

---

---

# Escalabilidade e Elasticidade com *AutoScaling e Elastic Load Balancer*

— Prof. Raphael Rodrigues Pereira —

---

---

# Conceitos

## **Escalabilidade:**

Capacidade de adicionar/remover recursos para aumentar a capacidade de uma aplicação.

## Escalabilidade vertical

(Adição de capacidade para um único recurso)

- + processador
- + memória
- + disco rígido



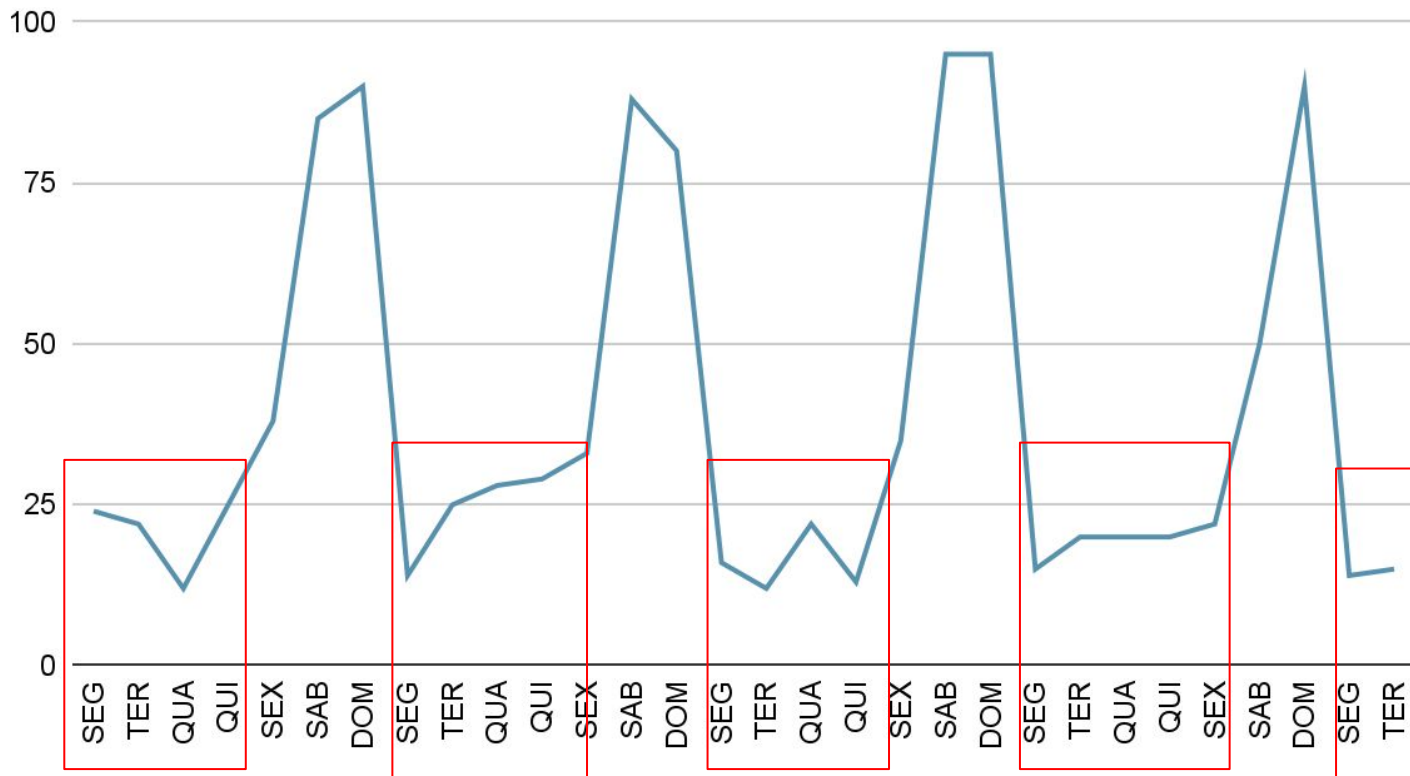
**8GB RAM**  
**2 CPU**

**16GB RAM**  
**4 CPU**

Para alterar os recursos é preciso desligar o servidor

**INDISPONIBILIDADE TEMPORÁRIA**

## % CPU Usage



Server = 10 CPUs

35% = 4 CPUs  
atenderia de Seg a Qui

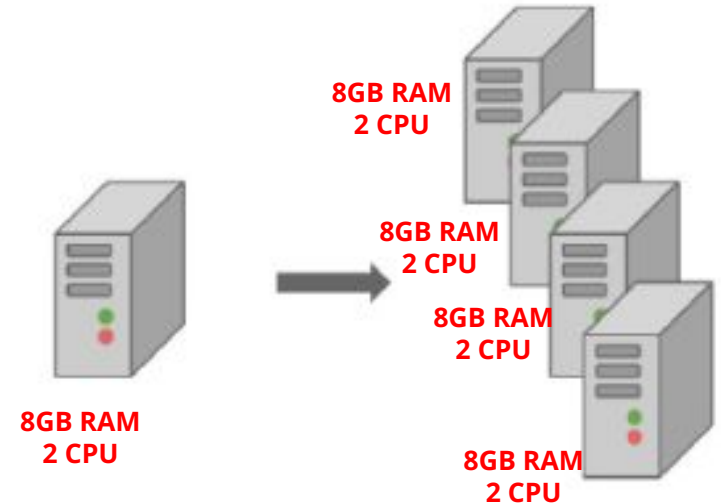
19 dias ocioso  
(63% do tempo)

# Qual melhor solução para isso?

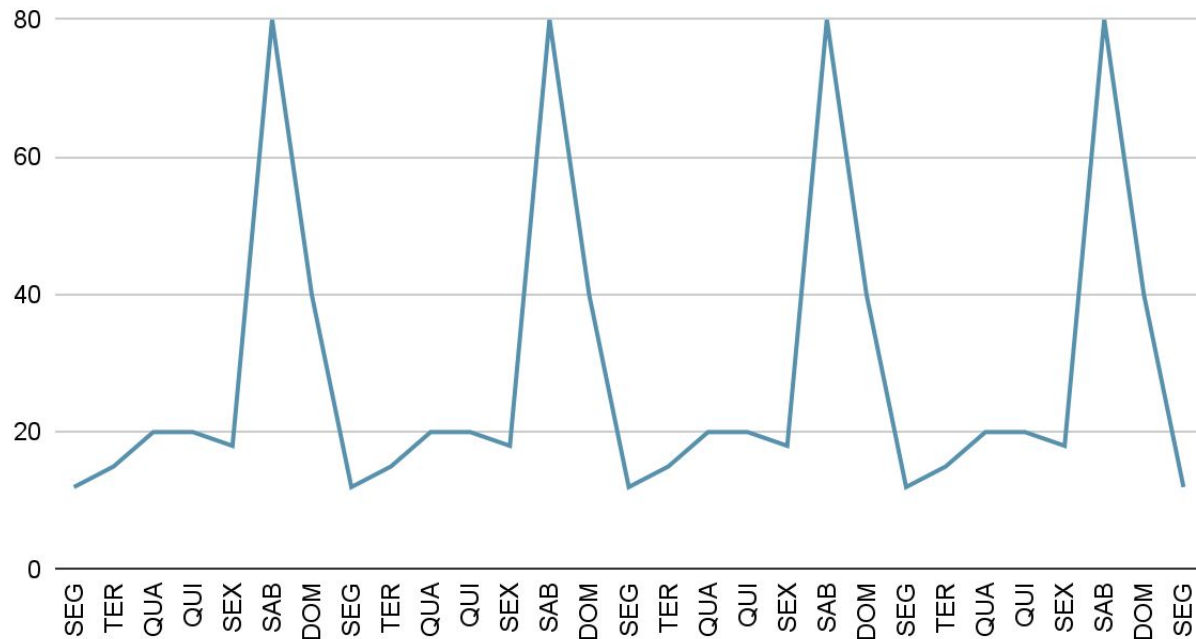
Horizontal

## Escalabilidade horizontal

(Adição de recursos ao ambiente computacional)



## % CPU Usage



Criar novas instâncias quando a carga chegar a 80%, por exemplo.

**Mas quem vai  
criar as novas  
instâncias?**



# Conceitos

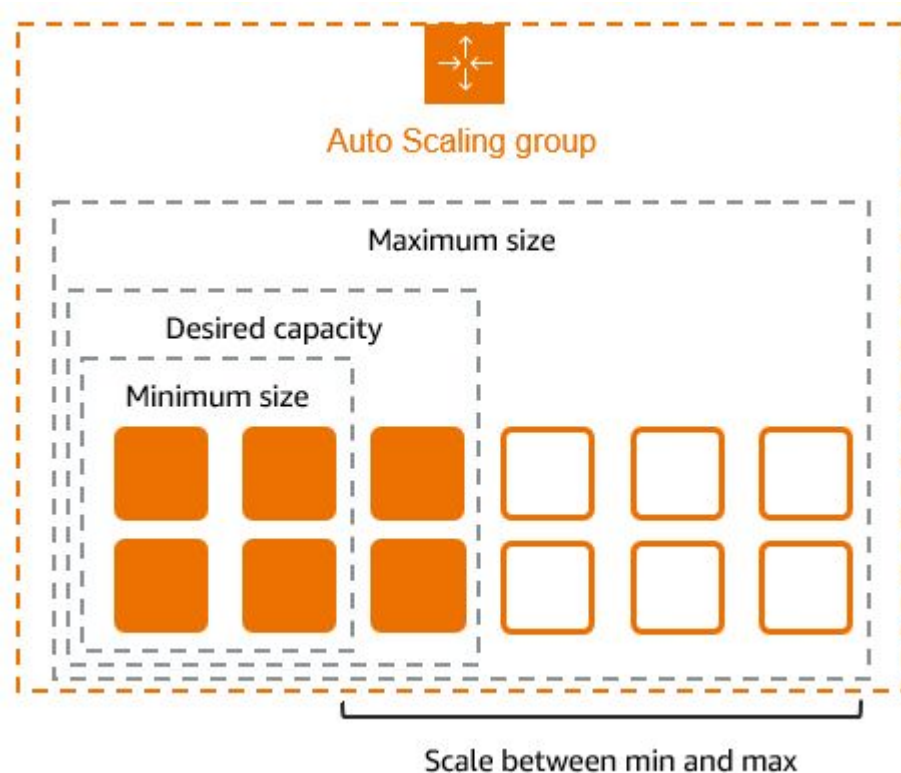
## Elasticidade:

Ajuste automático de recursos, sem intervenção manual, de acordo com a demanda em tempo real.



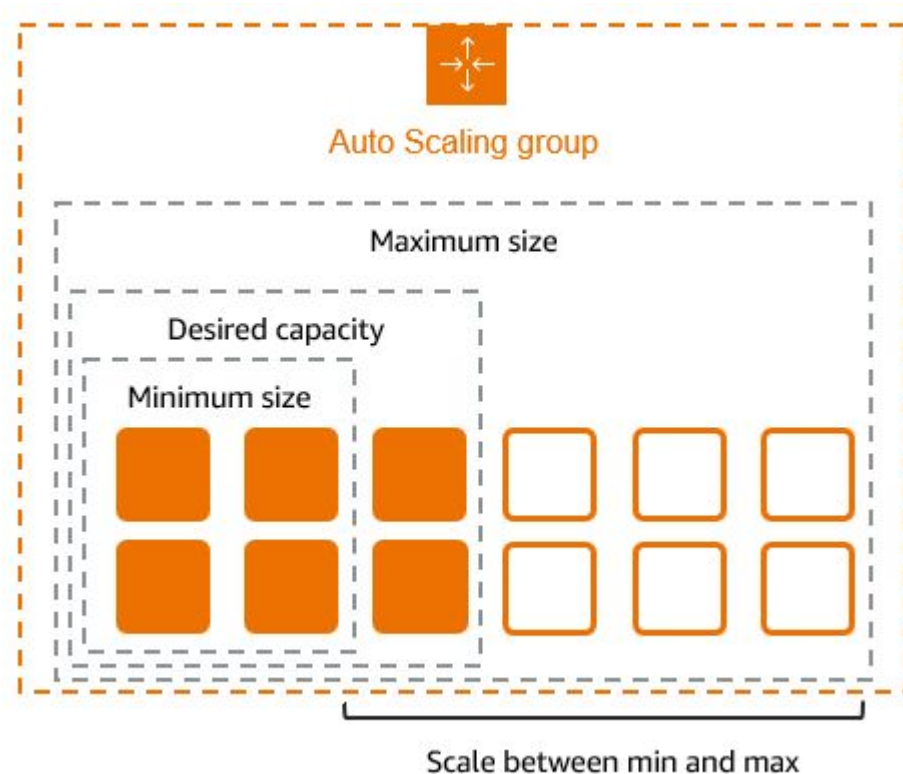
# Auto Scaling

- Define políticas para ajustar a quantidade de instâncias de acordo com a demanda.
- Obtém métricas do CloudWatch
- Funciona junto com ELB para manter o balanceamento e a alta disponibilidade.



# Auto Scaling - métricas para escalar

- CPU
- Rede
- Memória
- Filas Mensagerias (SQS, SNS)
- Qualquer métrica possível de ser mensurada pelo AWS CloudWatch



# Algumas implicações

A aplicação deve permanecer consistente então DB deve ficar separado - fora deste grupo de auto scaling (**RDS** / **DynamoDB** / Instância **EC2** ou On-Premisses para banco de dados )

Arquivos / Objetos também devem ficar separados (**S3**)

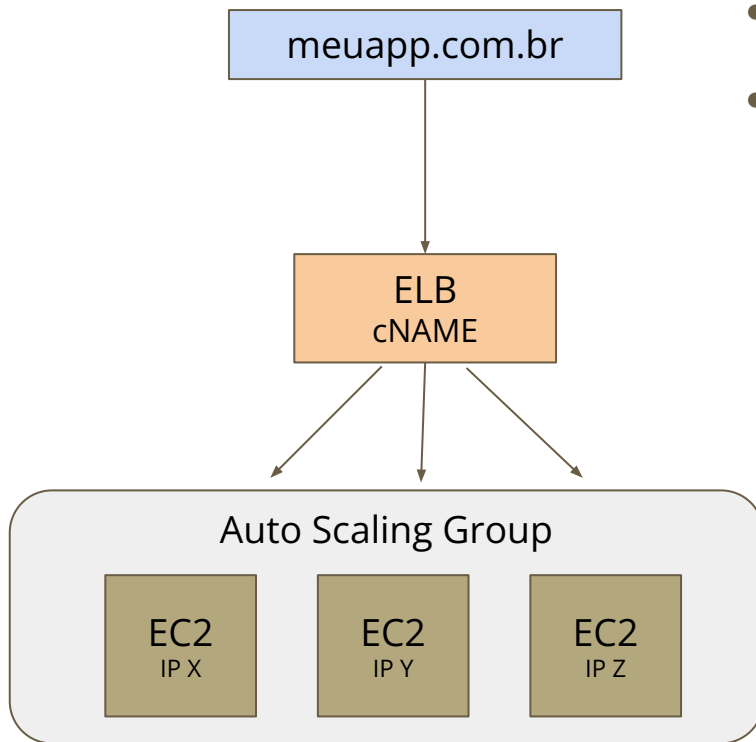
**Como é feito o  
direcionamento dos  
clientes para cada  
instância?**



## ELB - Elastic Load Balancer

- O ELB distribui o tráfego de rede entre várias instâncias.
- Contribui para a escalabilidade horizontal, aliviando sobrecarga em servidores individuais.

# ELB - Elastic Load Balancer - Health Check



- Auto Scaling informa para ELB cada instância que foi criada
- ELB confere se cada instância está OK - **Health Check**