Omar Gamal

Backend, Distributed Systems & AI Infrastructure Engineer | Go, Python, C++, Kubernetes, Terraform
Cairo, Egypt (UTC+2/+3) | +20 106 444 4598 | omargamal.dev@outlook.com
github.com/Omar0Gamal | linkedin.com/in/omar-gamal-091044168 | omar0gamal.vercel.app
Remote | Fluent English | Full-Time Available | Full EU & US Morning/Afternoon Overlap

Summary

Backend and Distributed Systems Engineer with 4+ years of production experience building low-latency microservices, cloud-native infrastructure, and high-throughput data engines in Go, Python, and C++. Designed a 14-service ERP/CRM architecture, operated 500+ client Linux production clusters (60% crash reduction), and built a C++17/CUDA graph-vector engine reaching 33,000+ QPS. Open-source contributor to llama.cpp.

Technical Skills

Languages: Go, Python, C++ (C++17), C#, Java, TypeScript, SQL, CUDA
Backend & Distributed Systems: gRPC, Protocol Buffers, REST, Gin, FastAPI, ASP.NET Core, sqlc, Saga, Transactional Outbox, Idempotency, Modular Monolith, Multi-Tenancy
Data & Messaging: PostgreSQL, Redis, NATS JetStream, SQLite, MongoDB, MySQL, Vector Indexing (HNSW, cuVS)
Cloud & Observability: AWS, Cloudflare (Workers, KV, R2), Terraform, Kubernetes (k8s/k3s), Helm, Docker, GitHub Actions CI/CD, Linux, Caddy, Prometheus, Grafana, Loki

Experience

Backend & DevOps Engineer **Jan 2022 – Mar 2026**
Attila Gaming *United States (Remote)*

- Engineered 15+ Java backend services and low-latency matchmaking modules for a multi-server gaming platform with 500+ concurrent clients; built automated instance spawning and packet-level anti-cheat detection, cutting latency by 40%.
- Owned production Linux infrastructure: GitHub Actions CI/CD, Docker deployments, off-site backups, and Prometheus/Grafana/Loki observability, reducing crashes by 60%.

Lead Backend Architect (Contract) **Sep 2025 – Nov 2025**
Ayist Group *Poland (Remote)*

- Architected a 14-service distributed platform (ERP, CRM, Document Management): domain boundaries, async event flows, and gRPC/Protobuf contracts with idempotent failure recovery, across a 6-engineer Go, Python, and React team.
- Built Terraform and Helm deployment pipelines for production Kubernetes clusters on Hetzner Cloud.

Projects

AxiomGraph – Hybrid Graph-Vector Database | *C++17, CUDA, Python, SQLite* github.com/Omar0Gamal/AxiomGraph

- Built an embedded C++17 engine with lock-striped CSR graph traversal, CPU HNSW, and NVIDIA cuVS GPU indices: 33,000+ QPS on GPU (8,600 on CPU) with 1M nodes in under 2 GB RAM.
- Reached 68.4% All-Gold-Found@10 on HotpotQA (+22.6% over dense+rerank baselines); exposed zero-copy Python bindings via `nanobind` and a concurrent HTTP server.

Cartwright – Distributed Saga Checkout Engine | *Go, C#, Python, gRPC, NATS, k8s, Terraform*
github.com/Omar0Gamal/cartwright-saga

- Built a Go Saga orchestrator, C# billing service, and Python worker over gRPC/NATS JetStream with transactional outbox, idempotency keys, backoff, and lease-based recovery.
- Verified zero double-charging via chaos tests that kill billing pods mid-flight; deployed to k3s with Terraform and Helm.

Nexus Commerce – Multi-Tenant Commerce Backend | *Go, sqlc, PostgreSQL, Redis* github.com/Omar0Gamal/Nexus-Commerce

- Built a Go modular monolith (25 domain modules, one distroless binary) with compile-time SQL (`sqlc`, `pgx`), Redis job queue, TTL inventory locks, and rate limiting; cross-domain calls run in one DB transaction.
- Designed edge multi-tenancy with a Cloudflare Worker + KV and a Caddy-verified shared-secret header; integrated Paymob payments and R2 media.

llama.cpp – RPC Server Bounds Validation (Open Source) | *C++* github.com/ggml-org/llama.cpp/pull/26933

- Fixed two critical out-of-bounds vulnerabilities (`GGML_OP_GET_ROWS` read, `GGML_OP_SET_ROWS` write) in the distributed RPC server via pre-execution tensor shape validation.

Education

Capital University **Cairo, Egypt**
B.Sc. in Computer Science | GPA: 3.92 / 4.0 (Dean's List)