Irving García Mendoza

Cloud Platform & AI Infrastructure Leader | Infrastructure & Information Security Manager

Remote — Mexico · hola@irving.bio

Professional Summary

Cloud platform and infrastructure leader with 15+ years spanning cloud operations, security engineering, and regulatory consulting — currently **Infrastructure & Information Security Manager at Clara, a pre-IPO LATAM fintech** operating a 70-account, two-region AWS Organization with a ~USD $2.9M annual cloud and SaaS footprint. Builds and runs the platforms the organization depends on: a six-agent **LLM operations fleet on AWS Bedrock** with policy-gated autonomy, an eight-cluster **Kafka (MSK)** estate, self-hosted **source control and inventory platforms** that replaced SaaS dependencies, event-driven **telemetry pipelines**, and the FinOps analytics stack. Delivers infrastructure as code with CI validation, documents decisions as architecture decision records, and holds every change to an auditable standard of approval, test, proof of success, and rollback.

Career Highlights

- Designed and operates six **LLM agents** (AWS Bedrock — Anthropic Claude and OpenAI GPT — plus self-hosted inference) serving infrastructure operations over Slack and an internal Matrix homeserver, including an **autonomous Kafka operations agent** that remediates authenticated alarms through registry-gated actions and patched a vulnerable service unattended in ~6.5 minutes.
- Executed a controlled **Kafka 3.6 to 3.9 upgrade of seven Amazon MSK clusters in ~21 hours with no recorded incidents**; raised ~1,950 topics to replication factor 3 (~400 GiB migrated, no partition unavailability); leads a right-sizing program with ~18% of run-rate verified in billing and ~50% modelled.
- Replaced SaaS and low-code dependencies with self-hosted, tested platforms: a **Steampipe inventory service** over 66 accounts, a **Forgejo source-control and CI platform** on ECS/Graviton, and a unit-tested Cloudflare Workers event gateway that retired a 42-workflow n8n server.
- Led and completed the migration of a full-mesh **VPC Peering estate (80+ connections) to AWS Transit Gateway** across two regions with atomic route cutover and instant rollback — zero service disruption.
- Built the FinOps analytics stack on CUR 2.0 (36 months, ~4.2B rows) that exposed a **~35% cost under-reporting defect**; produced the container-platform business case identifying **~31% run-rate savings**.

Core Competencies

AI Platform Engineering

- LLM agent architecture on AWS Bedrock (Anthropic Claude, OpenAI GPT, cross-region inference profiles) and self-hosted open-weight inference (llama.cpp)
- Agent security: IAM boundaries with cross-account roles, unit-tested guard hooks, toolset restriction, prompt contracts, negative-tested least privilege, signed webhook chains
- Policy-gated autonomous remediation, model portability with context-window and cost tuning, multi-surface delivery (Slack, Matrix, webhooks), MCP server development
- RAG architectures (Bedrock Knowledge Bases, S3 Vectors, Aurora pgvector, OpenSearch)

Platform & Infrastructure Engineering

- AWS multi-account Organizations (70 accounts), SCPs, StackSets, multi-region operations
- ECS on Graviton (capacity providers, Service Connect, EFS, RDS, KMS end to end), EKS, ROSA
- Infrastructure as Code: CloudFormation, Terraform, Terragrunt; CI validation with cfn-lint, ShellCheck, pytest, vitest; architecture decision records
- Self-hosted developer platforms: Forgejo with Actions runners, Matrix/Synapse homeserver
- Serverless on Cloudflare Workers (Workflows, Durable Objects, D1, KV, AI Gateway, Hono)

Data & Streaming Platforms

- Amazon MSK: version upgrades, replication and MinISR standards, partition-density remediation, right-sizing, centralized log forwarding, Managed Prometheus/Grafana

- OpenSearch lifecycle management, Aurora PostgreSQL, DocumentDB, DMS change-data-capture

Observability & Telemetry

- OpenTelemetry collector fleets, Vector pipelines (CloudWatch, Firehose, syslog over TLS), OpenSearch Dashboards and Alerting, CloudWatch OAM alerting standards, custom exporters

Security, FinOps & Governance

- GuardDuty, Prowler (OCSF), Cloudflare Zero Trust, KMS, IAM Identity Center, encryption by default, SSH elimination, WAF
- CUR 2.0 analytics (Athena/Glue/QuickSight), Savings Plans verification, AI-spend attribution, AWS Marketplace and vendor-contract administration

Professional Experience

Clara — Infrastructure & Information Security Manager

Pre-IPO fintech (corporate cards & spend management), LATAM — Mexico, Brazil, Colombia · 2020 – Present

Owns cloud infrastructure, platform engineering, and security posture for a regulated fintech operating a 70-account AWS Organization across three countries and two AWS regions.

AI Platform Engineering

- Designed and operates six **LLM agents** on Graviton EC2 (hardened AL2023, IMDSv2, private subnets, Secrets Manager) serving infrastructure operations over Slack and a self-hosted **Matrix homeserver** — conversational diagnostics, ECS and OpenShift operations, dead-letter queue redrive, incident management, FinOps reporting. Security model: IAM as the boundary (org-wide StackSet roles with cross-account ExternalId), unit-tested pre-execution guard hooks, toolset restriction per surface, and prompt contracts. Diagnostic agents are read-only; remediation is limited to explicitly authorized, gated actions verified by negative testing, with prompt-injection hardening that overrides operator requests.
- Built an **autonomous Kafka operations agent**: CloudWatch alarms pass through a Cloudflare Worker that validates SNS signatures and re-signs with HMAC into the agent, which triages and remediates through registry-gated actions with dry-run, throttling, and post-action verification (human approval retained only for unclean leader election). It patched a vulnerable internal service unattended in ~6.5 minutes, runs a daily replication-invariant audit, and was hardened with a maintenance-window gate after an alarm-storm incident (34 transitions in 59 minutes).
- Migrated the fleet's primary model from Anthropic Claude to **OpenAI GPT on Bedrock** (Converse API) with context compaction tuned to pricing tiers; runs auxiliary tasks on smaller models and one agent on a self-hosted open-weight endpoint — a vendor-portable agent architecture.
- Built an LLM-driven support-intake bot on **Cloudflare Workers** (Workflows, Durable Objects, D1, AI Gateway) converting Slack requests into structured, tracked tickets with a claim-and-escalation lifecycle (174 tickets routed to seven engineers in one quarter); built an LLM-assisted pull-request review webhook with signature verification and idempotent workflows.
- Decommissioned a 42-workflow **n8n** automation server by building a unit-tested TypeScript event gateway (SNS/EventBridge ingress, semantic de-duplication, D1 ledger; 23 test files); replaced third-party automations with auditable Workers.
- Evaluated and deployed an executive **multi-agent AI advisor** proof of concept for company leadership: hosting comparison across ECS Fargate, App Runner, OpenShift, and ECS on EC2 with four architecture decision records; a Bedrock provider adapter; roster-based SSO allow-listing; API image reduced from an estimated 8–10 GB to a measured 1.4 GB; an independent AI review gate before implementation; implementation reached a live UI in two working days.
- Deployed a self-hosted **agent memory platform** on ECS; designed a compliance-documentation **RAG architecture** (Bedrock Knowledge Bases on S3 Vectors, chosen over Aurora pgvector for cost and cold-start profile); governs organization Bedrock access and AI spend through inventory audits, per-principal API-key attribution, and model-tiering analysis.

Platform Engineering & Developer Tooling

- Built and operates the team's **self-hosted source-control and CI platform**: Forgejo with Actions runners (docker-in-docker sidecar) on ECS/Graviton across three availability zones with a managed-scaling capacity provider, RDS PostgreSQL with IAM authentication, EFS and S3 object storage, KMS encryption on every store, and an internal TLS 1.3 network load balancer reachable only through Zero Trust; hosts 25+ infrastructure repositories.
- Deployed an internal **Matrix homeserver** (Synapse, Element, identity server) as a 45-resource CloudFormation stack on ECS/EC2 with RDS PostgreSQL, configuration bootstrapped from S3, host-header-routed internal ALB, and RFC1918-only ingress.
- Administers the organization's infrastructure GitHub organization: CI validation with cfn-lint, ShellCheck, and unit tests; organization webhook to the pull-request review agent; authored 35+ infrastructure repositories (500+ commits) since May 2026.
- Builds hardened **Amazon Linux 2023 golden AMIs** (EC2 Image Builder, Graviton) on a quarterly cadence for all accounts and regions; directed lifecycle programs spanning DocumentDB engine upgrades (7 clusters), OpenSearch engine upgrades, EOL Lambda runtime remediation (78 functions, 1,283-function follow-on scoped), and ElastiCache service updates.

Data & Streaming Platforms

- Executed a controlled **Kafka 3.6 to 3.9 upgrade of seven Amazon MSK clusters** in ~21 hours with no recorded incidents under an AWS forced-upgrade window, after removing 104 abandoned topics (7,600 to 5,912 partition replicas) to unblock configuration changes and avoid broker additions; prepared staging by deleting 718 and recreating 368 topics to a new standard.
- Raised ~1,950 topics across 7 clusters to replication factor 3 (~400 GiB migrated under throttling with no partition unavailability), disabled unclean leader election fleet-wide, and established topic standards with a daily replication-invariant audit; responded to an AWS Health high-risk event on a production payments cluster.
- Leads the **MSK fleet right-sizing program** (~18% of run-rate verified in billing, ~50% modelled); centralized broker and connector logs from 4 accounts and 2 regions through Firehose and Vector with 396-day retention in two days; re-architected Kafka observability onto **Amazon Managed Prometheus and Grafana**.
- Raised OpenSearch lifecycle (ISM) coverage from 18.5% to 98.6% across a 1,703-index security-logging cluster; remediated write-saturation and shard-distribution incidents; replaced static OpenSearch administrator passwords with **IAM Identity Center–derived SigV4 credentials**.
- Operates data-integrity controls over financial pipelines: AWS DMS change-data-capture (13 replication instances, 213 tasks) with latency alerting across an estate of 86 Aurora PostgreSQL clusters, DocumentDB, OpenSearch, MSK, Databricks, MWAA (Airflow), and QuickSight.

Observability & Telemetry Pipelines

- Executed the observability vendor migration contract-to-contract; operates an **OpenTelemetry collector fleet** across EC2 and ECS hosts with configuration distributed via SSM Parameter Store, plus centralized cross-account CloudWatch observability spanning 64 accounts; is developing an organization-wide alerting standard on CloudWatch OAM with staged rollout and numbered decision records.
- Built event-driven **Vector** pipelines: firewall syslog over TLS with dual sinks (OpenSearch and OTLP); a **Prowler findings lake** streaming OCSF security findings through a cross-account, event-triggered Fargate pipeline into OpenSearch (726K findings, zero bulk errors, deterministic IDs and checksum verification); a custom exporter filling inotify and file-descriptor monitoring gaps fleet-wide.
- Investigated and documented telemetry incidents (a ~13,000 lines-per-second log flood, a reverse-proxy outage) with remediation evidence; catalogued collector configuration defects into a fleet remediation program.

Security & Identity Engineering

- Led the organization-wide **AWS GuardDuty** deployment (70 accounts, two regions, delegated administration, StackSet auto-enrollment, KMS-encrypted publishing) with centralized EventBridge alerting and an LLM-annotated daily digest running as an EKS CronJob with Pod Identity.
- Enforced **EBS encryption by default** across 54 accounts (107 account-region pairs); SSH disabled on 174 EC2 instances and baked into 57 ECS launch templates with continuous compliance reporting via SSM State Manager; vulnerable kernel module removed across 13 accounts.
- Built an **automated offboarding relay** on Cloudflare Workers deprovisioning password manager, BI, and Zero Trust access on ticket events (isolate-not-delete for audit retention); led an IAM Identity Center audit (65 accounts, 165 users, 52 permission sets); migrated 240 users from OpenVPN to **Cloudflare Zero Trust** and hardened WARP baselines.

- Ran the email anti-spoofing program (31 domains, 96% compliance across 18 SES sending domains); conducted a WAFv2 effectiveness review (21 Web ACLs, 65 load balancers) and contributed to the 66-account WAFv2 baseline.

Networking

- Led and completed the **VPC Peering to AWS Transit Gateway** migration (80+ connections, multi-region, cross-account) with atomic route cutover and instant-rollback design.
- Redesigned 6 partner site-to-site IPSec VPNs from static routing to **BGP dynamic routing** (IKEv2-only, automatic failover) validated on a staging canary; led the vendor-facing provider migration across Brazil, Mexico, and Colombia environments with Spanish-language runbooks.
- Built a production Cloudflare Workers reverse proxy across three country domains replacing a referral-link SaaS for the Growth team, with two independent review rounds and dry-run verification.

FinOps & Vendor Governance

- Stewards the ~USD $2.9M annual cloud and SaaS footprint: built the CUR 2.0 analytics stack (36-month backfill, ~4.2B rows) and a QuickSight dashboard with row-level security; exposed a ~35% under-reporting defect in Athena cost data; verified Savings Plans utilization at 97%; removed dormant resources (94 Lambda functions across 12 accounts, orphaned connectors, unused Elastic IPs).
- Replaced the primary cross-account inventory SaaS with a **self-hosted Steampipe platform** (66 accounts, cached inventory tables refreshed on a timer, VPN-only TLS access, consumed by agents through MCP), retaining the commercial service as fallback.
- Administers the **AWS Marketplace subscription portfolio** (11 products — Databricks, Auth0/Okta, Red Hat OpenShift, LaunchDarkly, Splunk, Fortinet, Turbot, and observability platforms); produced the ROSA-versus-EKS business case identifying ~31% run-rate savings.

Grow Mobility — Chief Security Officer & Infrastructure Manager

- Established and led the information-security function: web and mobile (iOS/Android) application security audits, API hardening, and infrastructure, database, and cloud security assessments.
- Built and administered the AWS environment with cloud-wide monitoring (New Relic).
- Designed and operated a data-lake platform (AWS Redshift, Glue, Athena) processing ~35 TB across 29 SQL databases and 270 DynamoDB tables for near-real-time analytics.

Advanced Solutions 2H Mexico — Solutions & Services Manager

- Designed and implemented client information-security solutions; managed medium- and long-term network re-engineering programs establishing foundations for control, administration, and performance-monitoring technologies.
- Evaluated and integrated security technologies: UTM appliances (ASIC/FPGA), on-the-fly and pre-boot encryption, token/software authentication, and SIEM/event-correlation platforms (ArcSight, NetIQ/Attachmate, Enterasys).
- Performed controlled real-attack product evaluations and contracted penetration testing.

Independent Information Security & Regulatory Consultant

- Delivered infrastructure and information-security consulting to financial-sector and government clients, including **Banjército** (banking), **ANA Seguros** (insurance), Alsis Funds, Solutrust (fiduciary services), Secretaría de Relaciones Exteriores, Secretaría del Trabajo y Previsión Social, SEDESOL/DICONSA, SKY/Novavisión, Universidad Latina (UNILA), Grupo THEOS, ITERA, McAfee, and Fortinet.
- Advised on **fintech regulatory frameworks** and **Latin-American data-protection law** (Mexico, Brazil, Chile, Colombia, Peru).
- Specialized in database query-filtering and record-transfer controls, network-anomaly detection, data-loss prevention (DLP), and IPv6 hardening for DNS, DHCP, and mail services.

Early Career — Systems Administration

- **Universidad y Escuela Internacional, S.C.** — Designed Active Directory with ACL-based segregation across administrative, academic, and student populations; administered ISA Server clusters with content filtering; integrated SUSE Linux and Informix/ADABAS-D systems.

- **EJE Technologies, S.A. de C.V.** — Remote administration of email, web hosting, and monitoring (Nagios/Cacti) on FreeBSD across SPARC and Intel x86 architectures.

Technical Skills

- **AI & Automation:** AWS Bedrock (Anthropic Claude, OpenAI GPT, cross-region inference profiles, Converse API), self-hosted inference (llama.cpp), Cloudflare AI Gateway, LLM agent architecture and security (guard hooks, gated remediation, signed webhook chains, prompt contracts), RAG (Bedrock Knowledge Bases, S3 Vectors, pgvector), MCP server development, Cloudflare Workers (Workflows, Durable Objects, D1, KV), n8n
- **Cloud & Platforms:** AWS (Organizations, IAM/Identity Center, EC2/Graviton, VPC/Transit Gateway, ECS/EKS, Lambda, RDS/Aurora PostgreSQL, DocumentDB, S3, EFS, MSK/Kafka, OpenSearch, DMS, MWAA/Airflow, CloudFront, API Gateway, Kinesis Data Firehose, Athena/Glue, QuickSight, Bedrock, Image Builder, SSM, EventBridge, Step Functions, Transfer Family), Red Hat OpenShift (ROSA), Databricks, Cloudflare (DNS at scale, Zero Trust, Workers)
- **Developer Platforms & IaC:** Forgejo with Actions CI, Matrix/Synapse, GitHub (organization administration, Actions), CloudFormation (StackSets), Terraform, Terragrunt, cfn-lint, ShellCheck, Python (pytest, boto3), Bash, TypeScript/Node.js (Hono, vitest), PowerShell, Ruby, Git, Docker, systemd
- **Observability:** OpenTelemetry Collector, Vector, Tsuga, Dynatrace, Splunk, Amazon Managed Prometheus, Managed Grafana, CloudWatch (cross-account OAM, Container Insights), OpenSearch Dashboards and Alerting, pgBadger
- **Security & Compliance:** GuardDuty, CloudTrail, Security Lake, Inspector, AWS Config, KMS, Secrets Manager, WAF (AWS / Fortinet), Cloudflare Zero Trust, Okta/Auth0, Prowler (OCSF), Sprinto, Steampipe/Turbot, SPF/DKIM/DMARC, network forensics (Wireshark/tshark)
- **Resilience & Data:** AWS Backup, snapshot/lifecycle policies, BCP/DRP, DMS/CDC data integrity, Kafka replication and MinISR standards, Kafka version upgrades, encryption at rest & in transit
- **Frameworks (working knowledge):** SOX (IT component) / ITGC / ITAC, SOC 2, ISO 27001, PCI DSS, NIST CSF, LGPD / LATAM data-protection law

Certifications & Professional Development

- *In progress:* CISA, CRISC, ISO 27001 Lead Implementer/Auditor
- Contributor to **b:Secure** information-security magazine
- Speaker at industry events including **b:Secure Conference** and the **Congreso de Seguridad y Software Libre**

Languages

- **Spanish** — Native
- **English** — Professional working proficiency (B2+)

Education

- Business Administration

Additional

- Open source: maintains a fork of an MCP diagram-rendering server with protocol-handshake and container fixes, delivered through pull requests with Node 20/22 CI.
- Independent work in financial markets: development of technical indicators (TradingView) and an algorithmic trading bot for automated order execution (Python, CCXT, Deribit).