

Auto Scaling Groups e Políticas Dinâmicas

Escalabilidade Horizontal vs Vertical

Horizontal



Adiciona mais instâncias menores
(Ex: de 2 para 4 EC2)

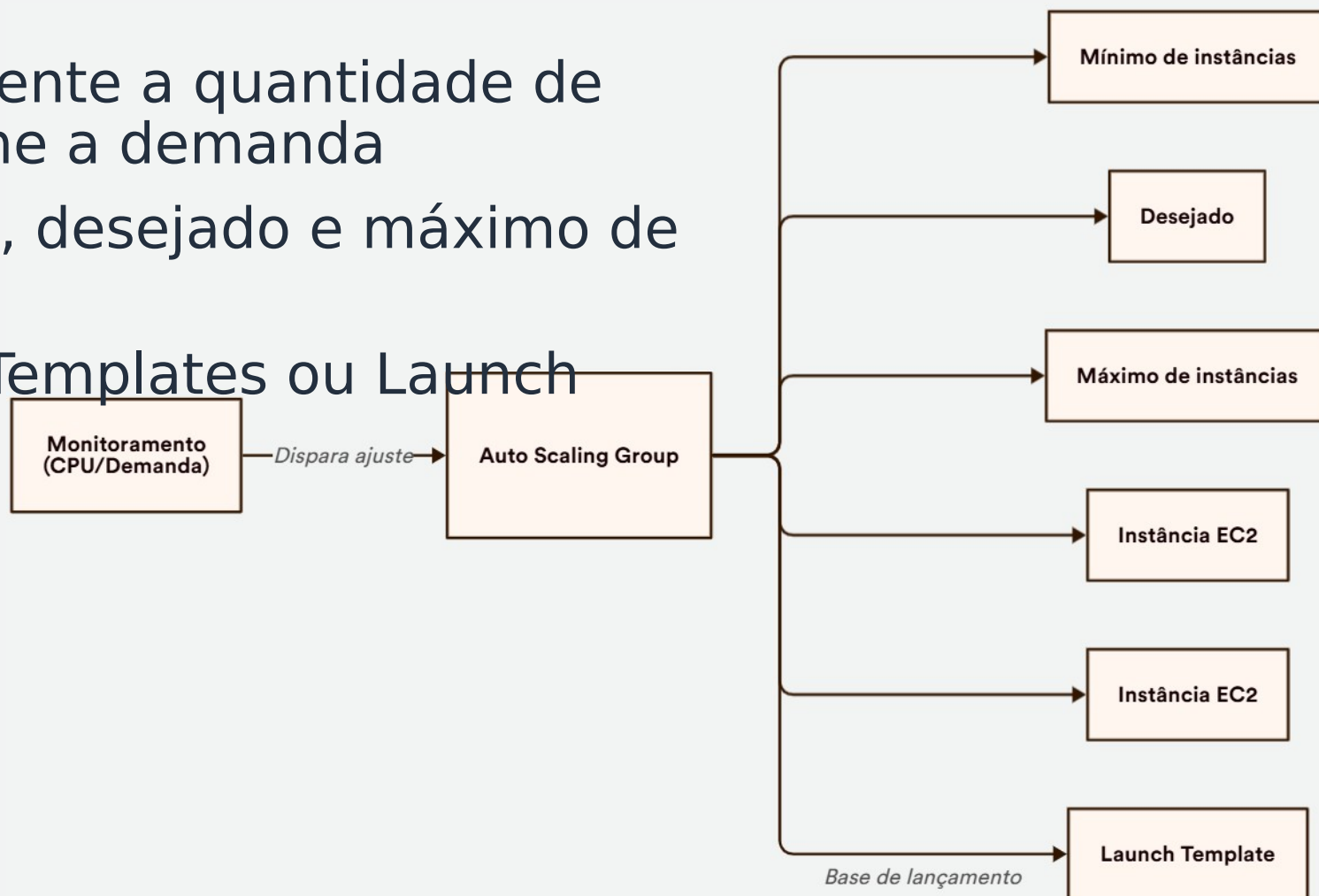
Vertical



Aumenta a capacidade da instância
(Ex: t3.micro → m5.large)

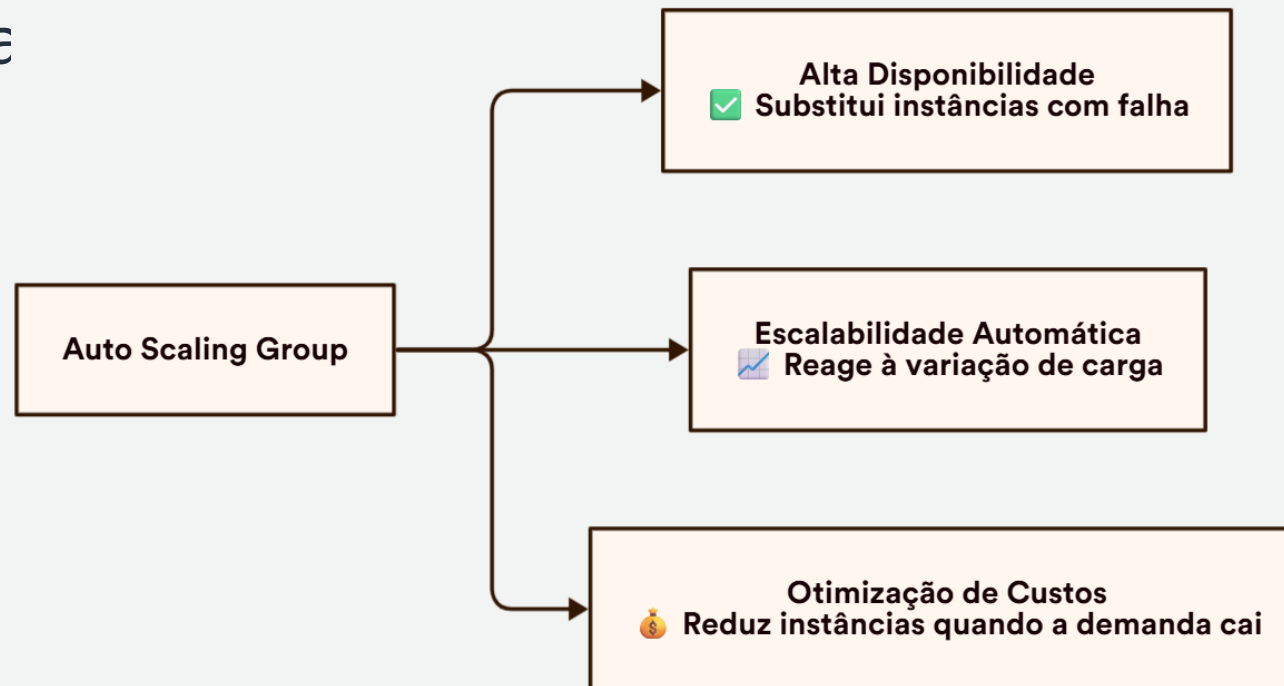
Auto Scaling Groups (ASG)

- Gerencia automaticamente a quantidade de instâncias EC2 conforme a demanda
- Define número mínimo, desejado e máximo de instâncias
- Funciona com Launch Templates ou Launch Configurations



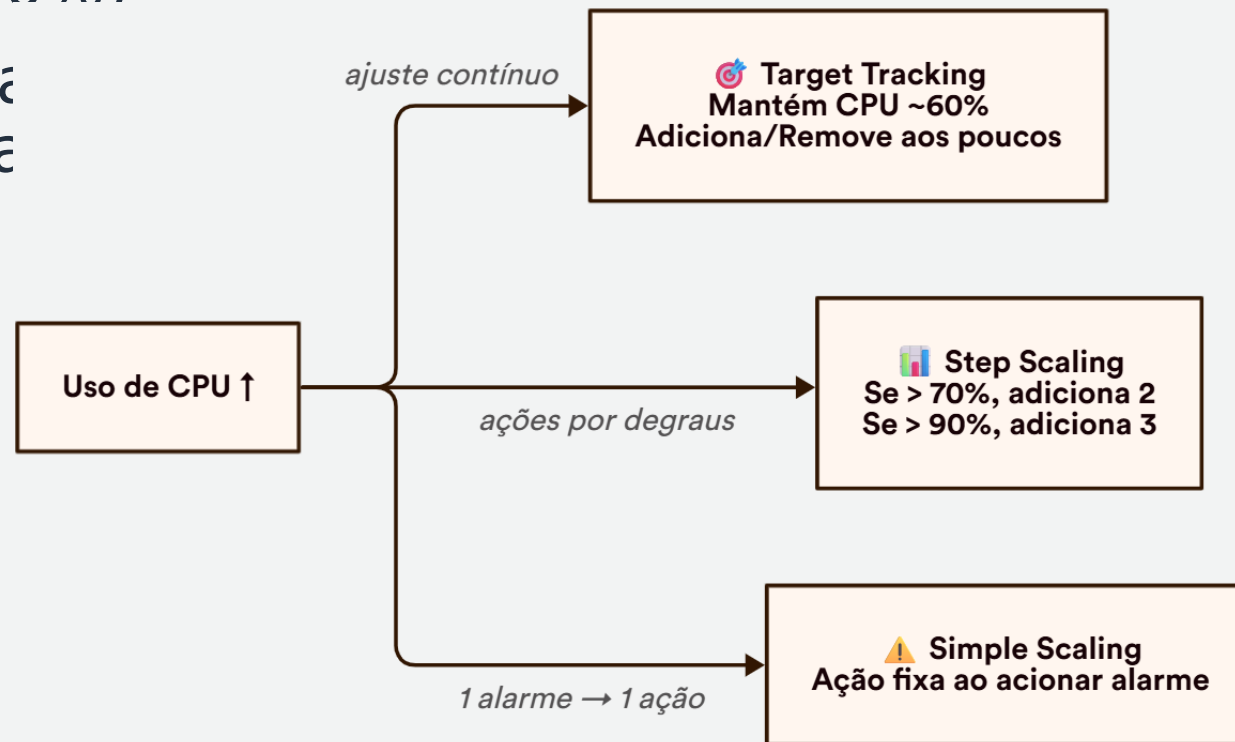
Vantagens do Auto Scaling

- Alta disponibilidade: substitui instâncias com falha automaticamente
- Escalabilidade automática: reage a aumento ou redução de carga
- Otimização de custos: reduz número de instâncias quando a demanda cai



Políticas de Escalonamento Dinâmico

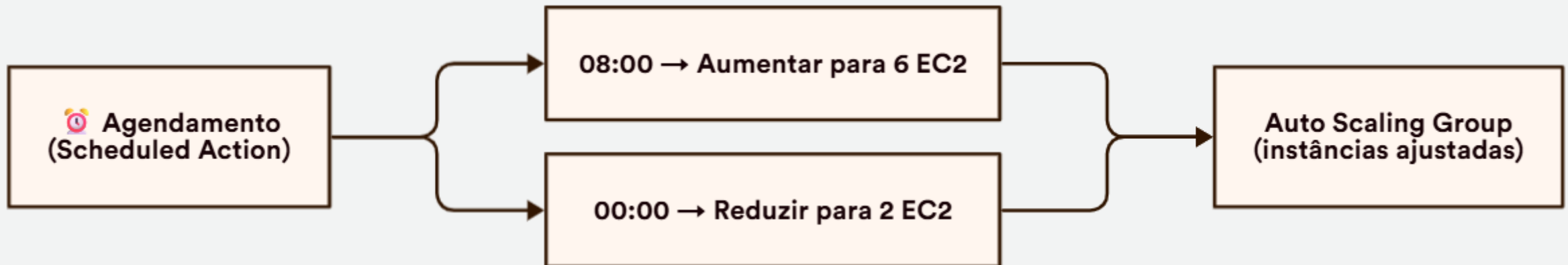
- Target Tracking: ajusta instâncias para manter uma métrica alvo (ex: 60% CPU)
- Step Scaling: define ações específicas por níveis de métrica (ex: +2 se > 70%)
- Simple Scaling: menos usado



;(mais antigo e

Escalonamento Baseado em Horário (Scheduled)

- Permite programar aumentos ou reduções com base em horários pré-definidos
- Ideal para cenários previsíveis como horário comercial vs madrugada



Boas Práticas e Dicas de Prova

- Target Tracking é o mais usado e mais cobrado — fácil de configurar e eficaz
- Step Scaling dá maior controle em ambientes com cargas variadas
- Sempre configure ASG em múltiplas AZs para alta disponibilidade
- Scheduled Scaling é útil em cargas previsíveis — mas não responde a eventos
- A prova cobra diferenças entre políticas, além de casos com health checks e substituições automáticas