

Shubham Tanwar

Backend & Systems Engineer · distributed systems, event-driven platforms, agent infrastructure

shubham@oryxa.in · theweirdlyodd.dev · linkedin.com/in/shubham-tanwar04 · github.com/Shubham04-oss

SUMMARY

Engineer who designs and builds systems end to end in Go, Rust and TypeScript: an event-driven multi-service SaaS platform with hybrid RAG and LLM classifiers, a calibrated cloud-architecture simulator, and an open-source runtime that puts people and several coding agents in one live room. Took an agent product to a paid pilot with 12 design partners. Strongest in stateful and event-driven backends, real-time collaboration and agent infrastructure; every system ships with its trade-offs and limits written down.

EXPERIENCE

Oryxa AI · Founder

Apr 2026 – Oct 2026

Pre-launch platform for teams working with AI agents. Go, Rust, TypeScript, AWS, Cloudflare.

- Architected a Go monorepo of ~12 event-driven services (auth, teams, agents, files, RAG, billing, realtime and others) that build into one static binary and deploy together or as separately scaled processes, selected by config; roughly 750 test files across Go and TypeScript.
- Gave each service its own SQS queue with a dead-letter queue and a Postgres outbox; versioned events (e.g. `auth.member.removed.v1`) are idempotent by event ID, and production refuses to boot while any required queue is missing instead of silently dropping deletions.
- Built real-time collaboration on Cloudflare Workers and Durable Objects over WebSockets, including a guest-to-account handoff that preserves workspace context. Go services reach the Worker through HMAC-signed, timestamped calls that purge private conversations when membership changes.
- Designed the delegation model: a team conversation is compressed into a typed brief, authority attaches only when a human commits it, and a per-user agent on Durable Objects runs it through durable Workflows, chasing follow-ups on its own alarms; events carry identities and state, never content.
- Built the RAG layer: hybrid retrieval (pgvector plus Postgres full-text, fused with reciprocal rank fusion, then reranked) over corpora that pin their embedding model and chunk policy, with pgvector, S3 Vectors or Bedrock knowledge bases as backends. Indexing runs with the member's own credentials, so no one can index what they cannot read.
- Built ingestion from GitHub, Notion, Slack, Drive and Gmail through Composio, with no tokens held by the service, and a Gemini judge that keeps promotions and noise out of the index.
- Built a privacy classifier for agent-to-agent answers: a provenance gate skips the model when no private tool ran; otherwise a small temperature-zero model flags private categories the answer carries that the question did not ask for, backed by an eval suite.
- Streamed concurrent agent output with per-workspace token budgets and Valkey-backed rate limiting.
- Prototyped a Rust API with BetterAuth: EdDSA tokens verified locally against cached JWKS; roles read at use time so a demotion applies at once.

ComStack · Co-founder

Oct 2025 – Feb 2026

Multiplayer agent teams for commerce, early in the category. Paid pilot with 12 design partners.

- Built Synq, a multi-tenant commerce platform: Go ops, commerce and agent services on Cloud Run over tenant-scoped Postgres, with pgvector for semantic catalog search and SKU classification and Valkey for caching and rate limits.
- Orchestrated agent workflows on Temporal with child workflows and tenant context propagated in workflow headers; async events over GCP Pub/Sub and Eventarc with dead-letter handling.
- Attacked inference cost: one agent operation consumed ~600k input and 30k output tokens. Noise-reducing classification and open-weight models cut it, but multiplayer traffic multiplied it faster than pricing could cover, so we wound the pilot down rather than keep subsidising usage.

Xtride · Systems Architect (Contract)

Sep 2025 – Oct 2025

- Built backend and spatial data systems for real-time driver telemetry on PostgreSQL/PostGIS, with H3/S2 indexing for dispatch and ETA pipelines.
- Built event-driven GPS ingestion with idempotent trip states; tuned pooling, queries and edge caching.

PROJECTS

Tectonic · cloud architecture simulator

Rust · Go · TypeScript

Private repository · demo and code walkthrough on request

- Deterministic discrete-event engine simulating AWS components (load balancers, RDS/Aurora, DynamoDB, Valkey, Lambda, SQS/SNS, autoscaling, failover) and Python, Node and Go runtimes on one simulated clock.
- Answers “where does it break, and why” as a range from a five-method ensemble (e.g. breaks at 400–450 rps because the database pool fills), streamed live over SSE to a web app.
- Go importer turns OpenTofu/Terraform files or plan JSON into runnable models, surfacing unknown inputs (traffic, CPU per request, query cost) as explicit gaps rather than hidden defaults.
- Calibrated against real hardware (Neoverse N1): on held-out experiments, p99 predictions within $\pm 25\%$ improved from **2/19 to 13/19** levels, with misses published alongside hits.

Oryxa Runtime · open-source multiplayer AI runtime

Rust · Apache 2.0

oryxa-vault.github.io/oryxa · github.com/Oryxa-Vault/oryxa · v0.7 pilot

- Puts several people and coding agents (Claude Code, Codex, OpenCode) in one live room without modifying any of the CLIs; one-command install, with browser and terminal views.
- An append-only event log makes late join, replay, recovery and audit one mechanism; each agent streams on its own lane, a cursor into that log, and rooms survive restarts.
- Declarative YAML connectors for ACP agents and frameworks (ADK, CrewAI, LangGraph, PydanticAI), with a probe command that runs a real turn before an agent joins a room.

SKILLS & EDUCATION

Languages Go, Rust, TypeScript, Python

Infra & data AWS (SQS, SNS, EventBridge), GCP (Cloud Run, Pub/Sub), Cloudflare Workers, Durable Objects & Workflows, OpenTofu/Terraform, PostgreSQL/PostGIS, Valkey/Redis, vector search, H3/S2

AI systems Hybrid RAG (pgvector, full-text, RRF, reranking), LLM classifiers and evals, Bedrock, Vertex AI/Gemini, open-weight models, LangGraph, ADK, ACP, MCP

Systems Event-driven architecture, transactional outbox, idempotency, Temporal, WebSockets/SSE, discrete-event simulation, agent runtimes

Education Computer Science, undergraduate studies, through Aug 2025