Engineer

Haridwar (U.K), India
+91 9027519437 · imvivekvermaa@gmail.com
github.com/imvivekvermaa · imvivekvermaa.in
imvivekvermaa.in/log

SUMMARY

Engineer who ships 0-to-1 production systems and works a layer below the app. Built a real-estate platform serving approx. 30K daily visitors; led an offline-first POS coordinating 11 terminals and a 9-month enterprise PLM system, both delivered end to end with direct client ownership. Comfortable where performance actually lives: diagnosed a 15-year-old legacy schema and cut its multi-minute reports to under 3 seconds. Now going deeper into LLM inference performance — benchmarking, KV-cache behaviour and serving cost — publishing the measured work daily at imvivekvermaa.in/log.

PROFESSIONAL EXPERIENCE

Stacx24 — Noida, India **June 2024 – June 2025**
Full Stack Developer

Led **3 production systems end to end**, with direct client ownership on each.

- **Led a 9-month Product Lifecycle Management build** for a textile manufacturer covering design → production → billing, managing a 2-engineer team and client communication directly. Serverless on **AWS (Lambda, RDS, Cognito, Amplify)** with hierarchical authorization and **WebSocket** real-time updates between departments.
- **Diagnosed the root cause of multi-minute report queries in a 15-year-old production database** — missing primary/foreign keys, untracked table relationships and no indexing. Proposed and drove the normalization + indexing plan, cutting those queries from several minutes to **under 3 seconds**.
- **Architected an offline-first Point of Sale system across 11 terminals**: a local database server powers every connected counter, sharing state in real time with role-based authorization — the business keeps selling through internet outages.
- **Initiated PropMatters** (real-estate platform): database schema design, **ISR/SSR** rendering strategy and **TanStack Query** prefetching for load performance, email/Google/LinkedIn authentication, and the property listing experience — multi-filter search including **radius-based filtering on an interactive map**.

PropMatters — Pune, India (freelance) **Feb 2026 – May 2026**
Full Stack Developer · listed on the company team page: propmatters.in/propmatters-teams

- **Returning client from Stacx24**, where I initiated the platform. Now serving approx. **30K daily visitors and 2,000+ registered users**.
- **Shipped LLM-based listing intake in production**: PDFs, images and free text → structured property records, removing manual data entry from the listing workflow.

SKAAY Data Labs — Noida, India (freelance) **Aug 2025 – Sep 2025**
Full Stack Developer

- Built an internal analytics dashboard over a set of **Infor CloudSuite** data APIs, giving the team one interface to explore curated data across multiple endpoints instead of querying each service separately.
- Implemented authentication and role-based authorization so each user only reached the data sources their permissions allowed.

PROJECTS

Where — real-time personal-safety app (in testing) **Nov 2025 – Present**
React Native (Expo) · Node/Express · PostgreSQL · SQLite · Socket.IO · FCM

- **Offline-first architecture**: SQLite cache, delta-sync orchestrator and an exponential-backoff operation queue — chat, contacts and notifications work with no connectivity.
- **Idempotent message delivery** via client-generated UUIDs over a hybrid **REST + WebSocket** sync layer with since-based delta fetching and 7-day rolling retention.
- Real-time SOS alerts delivered across **foreground, background and killed** app states, plus live location sharing over Socket.IO.

LLM Inference Engineering (current — updated weekly) **Aug 2026 – Present**

- **imvivekvermaa.in/log** — daily, link-backed public record of measured inference work: benchmarking, KV-cache behaviour, serving cost per million tokens.

SKILLS

Languages	TypeScript, JavaScript, Python, SQL
AI / LLM	OpenAI & Anthropic APIs, RAG, agent tooling, structured extraction from documents. Currently going deep on: vLLM, KV cache, quantization, inference benchmarking
Backend	Node.js, Express, PostgreSQL, MySQL, MongoDB, Drizzle & Sequelize ORMs, WebSockets
Infra	AWS (Lambda, RDS, Cognito, Amplify), Docker, serverless, self-hosted Supabase, Git
Frontend	Next.js, React, React Native (Expo)