

RAYER TUNG

AI-Native Staff Backend & Platform Engineer

Taipei, Taiwan | rayershiih@me.com | +886 983 753 408
linkedin.com/in/rayer-tung-56148850 | github.com/Rayer

PROFILE HIGHLIGHTS

- **AI-native engineering:** Builds multi-tenant, source-grounded LLM knowledge products and led a department-selected concept-oriented RAG approach that turned fragmented, long-lived enterprise documentation into coherent knowledge entries.
- **Developer platforms & adoption:** Shipped production API/SDK and compatibility platforms adopted by three internal teams at TSMC and 7+ business units at Binance, replacing legacy middleware and brittle UI-driven workflows.
- **Migration automation:** Automated rolling migration of 1,400+ production VMs; reduced a typical 6–7-node migration from two engineer-days to 1–2 largely unattended hours, with >99% first-run success and automatic retries.
- **High-scale platforms & cloud economics:** Built and operated systems at 20K–60K QPS and reduced product-attributed cloud costs by an estimated 65% through workload-aware cloud modernization.
- **Distributed reliability & hands-on ownership:** Designed and implemented placement, failover, and rebalancing; co-designed and implemented production transaction coordination and compensating recovery. Works hands-on from architecture and implementation through rollout, incident response, and multi-team adoption.

TECHNOLOGIES

Go, Java, Rust, Python, C++, SQL/NoSQL, Spring Boot, REST, Kafka, AWS, GCP, Kubernetes/GKE, Docker, Helm, Jenkins, Argo CD, CloudFormation, Lambda, DynamoDB, Cassandra, BigQuery, ELK, Grafana, Prometheus

PROFESSIONAL EXPERIENCE

TSMC — Section Manager, Fab Production Platform

Taipei & Hsinchu, Taiwan | Aug 2024 – May 2026

- Architected and engineered business-critical fab production systems from a shared Java/Go codebase across 60+ fab phases; reviewed critical changes and handled production incidents.
- Architected, implemented, and deployed a Java/Spring compatibility platform that continuously parsed evolving CORBA IDLs, converted nearly 1,000 operations across two major service domains into HTTP endpoints, and generated Springdoc OpenAPI specs plus Go/Java SDKs.
- The platform removed consumer-side IONA Orbix dependency while retaining CORBA behind a stable boundary and created a single future cutover point for a planned replacement broker/gRPC migration. This production-standard integration path has been adopted by three internal teams.
- Built an automated CentOS end-of-support migration platform for 1,400+ VMs to Ubuntu replacements across 6–12-node production groups, preserving exact WAR/application versions, configuration, and network/load-balancer behavior through rolling traffic shifts.
- Implemented the pipeline with Azure DevOps SDK/API orchestration, Python reconciliation, and PExpect; built and piloted in four weeks. A 6–7-node group moved from two engineer-days to 1–2 largely unattended hours, with >99% success on first automated run and transient provisioning failures recovered by automatic retries.
- Led a department-selected AI knowledge POC that compiled fragmented, long-lived domain documentation into concept-oriented LLM Wiki entries after fragment-level RAG failed to provide coherent system overviews.

Fusang — Technical Consultant, Fixed-Term Engagement

Remote, Taipei | Feb 2024 – Jul 2024

- Developed a Go digital-asset clearing engine covering matching, balance, settlement instructions, on-chain settlement, and reconciliation; delivered through UAT during the fixed-term engagement and subsequently entered production after the engagement.

Binance — Staff Software Engineer

Remote, Taipei | Jul 2022 – Jan 2024

- Owned the RFC, API contract, server integration layer, and Go/Java/Rust SDKs that exposed Themis's existing experiment lifecycle through programmable platform contracts, transforming a web-only A/B testing tool and replacing brittle UI-version-dependent Selenium integrations. Adopted by 7+ business units.
- Designed policy-bearing SDK behavior beyond wrappers: consumer-side prefiltering, local fallback, coordinated client/server caching, event routing, and pluggable metrics. Replaced a manual PM-operated KYC workflow with end-to-end SDK automation. Led requirement discovery, customer surveys, documentation, and onboarding.
- Supported the platform's latency-sensitive decision path at approximately 40,000 QPS while driving API/SDK integration across business units.

Appier — Senior Software Engineer — MLOps

Taipei, Taiwan | Feb 2021 – Jul 2022

- Owned the Dynamic Native Advertisement online recommendation, creative-assembly, and tracking platform at 20,000–60,000 QPS, around 100 ms P90, across US and Asia.
- Re-architected the Python/AWS Lambda serving path as Go services on GKE for a stable baseline plus 10–20x campaign traffic spikes; tuned KEDA autoscaling and gradual Route 53 canary traffic shifts, reaching comparable SLA and tail latency after stabilization.
- Partnered as the first production adopter for Infra's CloudWatch-to-self-hosted-ELK implementation. Billing tags captured all product-attributed AWS/GCP costs before and after; comparable traffic/ad-budget periods showed an estimated 65% reduction in product-attributed cloud costs.

Innova Solutions USA — Senior Software Engineer / Technical Lead, MCDS

Taipei, Taiwan | Jul 2017 – Feb 2021

- Directly implemented a Java HIPAA de-identification pipeline and representative test-data generation; established the project's first Jenkins CI pipeline.
- Co-designed and implemented the incremental migration of a company-hosted shared multi-tenant monolith into a shared AWS platform serving 50+ customers. Contributed the production transaction proxy, distributed locks, and explicit state machines for plugin-driven workflows across imperfect service boundaries; compensating actions restored prior state after partial failures.

Ruckus Wireless — Senior Software Engineer

Taipei, Taiwan | Oct 2012 – Jun 2017

- Primary IC and technical-spec author for a four-node controller platform formally rated to manage 30,000 APs, with production customers near the limit. Designed and implemented AP placement, overload redirection, failover, and rebalancing; remediated connection backlog and timeout issues found in scale certification.
- Partnered with PM and QA to define capacity/failover acceptance. QA executed 30K virtual-AP tests; analyzed results and contributed controller-aware test-harness integration.
- Owned public API/backward compatibility and releases for customer and internal integrations; modernized controller capabilities from C/C++ to Java using JNI where needed.

Earlier Experience

Wozai Info Tech / Spring House (CHT Subsidiary) / Gamania | Taipei | Jul 2007 – Sep 2012

Android and OpenID development; HamiBook e-book platform and DRM pipeline; MMORPG client and server engineering.

SELECTED PROJECTS

LLM Wiki Cloud — AI Knowledge Workspace: Turns Markdown files and web sources into structured, searchable, source-linked knowledge for grounded AI-assisted work. Shipped multi-tenant BYOK, ingest/compile/publish, tenant isolation, retrieval, and source traceability. [Demo: wiki.rayer.idv.tw](https://wiki.rayer.idv.tw)

MapSwitch — macOS Safari Extension: 2.6 MB privacy-focused extension for opening locations in Google Maps, OpenStreetMap, or Apple Maps; the App Store reports no data collection. [App Store](#)

EDUCATION & LANGUAGES

National Chung Hsing University — Food Science Studies | Chinese — Native | English — Professional Working Proficiency