

# *Load Balancer e API Gateway*

*Ramon Santos*

# Conteúdo

## 1 Camadas de redes no modelo TCP/IP

- ▶ Camadas de redes no modelo TCP/IP
- ▶ Proxy de encaminhamento vs proxy reverso
- ▶ Load balancer
- ▶ API Gateway
- ▶ Referências

# Camadas de redes no modelo TCP/IP

## 1 Camadas de redes no modelo TCP/IP

- Camada de aplicação:
  - contém os protocolos de alto nível;
  - HTTP, DNS, SMTP, FTP e etc.
- Camada de transporte:
  - responsável pela comunicação end-to-end entre dois dispositivos;
  - recebe dados da camada de aplicação, os divide em unidades menores e repassa para a camada de rede;
  - TCP e UDP.

# Camadas de redes no modelo TCP/IP

## 1 Camadas de redes no modelo TCP/IP

- Camada de internet (ou camada de rede):
  - faz o roteamento de pacotes entre diferentes redes;
  - IP e ICMP.
- Camada de enlace:
  - lida com a comunicação física entre os dispositivos;
  - Ethernet (via cabo), 802.11 (sem fio) e etc.

# Conteúdo

## 2 Proxy de encaminhamento vs proxy reverso

- ▶ Camadas de redes no modelo TCP/IP
- ▶ Proxy de encaminhamento vs proxy reverso
- ▶ Load balancer
- ▶ API Gateway
- ▶ Referências

# Forward Proxy (encaminhamento)

## 2 Proxy de encaminhamento vs proxy reverso

- Atua entre o cliente e a internet É usado para controle de acesso, filtragem de conteúdo, anonimato e cache.

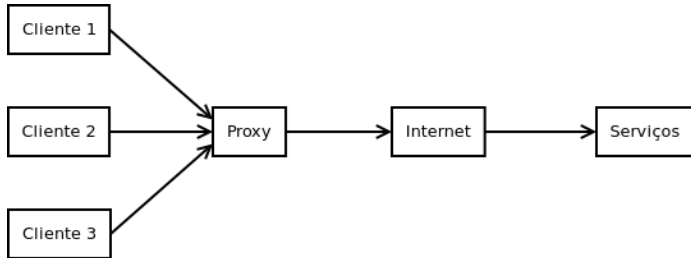


Figure: Forward proxy

# Proxy reverso

## 2 Proxy de encaminhamento vs proxy reverso

- Um proxy reverso é um servidor que recebe as requisições dos clientes e as redireciona para os servidores de backend apropriados.

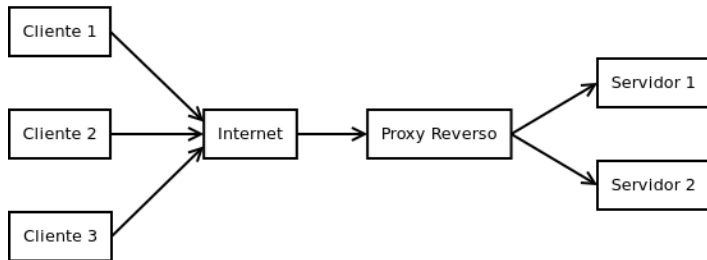


Figure: Proxy reverso

# Conteúdo

## 3 Load balancer

- ▶ Camadas de redes no modelo TCP/IP
- ▶ Proxy de encaminhamento vs proxy reverso
- ▶ **Load balancer**
- ▶ API Gateway
- ▶ Referências



# Load balancing

## 3 Load balancer

Load balancing (balanceamento de carga) é o processo de redistribuição da carga de trabalho entre os nós de um sistema distribuído para melhorar tanto a utilização dos recursos, quanto o tempo de resposta das tarefas, evitando também situações em que alguns nós fiquem muito carregados enquanto outros fiquem ociosos ou realizando pouco trabalho.

# Características de um load balancing

## 3 Load balancer

- Distribuir tráfego/carga entre múltiplos servidores.
- Possibilita a escalabilidade horizontal.
- Segurança: Separa tráfego público de tráfego privado.
- Pode atuar nas camadas de aplicação e transporte.
- Realizam health checks para verificar se um servidor consegue receber requisições.
- Possibilita manutenção de servidores sem downtime.

# Mecanismos de balanceamento

## 3 Load balancer

- **Round Robin:** Envia requisições sequencialmente entre os servidores.
- **Least Connections:** Escolhe o servidor com menos conexões ativas.
- **IP Hash:** Uma função hash é usada para determinar qual servidor deve receber a próxima request com base no endereço IP do cliente.
- **Weighted load balancing:** Usa o algoritmo Round Robin com diferentes pesos nos servidores. É útil quando determinado servidor conta com mais capacidade computacional que os demais.
- **Least Time:** Escolhe o servidor com o menor tempo médio de resposta.

# Ferramentas

## 3 Load balancer

- **Nginx:** Leve e extensível (com Lua). Muito usado para balancear carga HTTP.
- **HAProxy:** Além de ser usado para balanceamento de carga na camada de aplicação, é bastante popular no balanceamento de carga na camada de transporte. Como, por exemplo, balanceamento de conexão de leitura entre réplicas read-only do PostgreSQL.
- **Cloudflare Load Balancing:** É um serviço de balanceamento de carga global, que tem como principal característica o um roteamento inteligente e personalizável, onde a região geográfica, ou até mesmo as coordenadas do GPS do dispositivo que fez a requisição, podem ser considerados para encontrar o nó de menor latência.
- **AWS Elastic Load Balancing:**
  - **Application Load Balancer:** Balanceador de carga da camada de aplicação.
  - **Network Load Balancer:** Balanceador de carga da camada de transporte.

# Conteúdo

## 4 API Gateway

- ▶ Camadas de redes no modelo TCP/IP
- ▶ Proxy de encaminhamento vs proxy reverso
- ▶ Load balancer
- ▶ **API Gateway**
- ▶ Referências

# API Gateway

## 4 API Gateway

Um API Gateway é um ponto de entrada único para gerenciar e encaminhar requisições a diferentes serviços de uma aplicação em um sistema distribuído.

# Características de um API gateway

## 4 API Gateway

- Pode fazer balanceamento de carga.
- Atua na camada de aplicação.
- Faz tradução de protocolo.
- Monitoramento, Logging e Analytics.
- Gerenciamento da API:
  - roteia requisições para serviços mapeados;
  - versionamento de API;
  - adiciona headers e query params;
  - throttling e rate limiting;
  - transformação e agregação de resposta.
- Diminui carga de trabalho nos serviços:
  - pode responder a requisições com dados em cache;
  - pode lidar com autenticação e autorização;
  - validação de requisições;
  - não repassaria requisições de um ataque DDoS para serviços.

# Desafios e desvantagens

## 4 API Gateway

- Adiciona uma nova camada da arquitetura:
  - mais um componente para o time manter;
  - mais ferramentas que as pessoas do time devem conhecer;
  - cada request passa por uma nova camada, o que pode gerar latência adicional;
  - aumenta o custo financeiro com ferramentas gerenciadas ou infra em soluções self-hosted.
- É um ponto único de falha:
  - se o gateway cair, todos os serviços ficam inoperantes;
  - erros de configuração podem comprometer segurança e desempenho;
  - é necessário buscar alta disponibilidade, mediante técnicas como replicação e balanceamento, ou usar serviços gerenciados que suportem esse requisito.



# Ferramentas

## 4 API Gateway

- **Kong:** Funciona com Nginx e é extensível através de plugins escritos em Lua.
- **AWS API Gateway:** Serviço de API Gateway gerenciado da AWS.

# Conteúdo

## 5 Referências

- ▶ Camadas de redes no modelo TCP/IP
- ▶ Proxy de encaminhamento vs proxy reverso
- ▶ Load balancer
- ▶ API Gateway
- ▶ Referências

# Referências

## 5 Referências

- **Documentação no Nginx sobre load balancer:**  
<https://docs.nginx.com/nginx/admin-guide/load-balancer/http-load-balancer/>
- **Capítulo 20 do Livro *Engenharia de Confiabilidade do Google*:** (Existe a versão em português do livro pela Novatec e o livro está disponível de graça em inglês neste link <https://sre.google/sre-book/table-of-contents/>).
- **Post sobre API Gateway *What Is an API Gateway?*:**  
<https://www.paloaltonetworks.co.uk/cyberpedia/what-is-api-gateway>

*Obrigado!*