

UNVARNISHED REVIEWS

Kong vs. Apigee vs. TrueFoundry: The Per-Service Multiplier

Market Position: Big Picture

Kong, Apigee, and TrueFoundry sit at different points of the same growing category: gateways that manage, secure, and route traffic, increasingly including traffic to LLM providers, not just internal APIs. Kong and Apigee are the two dominant general-purpose API management platforms, both now extending into AI gateway functionality; TrueFoundry is a newer, AI-native entrant built from the start around routing to 250+ LLM providers, with API management as a secondary capability rather than the core product.

The finding that should change how any of the three gets budgeted for AI-routing use cases specifically: Kong and Apigee both price in a way that multiplies with the number of things you route, not just the volume of traffic. Kong charges per gateway service, and each LLM model you route counts as its own billable service. Apigee charges per environment, per region, a fee that compounds independently of call volume. TrueFoundry, running inside your own cloud infrastructure, charges neither, its cost is the infrastructure itself, flat regardless of how many models or endpoints you add.

Platform Ratings at a Glance

Platform	G2 Rating	Reviews (G2)	Note
TrueFoundry	4.6 / 5	~55	Highest-rated of the three; mid-market-skewed (48.1%)
Kong (Kconnect)	4.4 / 5	~322	Largest review base by a wide margin
Apigee	~4.4 / 5	19-89, varies by listing	G2 lists Apigee under multiple separate product pages (Apigee Edge, Apigee API Management, Google Apigee API Management Platform); ratings converge closely but review counts differ meaningfully by listing

TrueFoundry rates highest of the three despite the smallest review base. Apigee's split across multiple G2 listings is worth knowing before treating any single cited rating as the complete picture, confirm which specific listing a rating or review count refers to.

Kong: Fast and Flexible, With a Real Per-Service Multiplier for AI Routing

Kong Konnect Plus prices at roughly \$105/month per gateway service, with 1 million requests included and \$200 per additional million up to a 10 million monthly cap; Kong Gateway OSS remains permanently free and self-hosted for teams willing to run it themselves. Enterprise is a custom-contract tier with no public pricing, unlimited hybrid and dedicated gateways, unlimited AI Gateway models, and dedicated support; independent estimates place small Enterprise deployments at \$30,000-\$50,000 annually, scaling into six figures for larger implementations.

The mechanism specific to AI-routing budgets: each LLM model routed through Kong's AI Gateway counts as its own gateway service, billed at the same roughly \$105/month rate as a standard API endpoint. A team with 20 existing API services adding 5 LLM model integrations doesn't add a flat AI-routing fee, it adds 5 more services at full price, roughly \$525/month in new service fees before any request volume from those models. Kong's technical performance is independently benchmarked strongly, up to 137,850 RPS on a basic proxy configuration with 3.82ms added latency at p95, real speed, but the per-service structure means AI-routing costs scale with model count, not usage.

Apigee: Deep Google Cloud Integration, With Environment Fees That Compound by Region

Apigee prices on two independent dimensions: per-call fees starting at \$20 per million calls (dropping to \$13/million above 500M), and a separate per-environment, per-region fee running \$365/month at the Base tier up to \$3,431/month at the Comprehensive tier. A Standard subscription runs \$500/month with 15 million calls included; Enterprise is quote-only, with independent estimates placing typical annual spend at \$25,000-\$100,000+, and some sources citing \$250,000-\$500,000+ for larger multi-region enterprise deployments.

The mechanism that compounds specifically for multi-region or multi-environment deployments: The environment fee is charged per environment, per region, independent of call volume. A team running the same workload across two regions for redundancy or latency pays the environment fee twice, on top of identical call charges, a structural cost that has nothing to do with how much traffic is actually flowing. Users consistently praise Apigee's Google Cloud integration and analytics depth while citing cost and a steep learning curve as the most common complaints.

TrueFoundry: Runs Inside Your Own Cloud, Sidestepping Both Fee Structures

TrueFoundry prices on a seat-based model with a free tier (Developer, 3 users) and a Pro tier at \$499/month that includes RBAC, budget controls, rate limiting, and unlimited logging and retention at every tier, features that competitors commonly gate behind higher tiers. Enterprise is custom, with VPC and on-premises deployment available. The AI Gateway unifies access to 250+ LLM providers through a single OpenAI-compatible API, with MCP Gateway support for agentic workflows, a capability Kong and Apigee are still extending into rather than having built around from the start.

The structural difference from both Kong and Apigee: TrueFoundry's self-hosted Enterprise deployment runs entirely inside the customer's own cloud account (AWS, GCP, or Azure), eliminating both the per-service and per-environment fee structures entirely. Independent estimates place the marginal infrastructure cost at \$600-\$1,000/month, the actual compute and storage consumed, not a separate platform fee layered on top. This means TrueFoundry's cost stays flat as model count or endpoint count grows, the opposite of Kong's and Apigee's scaling patterns, though it requires a team comfortable operating infrastructure inside their own cloud account rather than consuming a fully managed SaaS product.

Pricing (July 2026)

Platform	Billing Unit	Entry Tier	What Scales the Bill
Kong	Per gateway service + per-request overage	Free (OSS, self-hosted)	Number of services/models routed
Apigee	Per environment/region + per-call	60-day evaluation sandbox	Number of environments/regions
TrueFoundry	Seat-based (SaaS) or infrastructure-only (self-hosted)	Free (Developer, 3 users)	Underlying compute/storage only, if self-hosted

TCO Comparison: 20 API Services + 5 LLM Models Routed, 10M Requests/Month

This scenario models a representative mid-size deployment adding AI-routing to an existing API footprint, the exact scenario where each platform's billing mechanism produces meaningfully different real costs.

Platform	Structural Fee	Volume Fee (10M req/mo)	Monthly Total
Kong (25 total services)	\$2,625 (service fees)	Included up to 10M cap	~\$2,625
Apigee (2 regions, Base tier)	\$730 (2 × \$365 environment)	\$200 (10M calls)	~\$930
TrueFoundry (self-hosted)	\$0 platform fee	\$600-\$1,000 (infrastructure only)	~\$600-\$1,000

Kong's service-count structure produces the highest bill in this specific scenario, since every one of the 25 routed services carries its own fee regardless of how much traffic each one handles. TrueFoundry's flat infrastructure cost is lowest, but reflects a distinctly different operating model, self-managing infrastructure inside your own cloud, not a like-for-like SaaS comparison. Apigee sits between the two, with its region-count multiplier mattering most for teams running true multi-region deployments rather than single-region setups.

Costs the Pricing Table Misses

- **Kong's per-service fee applies identically to a low-traffic LLM integration and a high-traffic one.** A rarely used specialized model costs the same \$105/month in service fees as your highest-volume production endpoint; audit for routed models that aren't earning their fee.
- **Apigee's environment fee is charged whether or not an environment is actively serving production traffic.** Staging and development environments carry the same per-environment fee as production; confirm which environments are actually necessary before multiplying regions.
- **TrueFoundry's self-hosted model shifts real cost from subscription to internal DevOps labor.** Running infrastructure inside your own cloud account means your team owns provisioning, scaling, and security patching, a real, ongoing cost that doesn't appear on the infrastructure bill itself.
- **Kong Enterprise and Apigee Enterprise contract ranges are wide because they're genuinely negotiated, not published.** The \$30,000-\$50,000 and \$25,000-\$100,000+ figures in this report are independent estimates; get a specific, itemized quote before budgeting either platform at enterprise scale.
- **None of the three platforms' AI Gateway features are fully mature at the same point in their product lifecycle.** Kong and Apigee are extending established API platforms into AI routing; TrueFoundry built AI routing first. Confirm current feature parity against your specific LLM provider and MCP requirements rather than assuming all three cover the same ground equally.

The Decision Framework

Choose Kong if: Sub-millisecond API routing performance and a mature plugin ecosystem are the primary requirement, and your AI-routing footprint is small enough that the per-service fee structure doesn't compound into a real cost problem.

Choose Apigee if: Deep Google Cloud integration, API monetization, and enterprise-grade analytics are real requirements, and your deployment is single-region or the multi-region environment fee is an acceptable tradeoff for the platform's full-lifecycle governance capability.

Choose TrueFoundry if: AI and LLM routing, not general API management, is the primary use case, your team is comfortable operating infrastructure inside your own cloud account, and you want cost that scales with actual infrastructure use rather than service or environment count.

Everyone: Before comparing headline rates, count how many services, models, or environments your actual deployment will require, then apply each platform's specific billing mechanism to that count. The platform with the lowest per-unit rate is not reliably the platform with the lowest real bill once your specific service, model, and region count is factored in.

The Bottom Line

Kong, Apigee, and TrueFoundry are all credible, well-reviewed platforms solving a related but distinctly different problem, Kong for high-performance API routing with AI Gateway extensions, Apigee for full-lifecycle enterprise API governance with AI Gateway extensions, and TrueFoundry for AI and LLM routing built from the ground up. The real finding this comparison surfaces isn't which platform is cheapest per request, it's that Kong's and Apigee's pricing structures both multiply with something other than usage, service count for Kong, environment and region count for Apigee, while TrueFoundry's self-hosted model ties cost to actual infrastructure consumption instead. Count your real service, model, and region footprint before comparing any two of these three.