



UNIVERSIDADE CATÓLICA DE PELOTAS
Mestrado em Engenharia Eletrônica e Computação

Escalabilidade

Disciplina: Sistemas Distribuídos

Pedro Tavares

Prof. Dr. Adenauer Corrêa Yamin

Pelotas, 04 de abril de 2020



Escalabilidade - Conceito

- Um sistema é dito **escalável** se consegue se manter eficiente quando há um aumento significativo no seu número de recursos e usuários;
- Muitas vezes é deixada de lado: sistema criado para resolver problema atual;
- Difícil de definir (longo-prazo);

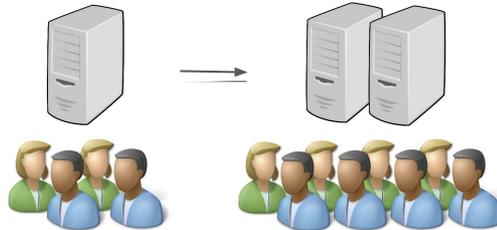


Escalabilidade - Desafios

1. Controlar o custo dos recursos físicos

Ex: Aumentar a quantidade de hardware (ex: servidores) à medida que o número de usuários do sistema aumenta;

Em geral, para que um sistema com n usuários seja escalável, a quantidade de recursos físicos exigida para suportá-los deve ser diretamente proporcional a n . Em outras palavras, se um servidor suporta, no máximo n usuários, ao adicionarmos outro servidor o sistema deve suportar até $2n$ usuários sem perder a eficiência.



Escalabilidade - Desafios

2. Controlar a perda de desempenho

Ex: Um sistema distribuído que faz constantes consultas à uma base de dados deve manter o tempo de resposta de uma consulta, mesmo se o número de dados aumentar consideravelmente.

Normalmente o tempo que leva para acessar dados hierarquicamente estruturados é $O(\log n)$, onde n é o tamanho do conjunto de dados. Para que um sistema seja **escalável**, a perda de desempenho máxima **não** deve ser maior do que isso.

Escalabilidade - Desafios

3. Impedir que os recursos de software se esgotem

Ex: Os endereços IP (criados nos anos 70) possuíam apenas 32 bits (IPv4), pois se subestimou a capacidade de crescimento da Internet. Mais tarde, foi necessária a criação de um novo protocolo IPv6 com 128 bits para suprir a necessidade de demanda.

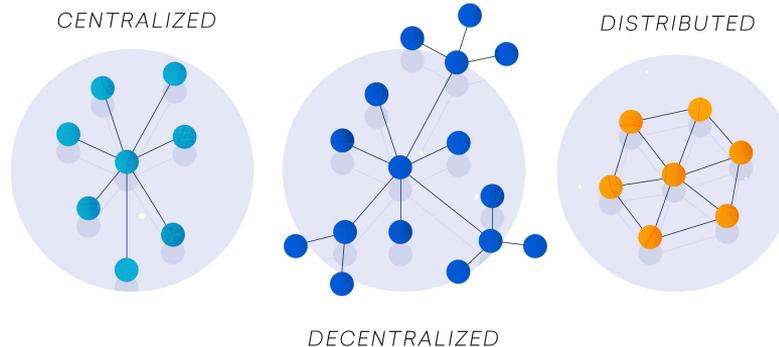
Por outro lado, superestimar também é ruim: dado extra não utilizado (mais armazenamento, aumenta overhead, etc.)

Escalabilidade - Desafios

4. Evitar gargalos de desempenho

Em geral, os algoritmos devem ser descentralizados para evitar a existência de gargalos de desempenho.

Ex: Ao invés de manter todos os dados (ou todo processamento) em um servidor central, distribuir em vários servidores para evitar sobrecarga do sistema.



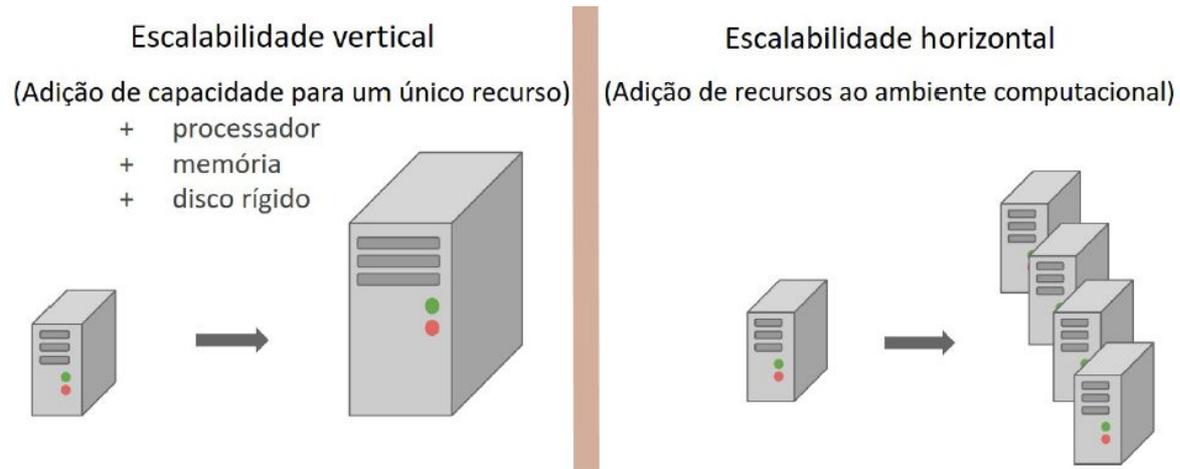
Escalabilidade - Tipos



- **Escalabilidade de Carga:** É a facilidade com que um sistema distribuído pode ser expandido usando todos os seus recursos para acomodar as demandas;
- **Escalabilidade Geográfica:** Um sistema é geograficamente escalável quando mantém sua utilidade e desempenho, independentemente dos pontos geográficos que compõem sua extensão;
- **Escalabilidade Administrativa:** Não importa a variação de informação que diferentes organizações necessitem compartilhar em um único sistema distribuído, este deve permanecer fácil de ser usado e gerenciado;

Métodos de adição de recursos

- **Escalar verticalmente (scale up)** significa adicionar recursos em um único nó do sistema (mais memória ou um disco rígido mais rápido).
- **Escalar horizontalmente (scale out)** significa adicionar mais nós ao sistema, tais como um novo computador com uma aplicação para clusterizar o software.



Escalabilidade - Técnicas

Algumas técnicas utilizadas para melhorar a escalabilidade de um sistema distribuído são:

- Replicação de dados;
- Uso de cache;
- Distribuição de vários servidores para a execução de tarefas recorrentes;

Referências

- [1] COULOURIS, George et al. - Sistemas Distribuídos: Conceitos e Projeto, 5ª ed. Bookman Editora, 2013.
- [2] Site da Wikipedia: <https://pt.wikipedia.org/wiki/Escalabilidade>

