

MASTERING KONG API GATEWAY

FROM CORE ARCHITECTURE TO
AI-POWERED LLM ORCHESTRATION



CORE ARCHITECTURE
& INTERNALS



SECURE & SCALE
APIS



KUBERNETES &
INGRESS CONTROLLER



AI GATEWAY &
LLM ORCHESTRATION



COST OPTIMIZATION &
OBSERVABILITY



DOCKER



KUBERNETES



DECK



TERRAFORM



HELM



APPLICATIONS
APIs, LLMs & MORE

EVAN RILEY

Mastering Kong API Gateway

From Core Architecture to AI-Powered LLM
Orchestration

Steve Publications

This book is available at <https://leanpub.com/masteringkongapigateway>

This version was published on 2026-07-20



This is a [Leanpub](#) book. Leanpub empowers authors and publishers with the Lean Publishing process. [Lean Publishing](#) is the act of publishing an in-progress ebook using lightweight tools and many iterations to get reader feedback, pivot until you have the right book and build traction once you do.

© 2026 Steve Publications

Contents

From Core Architecture to AI-Powered LLM Orchestration	1
Preface	2
Who Should Read This Book	2
How This Book Is Organized	3
What Is Not in This Book	3
Conventions Used in This Book	4
OSS versus Enterprise Features	4
Software Versions	4
Feedback and Corrections	5
Introduction: Why Kong Matters	6
Chapter 1: Architecture and Core Concepts	8
OpenResty and the Nginx + Lua Foundation	8
The Kong Request Lifecycle	9
Services, Routes, Upstreams, and Consumers	12
Configuration Models: Database-Backed, DB-less, Hybrid	17
The Plugin System Architecture	21
Chapter 2: Installation and Deployment	24
Prerequisites and Environment Setup	24
Docker and Docker Compose Deployments	24
Database-Backed Mode with PostgreSQL	24
DB-less Mode with YAML Configuration	25
Hybrid Mode: Control Plane and Data Plane	25
Kong Konnect Cloud Deployment	26
Choosing Your Deployment Model	27
Kong Manager: The Web UI for Kong Administration	27
Chapter 3: Routing, Services, and Traffic Management	29

CONTENTS

Defining Services and Routes	29
Route Matching Strategies	29
Upstreams and Load Balancing Algorithms	30
Health Checks: Active and Passive	31
Service Discovery Integration	31
API Versioning Strategies	32
Chapter 4: Authentication and Authorization	34
API Keys and Basic Authentication	34
HMAC Authentication	34
JWT Validation and Issuance	34
OAuth 2.0 and OIDC Flows	35
OpenID Connect with Keycloak	35
LDAP and Active Directory Integration	35
Access Control Lists (ACL)	35
Custom Authentication with Lua Plugins	36
Chapter 5: Rate Limiting, Throttling, and Traffic Shaping	37
Rate Limiting Algorithms: Local, Redis, Cluster, Sentinel	37
Per-Consumer and Per-Route Limits	37
Burst Handling and Token Bucket	38
Load Shedding Strategies	38
Rate Limit Response Headers	39
Chapter 6: Plugins Ecosystem	40
Plugin Phases and the Request Lifecycle	40
Built-in Plugins Overview	40
Configuring and Chaining Plugins	41
Plugin Priority and Execution Order	41
Writing Custom Lua Plugins	41
Third-Party Plugins from Kong Hub	41
Chapter 7: Security, TLS, and API Governance	43
Mutual TLS (mTLS) Configuration	43
Certificate Management and Auto-Renewal	43
IP Restriction and Geo-Blocking	44
Bot Detection and Abuse Prevention	44
Request/Response Body Filtering	44
CORS, CSRF, and Cross-Origin Security	45
Vulnerability Scanning Integration	45

Chapter 8: Observability, Logging, Monitoring, and Analytics 46

 Access Logs and Structured Logging 46

 Prometheus Metrics and Grafana Dashboards 47

 Distributed Tracing with OpenTelemetry 47

 Log Streaming to ELK, Splunk, Datadog 48

 Kong Analytics and Konnect Insights 48

 Custom Metrics with Plugins 49

Chapter 9: Kubernetes Integration 50

 Kong Ingress Controller Architecture 50

 Helm Chart Installation and Configuration 50

 Ingress Resources and Annotations 50

 Kong Gateway API Implementation 51

 Plugin Configuration via Custom Resources 51

 Multi-Cluster and Edge Deployment Patterns 52

Chapter 10: Automation, CI/CD, and GitOps 53

 The Admin API Reference 53

 deck Configuration Management 53

 Declarative Config and Version Control 54

 Terraform Provider for Kong 55

 CI/CD Pipeline Patterns 55

Chapter 11: Performance Tuning, Scalability, and High Availability . . . 57

 OpenResty and Nginx Tuning Parameters 57

 Connection Pooling and Keep-Alive Optimization 58

 Horizontal Scaling Strategies 58

 High Availability Architecture 59

Chapter 12: Kong AI Gateway and LLM Orchestration 61

 The AI Gateway Problem 61

 Getting Started with AI Gateway 61

 Intelligent Model Routing Strategies (Enterprise) 61

 Multi-Provider Orchestration 62

 Fallback and Retry Mechanisms 62

 Prompt Transformation and Template Management 63

 Response Caching for AI Workloads (Enterprise) 63

 Guardrails, Safety, and Content Moderation 63

 Token Tracking and Cost Optimization 64

 AI Observability and Governance 65

Chapter 13: AI Integrations and Framework Patterns 67

 OpenAI and Azure OpenAI Integration 67

 Anthropic Claude Integration 67

 Google Gemini Integration 68

 AWS Bedrock Model Access 68

 Self-Hosted Models: Ollama, vLLM, Hugging Face 69

 LangChain Orchestration with Kong 69

 LlamaIndex and RAG Patterns 70

 Streaming Responses and SSE 70

Chapter 14: Production Patterns, Migrations, and Troubleshooting . . . 71

 Application Integration Patterns 71

 gRPC and GraphQL Gateway Patterns 72

 Kafka Event Streaming with Kong 72

 Common Failure Modes and Troubleshooting 73

 Upgrade Strategies and Version Migration 73

 Real-World Migration Case Study: From Nginx Ingress to Kong 74

 Configuration Auditing and Compliance 74

 Production Checklist 75

Conclusion: The Future of API Infrastructure 76

 Key Takeaways 76

 Emerging Trends in API Architecture 76

 Where Kong Is Heading 76

 Final Recommendations 76

References 77

 Core Kong Documentation 77

 Kong Manager and RBAC 77

 AI Gateway and AI Plugins 77

 Installation and Deployment 77

 Kubernetes Integration 77

 Automation and Configuration Management 77

 Observability and Monitoring 78

 Underlying Technologies 78

 Community and Support 78

Index 79

 A 79

 B 79

C	79
D	79
E	79
F	79
G	80
H	80
I	80
K	80
L	80
M	80
N	80
O	81
P	81
R	81
S	81
T	81
U	81
V	81
W	82

From Core Architecture to AI-Powered LLM Orchestration

This book is your complete reference for mastering Kong API Gateway from the ground up. Whether you are deploying your first route, architecting a multi-region platform, or orchestrating traffic across dozens of LLM providers, this guide provides production-ready patterns, real configurations, and deep architectural understanding. You will learn how Kong works internally on OpenResty, how to deploy it in every supported mode, how to secure APIs at scale, how to run Kong as a Kubernetes ingress controller, and how to harness the AI Gateway for intelligent LLM routing, semantic caching, guardrails, and cost optimization. Every chapter delivers working examples in Docker, Kubernetes YAML, Helm charts, decK declarations, Terraform, and application code across Spring Boot, Node.js, Python, Go, .NET, gRPC, GraphQL, Kafka, LangChain, and LlamaIndex.

Preface

I wrote this book because I was tired of seeing teams reinvent the wheel.

Over the past decade, I have watched dozens of organizations build API infrastructure from scratch. They started with Nginx configuration files scattered across repositories, grew into custom middleware frameworks nobody fully understood, and eventually realized they had spent two years building something Kong already shipped. The pattern repeated itself regardless of industry, team size, or technical sophistication.

The problem was not that these teams were unskilled. It was that the landscape around API gateways had become incredibly complex, and the information about how to use tools like Kong effectively was fragmented across documentation sites, blog posts, GitHub issues, and community forums. There was no single resource that took you from “what is an API gateway” to “how do I run a multi-region, AI-powered Kong platform serving millions of requests per day.”

That is what this book attempts to be.

Who Should Read This Book

This book is designed for three audiences:

Platform and Infrastructure Engineers who are responsible for deploying, operating, and scaling Kong in production. You need to understand the internals, tune performance, design high-availability architectures, and troubleshoot incidents at 2 AM. This book gives you the depth you need.

API and Backend Developers who build services behind Kong and need to understand how routing, authentication, rate limiting, and plugins affect their code. You do not need to manage the Kong cluster itself, but you need to know how to design APIs that work well with gateway patterns.

AI and ML Engineers who are integrating LLMs into production applications and need a governance layer for AI traffic. The final chapters of this book are dedicated entirely to Kong AI Gateway, covering model routing, cost

optimization, guardrails, and integration with frameworks like LangChain and LlamaIndex.

The only prerequisite is familiarity with APIs, HTTP, and basic Linux concepts. No prior Kong experience is assumed. If you have never configured an API gateway before, the early chapters will bring you up to speed. If you are already running Kong in production, you will find advanced patterns and troubleshooting guidance that go beyond the official documentation.

How This Book Is Organized

The book is structured as a journey from fundamentals to mastery. You can read it cover to cover, or use it as a reference for specific topics.

Chapters 1 through 3 establish the foundation. You will learn how Kong works internally, its core architectural concepts, and how to deploy it in every supported mode. Chapter 4 covers authentication and authorization, while Chapter 5 tackles rate limiting and traffic shaping. Chapter 6 surveys the plugin ecosystem and shows you how to write custom plugins in Lua.

Chapters 7 through 10 focus on production operations. You will learn how to harden your APIs with security best practices, set up comprehensive observability, integrate Kong with Kubernetes, and automate configuration management with deckK, Terraform, and CI/CD pipelines. Chapter 11 covers performance tuning, scalability, and high-availability architectures.

The final chapters are dedicated to AI Gateway capabilities. Chapter 12 introduces Kong AI Gateway and explains how it solves the fundamental problems of LLM integration: vendor lock-in, cost unpredictability, reliability, and governance. Chapter 13 provides concrete integration patterns with specific providers and frameworks. Chapter 14 brings everything together with production patterns, migration strategies, troubleshooting guides, and a comprehensive checklist for running Kong at scale.

What Is Not in This Book

This book focuses exclusively on Kong API Gateway and its ecosystem. It does not attempt to compare Kong against every other API gateway on the market, though I occasionally reference alternatives where it helps clarify

design decisions. The emphasis is practical: how to use Kong effectively, not why you should choose it over competitors.

Similarly, this book does not cover AI model training, prompt engineering best practices, or the theoretical foundations of large language models. It assumes you already understand what LLMs are and focuses on the infrastructure layer that makes them production-ready.

Conventions Used in This Book

Code examples use a monospace font. Configuration files, command-line commands, and source code are presented in code blocks with syntax highlighting where appropriate.

Commands that should be run in a terminal are prefixed with `$` or `#` (for commands requiring root privileges). Do not type the prompt character itself.

Configuration examples use realistic but fictional hostnames, API keys, and credentials. Never copy these values into production environments. Always generate your own secure secrets using a proper key management solution.

Throughout the book, you will see two types of annotations:

- **Production Note:** Practical advice based on real-world deployment experience. These highlight pitfalls to avoid and patterns that work in large-scale environments.
- **Security Note:** Warnings about security implications of specific configurations or practices.

OSS versus Enterprise Features

Kong has both an open-source edition and a commercial Enterprise edition. Most core functionality is available in the open-source version, but some advanced features require a license. Throughout this book, Enterprise-only features are clearly labeled with “(Enterprise)” so you can distinguish between what is available in each edition.

If you are evaluating Kong for your organization, I recommend starting with the open-source edition and upgrading only when you need specific Enterprise capabilities. The core architecture and operational patterns are identical across both editions.

Software Versions

This book was written against Kong Gateway 3.9 and related ecosystem tools at their latest stable releases as of early 2026. Kong follows semantic versioning, and minor versions are generally backward-compatible. However, the API gateway space moves quickly, and new features are added regularly. Always consult the official documentation for the most current information, especially when planning upgrades or adopting newly released features.

The configurations and patterns in this book are designed to be forward-compatible where possible, but some specific syntax or plugin names may change in future Kong releases.

Feedback and Corrections

No book of this scope is perfect on first publication. If you find errors, outdated information, or areas that could be clarified, please report them through the official channels for this publication. I maintain a list of known corrections and updates online, and significant issues will be addressed in future editions.

I am particularly interested in hearing about your own Kong deployments: what patterns have worked well, what surprises you encountered, and what topics you wish had been covered. Real-world experience from the community is invaluable, and insights from readers often shape future revisions.

Thank you for choosing this book as your guide to Kong API Gateway. Let us get started.

Introduction: Why Kong Matters

In 2016, when Kong was first released as an open-source project, the microservices revolution was well underway but API infrastructure was still largely improvised. Teams were patching together Nginx configs, writing custom middleware, and struggling with authentication, rate limiting, and observability scattered across dozens of services. Kong arrived with a bold promise: a lightweight, high-performance, plugin-driven API gateway built on OpenResty that could sit in front of any service and handle all the cross-cutting concerns in one place.

Ten years later, that promise has only grown more relevant. Modern applications do not just expose HTTP APIs. They integrate with LLM providers, coordinate agents through protocol layers like MCP, stream events over Kafka, and communicate via gRPC and GraphQL alongside traditional REST. The infrastructure layer must keep pace, providing routing, security, observability, and governance for all of these traffic types from a single control plane.

Kong has evolved into exactly that platform. What began as an API gateway is now what the Kong team calls an “API and AI Gateway,” supporting universal LLM routing across OpenAI, Anthropic, AWS Bedrock, Google Vertex, Azure OpenAI, and self-hosted models, while maintaining its core strengths in HTTP, TCP, and gRPC traffic management. It runs everywhere: on bare metal, in Docker containers, across Kubernetes clusters via the Kong Ingress Controller, and through the cloud-managed Konnect control plane.

This book is organized to take you from zero to production mastery. The early chapters establish the foundation: how Kong works internally, its core entities, and every deployment mode. The middle chapters dive into features: routing strategies, authentication, rate limiting, plugins, security, and observability. Then we cover Kubernetes integration, automation, performance tuning, and operational excellence. The final chapters are dedicated entirely to AI Gateway capabilities, showing you how to route LLM traffic intelligently, optimize costs, enforce guardrails, and integrate with frameworks like LangChain and LlamaIndex.

If you are an API engineer, platform engineer, DevOps professional, or architect responsible for Kong in your organization, this book will become your

reference. If you are just starting out and want to understand what an API gateway is and why Kong is one of the most popular choices, this book will get you there as well. The only prerequisite is familiarity with APIs, HTTP, and basic Linux concepts. No prior Kong experience is assumed.

Let us begin by understanding what Kong actually is under the hood, and why its architecture makes it such a powerful platform.

Chapter 1: Architecture and Core Concepts

Before you can use Kong effectively, you must understand how it works. This chapter explains the architecture that makes Kong fast and flexible, the lifecycle of a request as it passes through the gateway, and the core entities that form the building blocks of every Kong deployment.

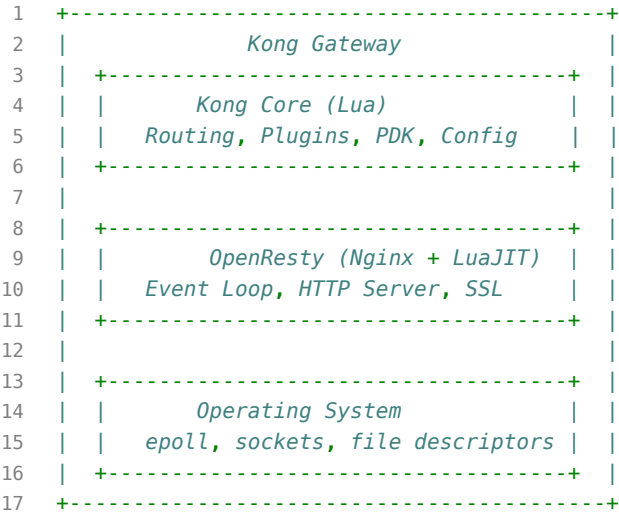
OpenResty and the Nginx + Lua Foundation

Kong is not a standalone server written from scratch. It is built on top of OpenResty, which itself is a bundle of Nginx with LuaJIT and a rich ecosystem of Lua libraries embedded directly into the Nginx event loop. This architectural choice is the foundation of everything Kong does well: high performance, low latency, and deep programmability without sacrificing throughput.

Nginx is renowned for its event-driven, non-blocking architecture. A small number of worker processes handle thousands of concurrent connections by multiplexing I/O events rather than spawning a thread per connection. OpenResty extends this model by embedding a Lua virtual machine in each worker process, allowing Lua code to run inside the request lifecycle with minimal overhead. The LuaJIT compiler translates Lua bytecode into native machine code at runtime, making it exceptionally fast for the kinds of operations Kong performs: routing decisions, header manipulation, authentication checks, and plugin logic.

The key insight is that Kong plugins are Lua scripts that hook into specific phases of the Nginx request processing pipeline. They run in-process, share the same event loop as the proxy itself, and have access to a stable Plugin Development Kit (PDK) that abstracts away the underlying Nginx complexity. This means Kong can be extended with custom logic without requiring external service calls, without adding network latency, and without sacrificing the performance characteristics of Nginx.

Here is a simplified view of the stack:



This architecture has several important implications:

- Plugins execute with microsecond-level latency because they run in the same process as Nginx.
- A single Kong node can handle tens of thousands of requests per second on modest hardware.
- Custom logic is written in Lua, a lightweight scripting language designed for embedding.
- Configuration changes are loaded into shared memory, allowing workers to pick up updates without restarts.
- The event loop model means CPU usage scales linearly with traffic, not with concurrent connections.

Understanding this stack matters when you tune Kong for performance, write custom plugins, or troubleshoot issues. Many operational decisions trace back to how OpenResty and Nginx handle events, workers, and memory.

The Kong Request Lifecycle

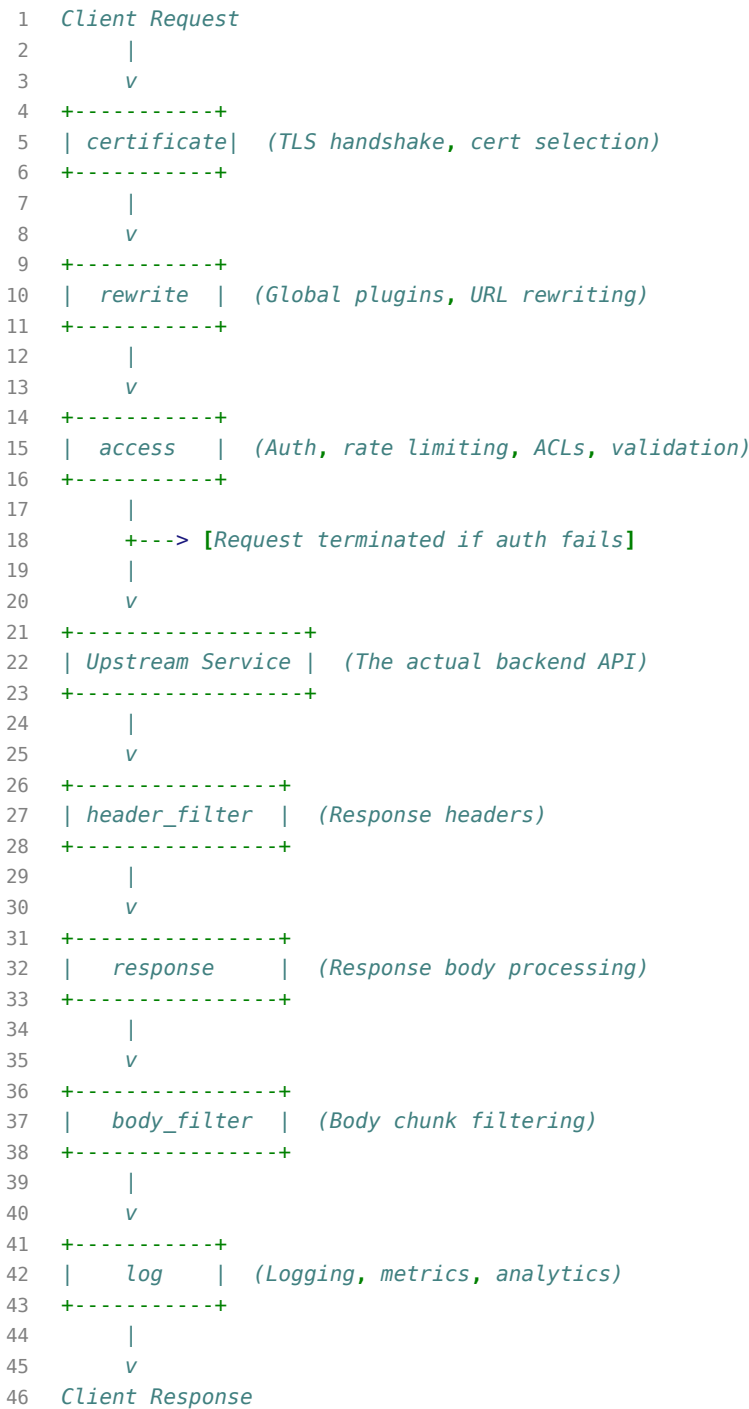
When a client sends a request to Kong, it passes through a well-defined sequence of phases. Each phase has a specific purpose, and plugins can hook into these phases to modify the request or response, enforce policies, collect

metrics, or terminate the request early. This lifecycle is the backbone of Kong's plugin system and understanding it is essential for using Kong effectively.

The complete request lifecycle consists of the following phases:

1. **init_worker**: Runs once when each Nginx worker process starts. Plugins use this phase to initialize connections, load caches, or set up timers. This phase runs outside any specific request context.
2. **certificate**: Executes during TLS handshake when a client connects over HTTPS. Plugins here can select server certificates, validate client certificates for mTLS, or inspect SNI values. The request context is cleared after this phase, so data cannot be passed forward.
3. **rewrite**: Runs early in the request processing before Kong determines which consumer and service are involved. Global-scoped plugins execute here. This is where URL rewriting, early request modifications, and global policies happen.
4. **access**: The most important phase for authentication and authorization. Kong has already matched the request to a route and service by this point, and all applicable plugins execute in priority order. Authentication plugins validate credentials, rate limiting plugins check quotas, IP restriction plugins verify allowed addresses, and so on. Any plugin can terminate the request with an error response during this phase.
5. **header_filter**: Runs after Kong receives the response headers from the upstream service but before sending them to the client. Plugins can add, modify, or remove response headers here. This is where response transformation plugins operate.
6. **response**: Runs while Kong is receiving the response body from the upstream. Plugins can inspect or transform the response body in this phase. For streaming responses, this phase may execute multiple times as chunks arrive.
7. **body_filter**: Similar to response but focused on filtering and transforming body content. This phase runs for each chunk of the response body.
8. **log**: Runs after Kong has sent the complete response to the client. Plugins use this phase for logging, metrics collection, and analytics. Since the response is already sent, plugins cannot modify anything here; they can only observe and record.

Here is a visual representation of how a request flows through Kong:



The access phase is where most of Kong's security and traffic control plugins operate. When multiple plugins execute in this phase, they run in a

specific order determined by their priority values. Plugins with lower priority numbers run first. This ordering matters when plugins depend on each other: authentication must complete before authorization, rate limiting should happen after authentication so limits can be applied per consumer, and so on.

Services, Routes, Upstreams, and Consumers

Kong organizes its configuration around four core entities: Services, Routes, Upstreams, and Consumers. Understanding how these entities relate to each other is fundamental to using Kong correctly.

Services

A Service represents an upstream backend API or microservice that Kong proxies traffic to. It defines the connectivity details: the protocol (HTTP, HTTPS, TCP, TLS, GRPC, GRPCS), the host or hostname, the port, and optionally a path prefix. A Service is essentially a logical abstraction of your backend infrastructure.

When you create a Service in Kong, you are telling Kong where to send requests after they have been matched by a Route and processed by plugins. The Service configuration includes timeout settings (`connect_timeout`, `read_timeout`, `write_timeout`), retry behavior (`retries` count), and connection pooling options (`keepalive`).

Here is how you define a Service using the Admin API:

```
1 curl -i -X POST http://localhost:8001/services \  
2   --data "name=user-service" \  
3   --data "url=http://user-api.internal:8080"
```

The response includes a generated ID and confirms the Service is created:

```
1 {  
2   "id": "a1b2c3d4-e5f6-7890-abcd-ef1234567890",  
3   "name": "user-service",  
4   "url": "http://user-api.internal:8080",  
5   "protocol": "http",  
6   "host": "user-api.internal",  
7   "port": 8080,  
8   "path": "/",  
9   "created_at": 1700000000  
10 }
```

You can also specify these fields individually instead of using the url shorthand:

```
1 curl -i -X POST http://localhost:8001/services \  
2   --data "name=payment-service" \  
3   --data "protocol=https" \  
4   --data "host=payments.internal" \  
5   --data "port=443" \  
6   --data "path=/api/v1" \  
7   --data "connect_timeout=5000" \  
8   --data "read_timeout=30000" \  
9   --data "write_timeout=30000" \  
10  --data "retries=3"
```

Services are the anchor point for many Kong features. Plugins can be attached to a Service to apply policies to all traffic destined for that backend, regardless of which Route the traffic entered through.

Routes

A Route defines how incoming client requests are matched and mapped to Services. While a Service answers the question “where should requests go?”, a Route answers “which requests should go there?” A single Service can have multiple Routes, allowing you to expose the same backend through different URL patterns, hostnames, or protocols.

Routes support several matching criteria:

- **paths:** URL path prefixes or regular expressions
- **hosts:** Host header values (domain names)
- **methods:** HTTP methods (GET, POST, PUT, DELETE, etc.)

- **headers:** Custom header name-value pairs
- **protocols:** Supported protocols (http, https, grpc, grpcs)
- **snis:** Server Name Indication values for TLS routing
- **expression:** Advanced matching using Kong's expression language

When a request arrives at Kong, the router evaluates all configured Routes and selects the one with the highest priority that matches the request. Priority is calculated based on how specific the matching criteria are: a Route that matches both hosts and paths has higher priority than one that matches only paths, and exact matches beat wildcard or regex matches.

Here is how to create a Route for the user-service we defined earlier:

```
1 curl -i -X POST http://localhost:8001/services/user-service/routes \
2   --data "name=user-api-route" \
3   --data "paths[]=api/users" \
4   --data "methods[]=GET" \
5   --data "methods[]=POST"
```

This Route matches any GET or POST request to paths starting with /api/users and forwards it to the user-service. By default, Kong strips the matching path prefix before forwarding the request to the upstream, so a request to /api/users/123 becomes /123 at the backend. You can disable this behavior with `strip_path=false` if your backend expects the full path.

For more complex routing, you can use regular expressions in paths:

```
1 curl -i -X POST http://localhost:8001/services/user-service/routes \
2   --data "name=user-id-route" \
3   --data 'paths[]=~/api/users/\d+$' \
4   --data "regex_priority=100"
```

The `regex_priority` field controls the order in which regex routes are evaluated when multiple regex patterns match the same request. Higher values mean higher priority.

Kong also supports an expression-based router that provides more flexible matching using a domain-specific language. This allows you to combine conditions with logical operators:

```
1 curl -i -X POST http://localhost:8001/routes \
2   --data "name=advanced-route" \
3   --data 'expression=(http.path ~= "/api/v2/.*)" && (http.method == "POST")' \
4   --data "service=user-service"
```

This expression matches any POST request to paths starting with `/api/v2/`, using the `~=` prefix match operator.

Upstreams

While a Service represents a single backend endpoint, an Upstream provides load balancing and health checking across multiple instances of a service. An Upstream is essentially a virtual hostname that resolves to a pool of Targets (individual backend servers). When you configure a Service to use an Upstream instead of a direct URL, Kong distributes traffic across the healthy targets using the configured load balancing algorithm.

Upstreams are essential for production deployments where services run as multiple replicas. They enable:

- Load balancing across multiple backend instances
- Active and passive health checking
- Automatic failover when targets become unhealthy
- Integration with service discovery systems (Consul, etcd, Kubernetes DNS)

Here is how to create an Upstream with Targets:

```
1 # Create the upstream
2 curl -i -X POST http://localhost:8001/upstreams \
3   --data "name=user-api-upstream" \
4   --data "algorithm=round-robin"
5
6 # Add targets
7 curl -i -X POST http://localhost:8001/upstreams/user-api-upstream/targets \
8   --data "target=10.0.1.10:8080" \
9   --data "weight=100"
10
11 curl -i -X POST http://localhost:8001/upstreams/user-api-upstream/targets \
12   --data "target=10.0.1.11:8080" \
13   --data "weight=100"
14
15 curl -i -X POST http://localhost:8001/upstreams/user-api-upstream/targets \
16   --data "target=10.0.1.12:8080" \
17   --data "weight=50"
```

This creates an upstream with three targets. The first two have equal weight (100), while the third has half the weight (50), meaning it will receive approximately one-third of the traffic under round-robin load balancing.

Kong supports several load balancing algorithms:

- **round-robin**: Distributes requests evenly in a circular pattern
- **least-connections**: Sends requests to the target with the fewest active connections
- **consistent-hashing**: Uses a hash of request attributes (IP, header, path, etc.) to consistently route to the same target
- **latency-based**: Routes to targets with the lowest response latency
- **request-count**: Distributes based on total request counts

To use an Upstream in a Service, you set the Service's host to the upstream name:

```
1 curl -i -X PATCH http://localhost:8001/services/user-service \
2   --data "host=user-api-upstream" \
3   --data "port=8080" \
4   --data "protocol=http"
```

Now Kong will load balance requests across all healthy targets in the user-api-upstream.

Consumers

A Consumer represents an identity that accesses your APIs through Kong. Consumers are not necessarily individual users; they can be applications, services, teams, or any entity that needs to authenticate and consume your APIs. Most of Kong's authentication plugins work by associating credentials with Consumers.

Consumers enable per-identity policies: different rate limits for different API keys, access control lists scoped to specific consumers, custom plugin configurations applied only to certain identities.

Here is how to create a Consumer and associate an API key credential:

```
1 # Create the consumer
2 curl -i -X POST http://localhost:8001/consumers \
3   --data "username=mobile-app" \
4   --data "custom_id=app-mobile-prod-001"
5
6 # Add an API key credential
7 curl -i -X POST http://localhost:8001/consumers/mobile-app/key-auth \
8   --data "key=mobile-app-secret-key-2024"
```

Now any request that includes the header `X-API-Key: mobile-app-secret-key-2024` will be authenticated as the `mobile-app` consumer. You can then apply rate limiting, ACLs, or other policies specifically to this consumer.

Consumers can have multiple types of credentials (API keys, JWT secrets, OAuth2 tokens, etc.), and they can belong to groups for easier management. The `custom_id` field is particularly useful when you need to correlate Kong consumers with identities in your own systems.

Configuration Models: Database-Backed, DB-less, Hybrid

Kong supports three primary configuration models, each suited to different operational requirements. Choosing the right model is one of the most important architectural decisions for a Kong deployment.

Database-Backed Mode (Traditional)

In database-backed mode, Kong stores all its configuration (Services, Routes, Plugins, Consumers, etc.) in a PostgreSQL database. Every Kong node connects to the database and synchronizes its in-memory cache with the stored configuration. The Admin API writes directly to the database, allowing dynamic configuration changes at runtime.

This is the most flexible mode. You can:

- Use the full Admin API for CRUD operations on all entities
- Make configuration changes without restarting Kong nodes
- Use Kong Manager (the web UI) for visual management
- Leverage plugins that require database storage

The tradeoff is operational complexity: you must deploy, manage, and maintain a PostgreSQL cluster. For production deployments, this typically means a replicated PostgreSQL setup with backups, monitoring, and failover procedures.

Here is a minimal Docker Compose configuration for database-backed Kong:

```
1  version: "3.8"
2  services:
3    kong-database:
4      image: postgres:15-alpine
5      restart: always
6      environment:
7        POSTGRES_DB: kong
8        POSTGRES_USER: kong
9        POSTGRES_PASSWORD: kongpassword
10     volumes:
11       - kong-db-data:/var/lib/postgresql/data
12     networks:
13       - kong-net
14
15   kong-migrations:
16     image: kong:3.9
17     restart: on-failure
18     command: "kong migrations bootstrap"
19     environment:
20       KONG_DATABASE: postgres
21       KONG_PG_HOST: kong-database
22       KONG_PG_PASSWORD: kongpassword
23     depends_on:
24       - kong-database
25     networks:
26       - kong-net
27
28   kong:
29     image: kong:3.9
30     restart: always
31     ports:
32       - "8000:8000"
33       - "8443:8443"
34       - "8001:8001"
35       - "8444:8444"
36     environment:
37       KONG_DATABASE: postgres
38       KONG_PG_HOST: kong-database
39       KONG_PG_PASSWORD: kongpassword
40       KONG_PROXY_ACCESS_LOG: /dev/stdout
41       KONG_ADMIN_ACCESS_LOG: /dev/stdout
```

```
42     KONG_PROXY_ERROR_LOG: /dev/stderr
43     KONG_ADMIN_ERROR_LOG: /dev/stderr
44     depends_on:
45       - kong-migrations
46     networks:
47       - kong-net
48
49     volumes:
50       kong-db-data:
51
52     networks:
53       kong-net:
54         driver: bridge
```

The migrations step initializes the database schema before Kong starts. In production, you would run migrations as a separate job during upgrades rather than bootstrapping every time.

DB-less Mode

DB-less mode eliminates the database dependency entirely. Kong runs with all configuration loaded from a declarative YAML or JSON file into in-memory storage. The Admin API becomes read-only (you cannot create or modify entities through it), and all configuration changes are made by updating the declarative file and reloading Kong.

This mode is ideal for:

- Containerized environments where statelessness is preferred
- GitOps workflows where configuration lives in version control
- Simple deployments that do not need dynamic Admin API updates
- Kubernetes deployments managed by the Kong Ingress Controller

The tradeoffs are:

- No write access to the Admin API (configuration must be managed externally)
- Some plugins that require database storage do not work
- Configuration reloads can briefly impact performance as Kong rebuilds its routing tables

Here is a DB-less configuration file:

```
1  _format_version: "3.0"
2  services:
3    - name: user-service
4      url: http://user-api.internal:8080
5      routes:
6        - name: users-route
7          paths:
8            - "/api/users"
9          methods:
10             - GET
11             - POST
12      plugins:
13        - name: prometheus
14  consumers:
15    - username: mobile-app
16      key-auth_credentials:
17        - key: mobile-api-key-12345
```

To run Kong in DB-less mode with this configuration:

```
1  version: "3.8"
2  services:
3    kong:
4      image: kong:3.9
5      restart: always
6      ports:
7        - "8000:8000"
8        - "8443:8443"
9        - "8001:8001"
10     volumes:
11       - ./kong.yml:/usr/local/kong/declarative/kong.yml
12     environment:
13       KONG_DATABASE: "off"
14       KONG_DECLARATIVE_CONFIG: /usr/local/kong/declarative/kong.yml
15       KONG_PROXY_ACCESS_LOG: /dev/stdout
16       KONG_ADMIN_ACCESS_LOG: /dev/stdout
```

Kong polls the declarative configuration file for changes (default interval is 60 seconds) and automatically reloads when it detects updates. You can also trigger a reload via the Admin API endpoint `POST /config`.

Hybrid Mode

Hybrid mode combines the best of both worlds by separating Kong into two roles: Control Plane (CP) nodes and Data Plane (DP) nodes. Control Plane nodes

connect to the database, serve the Admin API, and manage configuration. Data Plane nodes run in DB-less mode, receive their configuration from the Control Plane via a secure WebSocket connection, and handle all proxy traffic.

This architecture enables:

- Deploying Data Planes in isolated networks, edge locations, or different cloud regions without exposing them to the database
- Scaling Data Planes independently of the Control Plane
- Centralized management with distributed execution
- Improved security by keeping the database behind a secure network

In hybrid mode, communication between Control Plane and Data Plane is secured with mutual TLS. Data Planes fetch their configuration from the Control Plane and can continue serving traffic with their last known configuration if connectivity is temporarily lost.

Here is how you configure a Control Plane node in `kong.conf`:

```
1  role = control_plane
2  database = postgres
3  pg_host = 127.0.0.1
4  pg_user = kong
5  pg_password = kongpassword
6  pg_database = kong
7  cluster_cert = /etc/kong/certs/tls.crt
8  cluster_cert_key = /etc/kong/certs/tls.key
```

And a Data Plane node:

```
1  role = data_plane
2  database = off
3  cluster_control_plane = cp-node.internal:8005
4  cluster_cert = /etc/kong/certs/tls.crt
5  cluster_cert_key = /etc/kong/certs/tls.key
6  cluster_server_name = cp-node.internal
```

The `cluster_cert` and `cluster_cert_key` are shared certificates used for mTLS between CP and DP nodes. The Control Plane listens on port 8005 by default for Data Plane connections.

Kong Konnect, the cloud-managed control plane offering, uses a hybrid mode architecture where Kong hosts the Control Plane in their SaaS environment and you deploy your own Data Planes wherever needed.

The Plugin System Architecture

Plugins are what make Kong extensible and powerful. At their core, they are Lua modules that hook into the request lifecycle phases we discussed earlier, adding functionality without requiring you to modify Kong's core code. This design is not accidental; it reflects a deliberate architectural choice to keep the gateway itself lean while making it trivially easy to extend.

The plugin system rests on four interlocking principles that shape how you should think about using Kong. First, plugins are phase-based: each one declares which lifecycle phases it participates in, whether that is access for authentication, rewrite for URL manipulation, log for metrics collection, or any combination thereof. Second, they execute in a strict priority order determined by numeric values, with lower numbers running first. This ordering matters when plugins depend on each other; authentication must complete before authorization can make decisions, and rate limiting should happen after authentication so limits can be applied per consumer rather than per IP address. Third, plugins operate within a scope hierarchy that moves from broad to specific: global configuration applies to all traffic, service-level configuration targets a particular backend, route-level configuration narrows further to specific endpoints, and consumer-level configuration personalizes behavior for individual identities. Fourth, all plugins interact with Kong through the Plugin Development Kit (PDK), a stable API layer that abstracts away the underlying Nginx complexity. This means plugins written today will continue to work after Kong upgrades, and you can reason about plugin behavior without needing to understand every detail of OpenResty internals.

When Kong starts up, it loads all enabled plugins into memory. For each incoming request, Kong determines which plugins apply by examining the matched route and authenticated consumer, then executes them in the appropriate phases according to their priority values. This happens entirely within the Nginx event loop, which is why plugin execution adds only microseconds of latency rather than milliseconds.

The scope hierarchy has an important implication for how configuration works: when the same plugin type is configured at multiple levels, Kong does not run it multiple times. Instead, the most specific configuration wins. If you configure rate limiting globally at 100 requests per minute but also on a specific route at 10 requests per minute, requests matching that route are limited to 10 per minute while all other traffic follows the global limit. This precedence

model prevents duplicate processing and ensures predictable behavior, but it can be subtle when you first encounter it. Always remember: route beats service, service beats consumer, and consumer beats global.

Plugins are controlled via the `plugins` directive in `kong.conf`. Setting `plugins = bundled` loads every plugin that ships with Kong, which is fine for development but wasteful in production where each loaded plugin consumes memory even if never configured. A more disciplined approach specifies only what you need:

```
1 plugins = bundled,key-auth,rate-limiting,prometheus
```

You can also explicitly disable plugins you know you will not use by prefixing their names with a minus sign:

```
1 plugins = bundled,-grpc-gateway,-graphql-post
```

This reduces Kong's memory footprint and attack surface. In large production deployments, I recommend auditing your plugin list quarterly and removing anything that has not been actively used.

Custom plugins extend this model further by letting you place your own Lua code in the `custom_plugins` directory and reference them in the `plugins` list. We will cover custom plugin development in depth in Chapter 6, including how to structure your code, handle errors gracefully, and test plugins before deploying them to production.

The plugin system is genuinely Kong's defining feature. It is what transforms a capable reverse proxy into a full-featured API platform. Authentication, rate limiting, traffic shaping, observability, AI routing, protocol translation – all of these are just plugins in Kong's vocabulary. Once you internalize how the plugin model works, you can approach almost any API infrastructure problem by asking whether it can be solved as a plugin rather than as application code.

With this foundation in place, we are ready to start deploying Kong. The next chapter covers every installation and deployment option in detail, from simple Docker setups to production-grade hybrid architectures.

Chapter 2: Installation and Deployment

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Prerequisites and Environment Setup

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Docker and Docker Compose Deployments

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Database-Backed Docker Deployment

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

DB-less Docker Deployment

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Database-Backed Mode with PostgreSQL

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

PostgreSQL Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Database Initialization

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Connection Pooling

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

DB-less Mode with YAML Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Complete Declarative Configuration Example

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Hot Reload Behavior

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Hybrid Mode: Control Plane and Data Plane

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Architecture Overview

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Generating Cluster Certificates

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Control Plane Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Data Plane Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Hybrid Mode Benefits

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong Konnect Cloud Deployment

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Getting Started with Konnect

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Konnect Data Plane Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Konnect vs Self-Managed

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Choosing Your Deployment Model

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong Manager: The Web UI for Kong Administration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Enabling Kong Manager in Docker Compose

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Securing Kong Manager with RBAC and Basic Auth

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Creating RBAC Users with Custom Roles

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Using OIDC for Kong Manager Authentication (Enterprise)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong Manager Workflows: Key Tasks

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong Manager and Declarative Configuration: Best Practices

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 3: Routing, Services, and Traffic Management

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Defining Services and Routes

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Service Configuration Options

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Route Configuration Options

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Route Matching Strategies

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Path-Based Routing

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Host-Based Routing

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Header-Based Routing

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Method-Based Routing

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Combined Matching

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Expression-Based Routing

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Upstreams and Load Balancing Algorithms

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Creating Upstreams and Targets

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Load Balancing Algorithms

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Health Checks: Active and Passive

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Active Health Checks

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Passive Health Checks (Circuit Breakers)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Service Discovery Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

DNS-Based Discovery

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Consul Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

etcd Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Zookeeper Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kubernetes Service Discovery

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

API Versioning Strategies

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

URI Path Versioning

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Header-Based Versioning

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Content Negotiation Versioning

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Query Parameter Versioning

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Versioning Best Practices with Kong

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Path Rewriting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Header Transformation

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

CORS Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 4: Authentication and Authorization

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

API Keys and Basic Authentication

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Key Authentication

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Basic Authentication

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

HMAC Authentication

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

JWT Validation and Issuance

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Validating External JWTs

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

JWT Claims Validation

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

OAuth 2.0 and OIDC Flows

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Authorization Code Flow

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Client Credentials Flow

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

OpenID Connect with Keycloak

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

LDAP and Active Directory Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Access Control Lists (ACL)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Custom Authentication with Lua Plugins

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 5: Rate Limiting, Throttling, and Traffic Shaping

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Rate Limiting Algorithms: Local, Redis, Cluster, Sentinel

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Local Rate Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Redis Rate Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Redis Cluster and Sentinel Support

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Cluster Rate Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Per-Consumer and Per-Route Limits

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Per-Consumer Rate Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Per-IP Rate Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Per-Route Rate Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Burst Handling and Token Bucket

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Sliding Window Rate Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Token Bucket Behavior

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Load Shedding Strategies

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Response Rate Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Priority-Based Rate Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Circuit Breaking with Health Checks

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Rate Limit Response Headers

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 6: Plugins Ecosystem

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Plugin Phases and the Request Lifecycle

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Built-in Plugins Overview

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Authentication and Authorization Plugins

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Traffic Control Plugins

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Transformation Plugins

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Observability and Logging Plugins

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AI-Specific Plugins (AI Gateway)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Configuring and Chaining Plugins

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Plugin Priority and Execution Order

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Writing Custom Lua Plugins

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Plugin Structure

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Example: Response Header Injection Plugin

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Example: External Service Validation Plugin

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Third-Party Plugins from Kong Hub

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 7: Security, TLS, and API Governance

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Mutual TLS (mTLS) Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Setting Up mTLS with Kong

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

mTLS with Kubernetes

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Certificate Management and Auto-Renewal

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Uploading Certificates via Admin API

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Certificate Rotation Strategy

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

IP Restriction and Geo-Blocking

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Basic IP Allow List

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

IP Deny List

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Per-Consumer IP Restrictions

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Geo-Blocking

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Bot Detection and Abuse Prevention

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Request/Response Body Filtering

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Request Size Limiting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Body Content Filtering with Custom Plugin

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

CORS, CSRF, and Cross-Origin Security

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

CORS Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

CSRF Protection

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Vulnerability Scanning Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 8: Observability, Logging, Monitoring, and Analytics

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Access Logs and Structured Logging

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Default Log Format

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Custom Log Formats

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Structured JSON Logging

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Log Streaming to External Systems

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Log What Matters

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Prometheus Metrics and Grafana Dashboards

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Enabling Prometheus Metrics

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Key Metrics

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Prometheus Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Grafana Dashboard

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Distributed Tracing with OpenTelemetry

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Enabling OpenTelemetry

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Tracing Flow

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Example Trace

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Log Streaming to ELK, Splunk, Datadog

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

ELK Stack (Elasticsearch, Logstash, Kibana)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Splunk Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Datadog Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong Analytics and Konnect Insights

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Custom Metrics with Plugins

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 9: Kubernetes Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong Ingress Controller Architecture

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Helm Chart Installation and Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Installing Kong Ingress Controller with Helm

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Custom Values File

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Ingress Resources and Annotations

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Basic Ingress

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong-Specific Annotations

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong Gateway API Implementation

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

GatewayClass and Gateway

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

HTTPRoute Resources

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Advanced Routing with HTTPRoute

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

TCPRoute and UDPRoute

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Plugin Configuration via Custom Resources

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

KongPlugin (Namespace-Scoped)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

KongClusterPlugin (Cluster-Scoped)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Plugin on HTTPRoute

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

KongConsumer for Authentication

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Multi-Cluster and Edge Deployment Patterns

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Global Ingress with Multiple Clusters

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Edge Caching Pattern

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 10: Automation, CI/CD, and GitOps

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

The Admin API Reference

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Core Endpoints

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Creating Configuration via Admin API

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Admin API Security

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

deck Configuration Management

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Installing deck

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

deck Configuration File Format

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Core deck Commands

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

deck with Konnect

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Multi-File Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Declarative Config and Version Control

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Repository Structure

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Environment-Specific Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Terraform Provider for Kong

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Official Konnect Provider

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Self-Managed Provider (kevholditch/kong)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Terraform + deck Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

CI/CD Pipeline Patterns

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

GitHub Actions Workflow

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

GitOps with Argo CD

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Environment Promotion Strategies

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 11: Performance Tuning, Scalability, and High Availability

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

OpenResty and Nginx Tuning Parameters

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Worker Processes

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Worker Connections

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

File Descriptor Limits

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Shared Memory Zones

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Custom Nginx Template

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Connection Pooling and Keep-Alive Optimization

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Service-Level Keepalive

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Upstream Connection Pool Settings

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

HTTP/2 to Upstream

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Performance War Story: The Plugin That Killed Our Black Friday

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Horizontal Scaling Strategies

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Database-Backed Scaling

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

DB-less Scaling

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Hybrid Mode Scaling

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kubernetes HPA

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

High Availability Architecture

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Load Balancer in Front

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

PostgreSQL High Availability

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Multi-Region Architecture

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Disaster Recovery

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 12: Kong AI Gateway and LLM Orchestration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

The AI Gateway Problem

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Getting Started with AI Gateway

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

OSS vs Enterprise Plugin Availability

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Basic LLM Routing

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Using deckK for AI Gateway Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Intelligent Model Routing Strategies (Enterprise)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Multi-Target Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Load Balancing Algorithms

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Multi-Provider Orchestration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Provider Format Translation

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Model Aliases

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Fallback and Retry Mechanisms

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Automatic Failover

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Retry Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Prompt Transformation and Template Management

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

System Prompt Injection

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Prompt Templates

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Response Caching for AI Workloads (Enterprise)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AI Semantic Cache Configuration (Enterprise)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Guardrails, Safety, and Content Moderation

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AI Prompt Guard Plugin (OSS)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AI Semantic Prompt Guard (Enterprise)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AWS Bedrock Guardrails Integration (Enterprise)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Token Tracking and Cost Optimization

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Token Statistics Logging (OSS)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AI Rate Limiting by Tokens (Enterprise)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Cost Optimization Strategies

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AI Observability and Governance

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AI-Specific Metrics

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AI Request Logging

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Governance Policies

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Policy-as-Code with Open Policy Agent (OPA)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Centralized Prompt Template Management

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Model Drift Detection via Kong Metrics

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Compliance Auditing for AI Decisions

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 13: AI Integrations and Framework Patterns

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

OpenAI and Azure OpenAI Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Basic OpenAI Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Azure OpenAI Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Application Code (Python)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Anthropic Claude Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Anthropic Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Multi-Provider with Anthropic Fallback (Enterprise)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Google Gemini Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Gemini Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AWS Bedrock Model Access

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Bedrock Configuration for Claude

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Bedrock with Guardrails

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

IAM Configuration for EKS

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Self-Hosted Models: Ollama, vLLM, Hugging Face

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Ollama Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

vLLM Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Hugging Face Inference Endpoints

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

LangChain Orchestration with Kong

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Basic LangChain Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Multi-Model LangChain with Kong Routing

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

LangGraph with Kong AI Gateway

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

LlamaIndex and RAG Patterns

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

RAG Pipeline with Kong

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Redis Vector Store with Kong and LlamaIndex

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Streaming Responses and SSE

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Enabling Streaming

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Application Code with Streaming

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Chapter 14: Production Patterns, Migrations, and Troubleshooting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Application Integration Patterns

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Spring Boot Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Node.js Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Python (FastAPI) Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Go Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

.NET Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

gRPC and GraphQL Gateway Patterns

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

gRPC Gateway Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

GraphQL Gateway Configuration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kafka Event Streaming with Kong

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

TCP Proxy for Native Kafka Clients

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

HTTP-to-Kafka Mediation with kafka-upstream Plugin

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong Event Gateway for Native Kafka Protocol Management

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kafka Consumer Group Handling

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Common Failure Modes and Troubleshooting

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

No Route Matched (404)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Upstream Request Failed (502/503/504)

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Plugin Errors

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Performance Issues

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Upgrade Strategies and Version Migration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Safe Upgrade Process

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Blue-Green Upgrade

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Version-Specific Considerations

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Real-World Migration Case Study: From Nginx Ingress to Kong

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Configuration Auditing and Compliance

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

deck Drift Detection

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Audit Logging

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Production Checklist

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Conclusion: The Future of API Infrastructure

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Key Takeaways

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Emerging Trends in API Architecture

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Where Kong Is Heading

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Final Recommendations

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

References

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Core Kong Documentation

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kong Manager and RBAC

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

AI Gateway and AI Plugins

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Installation and Deployment

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Kubernetes Integration

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Automation and Configuration Management

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Observability and Monitoring

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Underlying Technologies

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Community and Support

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

Index

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

A

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

B

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

C

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

D

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

E

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

F

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

G

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

H

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

I

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

K

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

L

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

M

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

N

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

O

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

P

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

R

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

S

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

T

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

U

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

V

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.

W

This content is not available in the sample book. The book can be purchased on Leanpub at <https://leanpub.com/masteringkongapigateway>.