



# Load Balancing e Escalabilidade

CLOUD COMPUTING

# Load Balancing



Técnica para otimizar o uso de recursos na cloud



Evitar que um recurso específico fique sobrecarregado



Distribuir o trabalho entre servidores, VMs, containers



Network Load Balancing



Application Load Balancing



Database Load Balancing



Ótimo para lidar com picos de requisições e falhas em servidores



# Vantagens



Melhora a performance



Reduzir a carga de cada recurso possibilitando maior performance



Maior disponibilidade, evitando recursos com falha



Escalabilidade



Utilização eficiente dos recursos



# Desvantagens



Complexo



Custo alto



O load balancing em si é um ponto sensível a falha



Segurança



# Vários servidores ou apenas um melhorado?



Quando as requisições da sua aplicação aumentam, existem dois caminhos



Melhorar o servidor



Adicionar vários servidores e distribuir a carga entre eles



Melhorar o servidor normalmente reduz o problema temporariamente



# Hardware e Software

---



Load balancing pode ser feito em ambos



Algoritmo que direciona carga ou hardware que analisa o tráfego da rede

# Escalabilidade

---

Formas de  
aumentar o poder  
de processamento  
dos recursos

Pode ser feita de  
forma horizontal  
ou vertical

# Escalabilidade Vertical

---



AUMENTAR OS  
RECURSOS DE UM  
ÚNICO SERVIDOR



COMO AUMENTAR A  
RAM OU CPU DE UM  
SERVIDOR



RÁPIDO PARA  
EXECUTAR



UMA HORA CHEGARÁ  
NO LIMITE



# Escalabilidade Horizontal

---



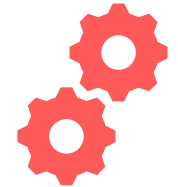
Adicionar mais recursos  
e dividir a carga de  
trabalho



Como por exemplo subir  
mais uma instância de  
um servidor na nuvem



Flexível com cargas de  
trabalho que variam e  
promove redundância



Requer load balancing  
que é uma técnica  
complexa