

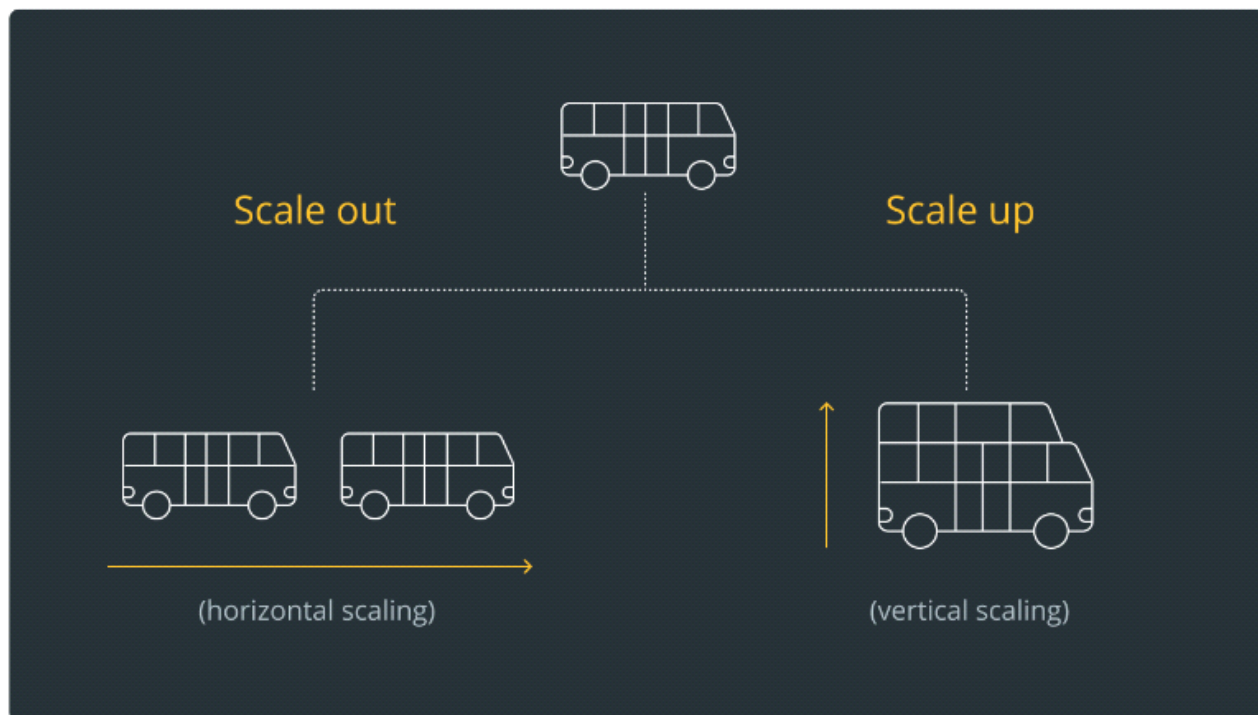
01.Escalabilidade, Elasticidade e Disponibilidade

terça-feira, 5 de dezembro de 2023

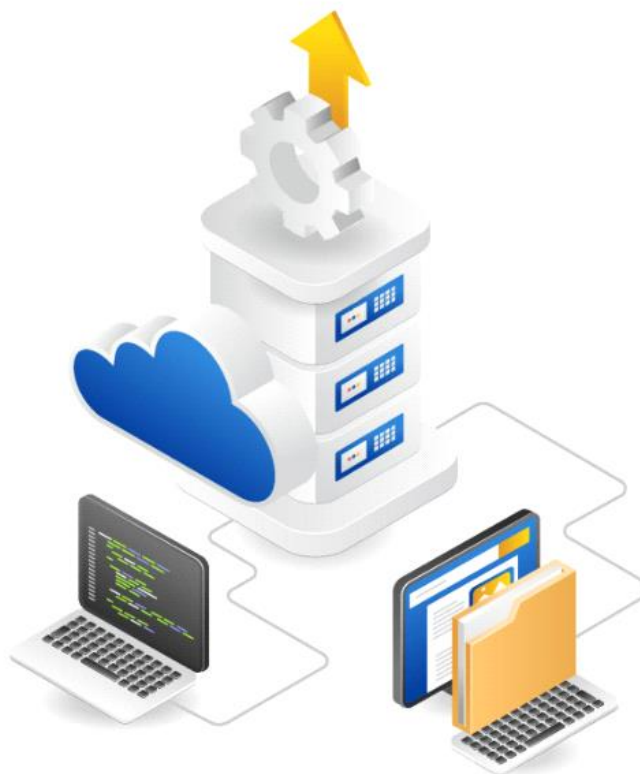
08:26

AutoScaling

- O **Amazon Auto Scaling** é um serviço da AWS que permite o dimensionamento automático de recursos para manter a performance e a disponibilidade de suas aplicações. Ele funciona monitorando continuamente suas aplicações e ajustando a capacidade para manter a performance estável e previsível ao menor custo possível.
- O **Auto Scaling** realiza **verificações de saúde em suas instâncias EC2** e substitui automaticamente as instâncias que não estão saudáveis.
- O Auto Scaling **está integrado com uma série de serviços da AWS**, incluindo Amazon CloudWatch, Amazon SNS, AWS CloudFormation, entre outros.
- O Auto Scaling permite **dimensionar vários recursos, não se limitando apenas às instâncias EC2**. Você também pode dimensionar serviços como Amazon DynamoDB, Amazon Aurora, Amazon ECS, e Amazon RDS.
- **Escalabilidade Vertical** - Aumento da capacidade da instancia
- **Escalabilidade Horizontal (Scaling OUT)** - Aumento no número de instancias (**AutoScaling + Elastic Load Balancing (ELB)**)
- **Scaling IN** - Removendo instancias desnecessárias



- **Elasticidade** - Adaptação automática as mudanças (Load/Carga) - (**AutoScaling + ELB**)
 - Adaptação de acordo com a demanda



- **Disponibilidade** - Disponível em vários locais quando falhar - (AutoScaling)
 - Tolerante a falhas



How autoscaling works

