



# Problema

---

- Como processar grandes volumes de dados?
- Como escalar?



# O que é escalar?

---

Aumentar a capacidade!

Preferencialmente de forma:

- Rápida
- Barata
- Segura



# Como escalar

---

- Aumentar a capacidade do servidor: escalar verticalmente
- Aumentar o número de servidores: escalar horizontalmente



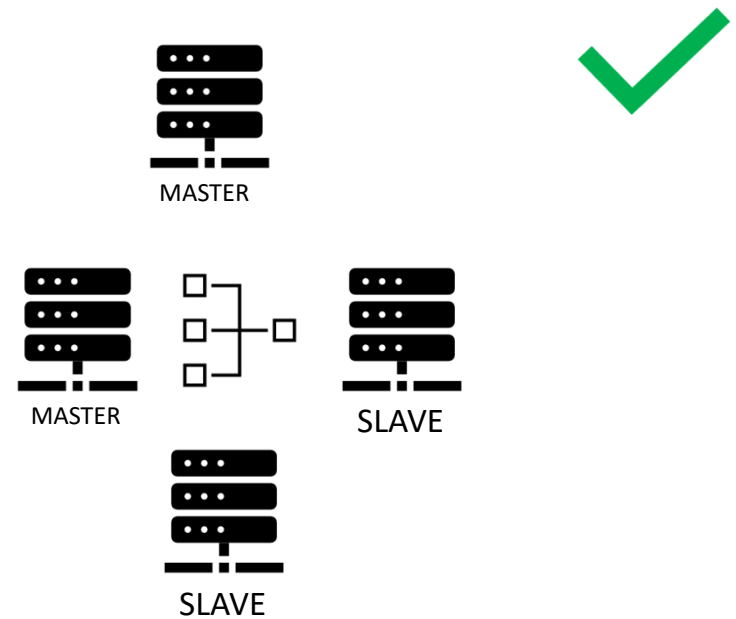
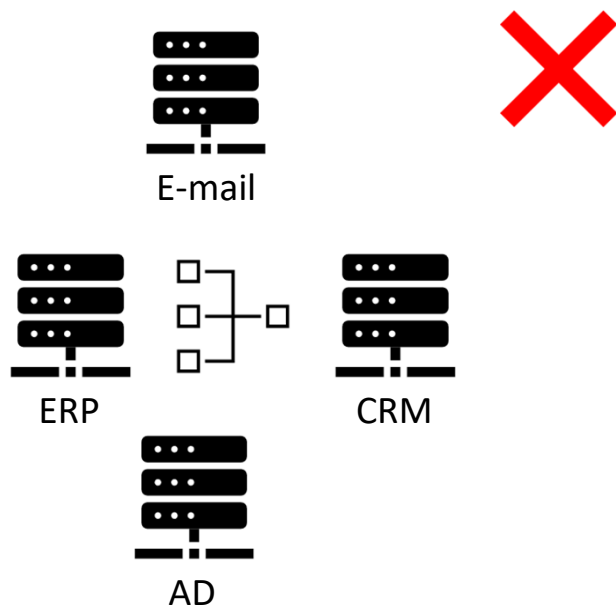
# Commodity Hardware

---

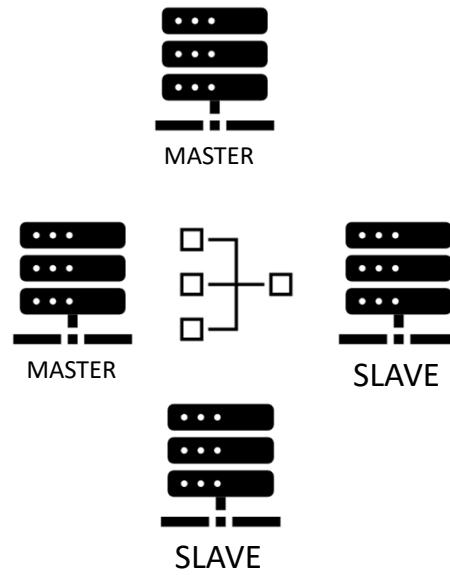
- Commodity Hardware:
  - Barato
  - Compatível
  - Amplamente Disponível
  - Intercambiável

# Cluster

- Conjunto de servidores que operam em conjunto com o mesmo objetivo



# Nós

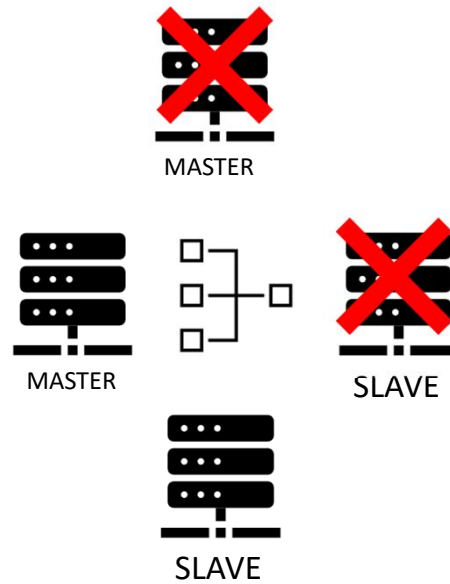




## Requisitos de uma Aplicação Orientada a Dados Moderna:

- Confiável
- Escalável
- Sustentável

# Confiável



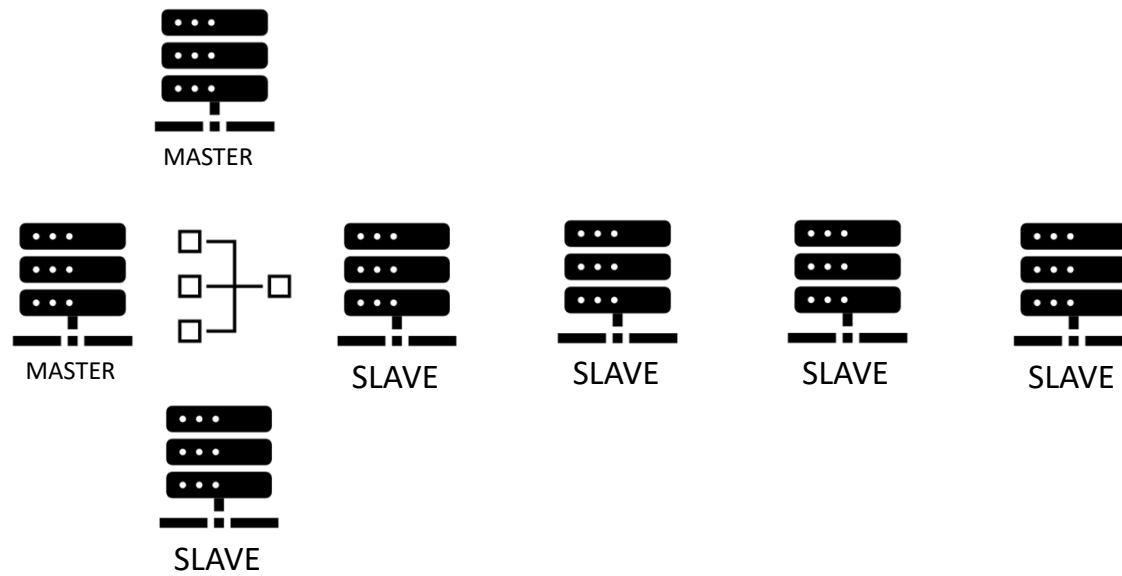




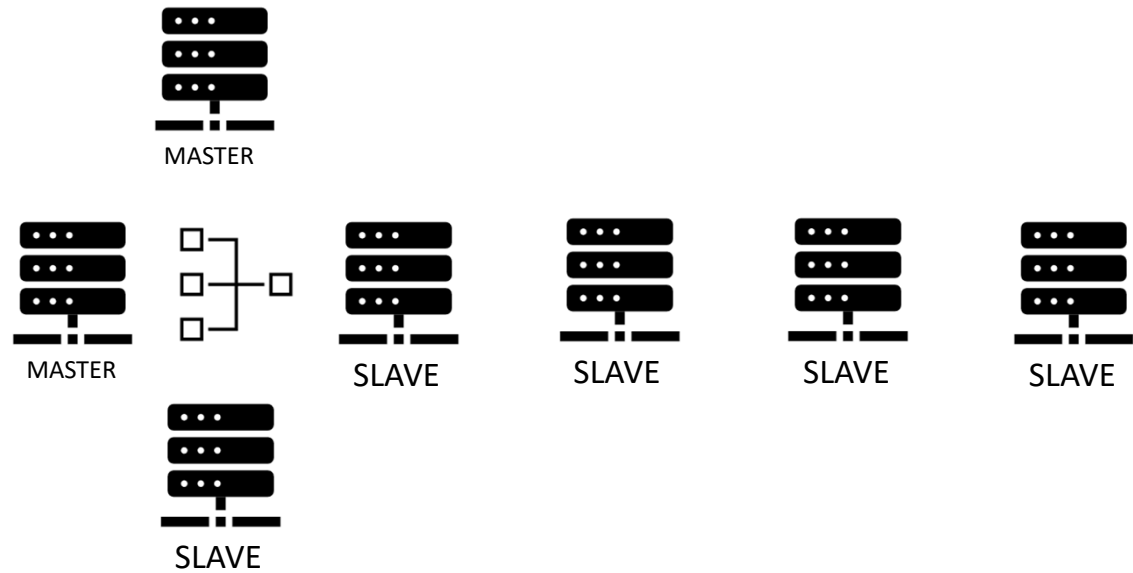
# State Management

- Em caso de falha, o sistema manterá o estado na recuperação
  - Sem manter estado: o processo (ex. ingestão de dados) é reiniciado
  - Com manutenção de estado, ele continua de onde parou

# Escalável



# Sustentável

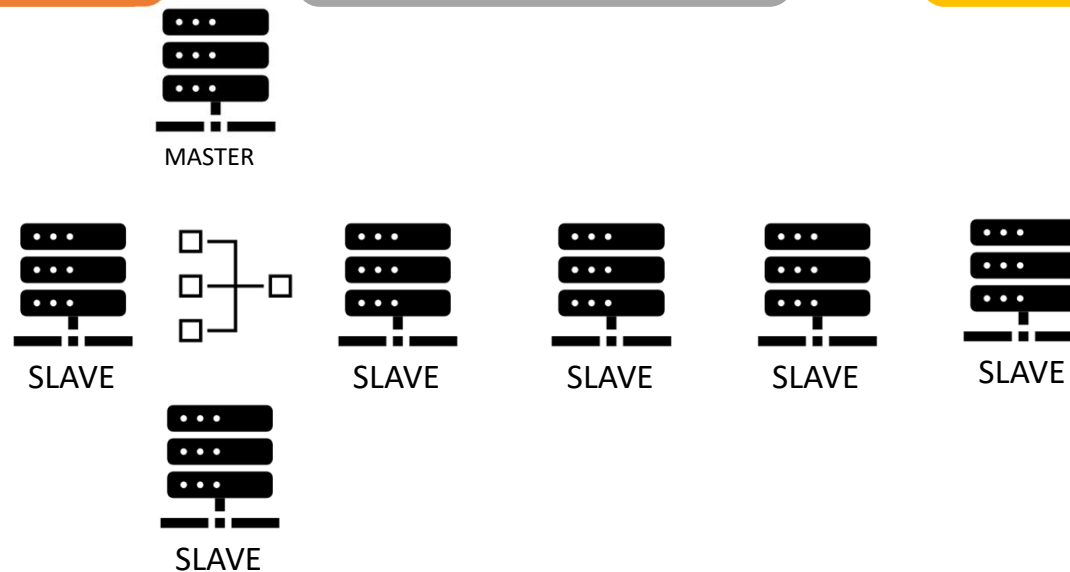


# Mutation

Suporte ao sistema em  
alterar dados

Em sistemas  
distribuídos isso pode  
não ser tão simples

Alguns formatos de  
arquivo, como colunas,  
o processo também  
pode ser complexo



Não se pode ter tudo!

Consistência

Disponibilidade