

ACH2147 - Desenvolvimento de Sistemas de Informação Distribuídos

Aula 03 – Objetivos dos SDs (cont.)

Norton Trevisan Roman

17 de março de 2022

Objetivos dos sistemas distribuídos:

- Disponibilização de recursos compartilhados
- Transparência de distribuição
- Abertura
- **Escalabilidade**

Escalabilidade em SDs

Muitos desenvolvedores de sistemas distribuídos modernos usam o adjetivo “escalável” sem deixar claro o **porquê** deles escalarem.

Escalabilidade em SDs

Muitos desenvolvedores de sistemas distribuídos modernos usam o adjetivo “escalável” sem deixar claro o **porquê** deles escalarem.

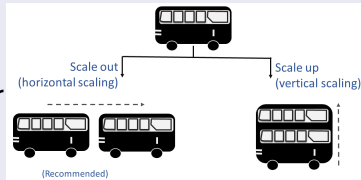
Escalabilidade possui pelo menos 3 componentes:

- Número de usuários e/ou processos – **escalabilidade de tamanho**
- Distância máxima entre nós – **escalabilidade geográfica**
- Número de domínios administrativos – **escalabilidade administrativa**

Escalabilidade: Componentes

Escalabilidade de tamanho

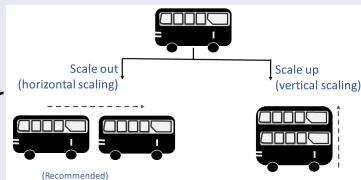
- Devemos poder adicionar usuários e recursos sem notar perda de desempenho



Escalabilidade: Componentes

Escalabilidade de tamanho

- Devemos poder adicionar usuários e recursos sem notar perda de desempenho



Escalabilidade geográfica

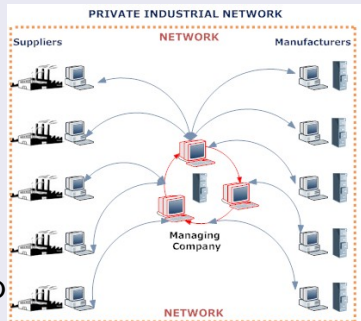
- Não podemos notar atrasos de comunicação, mesmo que usuários e recursos estejam longes uns dos outros



Escalabilidade: Componentes

Escalabilidade administrativa

- O sistema deve ser gerenciável mesmo que se espalhe por diferentes organizações
- Permite que um programa rodando em uma organização acesse recursos em outra
 - Potenciais problemas de segurança



Escalabilidade em SDs

Problema

- A maior parte dos sistemas escalam apenas (e até certo ponto) em tamanho.

Escalabilidade em SDs

Problema

- A maior parte dos sistemas escalam apenas (e até certo ponto) em tamanho.
- Solução(?!): Múltiplos servidores potentes operando independentemente em paralelo

Escalabilidade em SDs

Problema

- A maior parte dos sistemas escalam apenas (e até certo ponto) em tamanho.
- Solução(?!): Múltiplos servidores potentes operando independentemente em paralelo

Hoje em dia, o desafio ainda reside em conseguir escalabilidade geográfica e administrativa.

Dificuldades para obtenção de escalabilidade

- Características de um sistema completamente descentralizado:
 - Nenhuma máquina tem informação completa sobre o estado do sistema
 - Máquinas tomam decisões baseadas apenas em informação local
 - Falhas em uma máquina não devem arruinar a execução do algoritmo
 - Não é possível assumir a existência de um relógio global

Técnicas de escalabilidade

Esconda latências de comunicação

- Utilize **comunicação assíncrona**
- Mantenha diferentes *handlers* para tratamento de mensagens recebidas
- **Problema:** nem toda aplicação se encaixa nesse modelo

Técnicas de escalabilidade

Esconda latências de comunicação

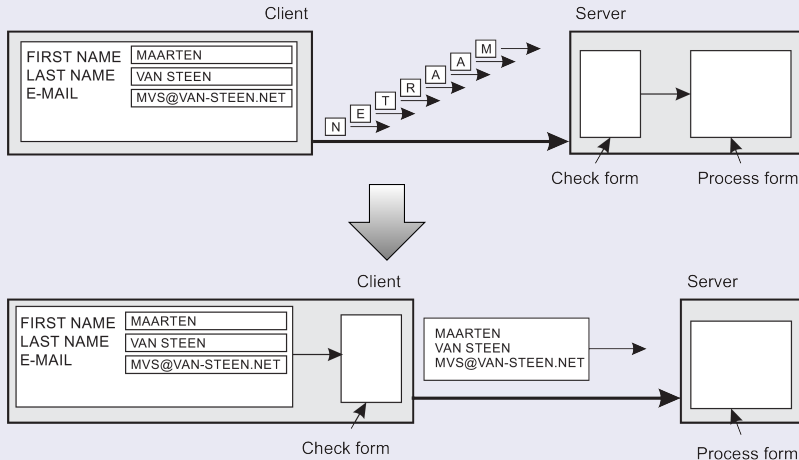
- Utilize **comunicação assíncrona**
- Mantenha diferentes *handlers* para tratamento de mensagens recebidas
- **Problema:** nem toda aplicação se encaixa nesse modelo

Ideia geral

Não espere por respostas – faça outra coisa

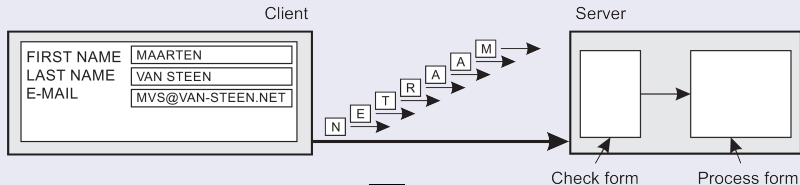
Técnicas de escalabilidade

Mova a computação para o lado do cliente

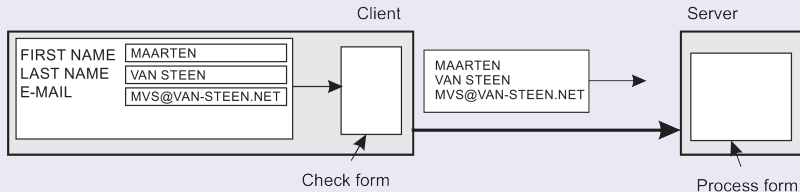


Técnicas de escalabilidade

Mova a computação para o lado do cliente



Se latência é um problema,
evite múltiplas requisições



Técnicas de escalabilidade

Distribua os dados e a computação

- Mova a computação para os clientes (ex: Javascript, Applets Java, etc.)
- Use serviços de nomes descentralizados (DNS)
- Use sistemas de informação descentralizados (WWW)

Replicação/caching

Disponibilize cópias dos dados em diferentes máquinas:

- Bancos de dados e sistemas de arquivos replicados
- Sites web “espelhados”
- Caches web (nos navegadores e nos *proxies*)
- Cache de arquivos (no servidor e nos clientes)

Técnicas de escalabilidade: replicação

Aplicar técnicas de replicação é fácil

Técnicas de escalabilidade: replicação

Aplicar técnicas de replicação é fácil

Exceto por alguns detalhes...

- Manter múltiplas cópias (em cache ou replicadas) leva a **inconsistências**: a modificação em uma cópia a torna diferente das demais

Técnicas de escalabilidade: replicação

Aplicar técnicas de replicação é fácil

Exceto por alguns detalhes...

- Manter múltiplas cópias (em cache ou replicadas) leva a **inconsistências**: a modificação em uma cópia a torna diferente das demais
- Manter as cópias consistentes requer **sincronização global** em cada modificação

Técnicas de escalabilidade: replicação

Aplicar técnicas de replicação é fácil

Exceto por alguns detalhes...

- Manter múltiplas cópias (em cache ou replicadas) leva a **inconsistências**: a modificação em uma cópia a torna diferente das demais
- Manter as cópias consistentes requer **sincronização global** em cada modificação
- Sincronização global impossibilita soluções em larga escala

Técnicas de escalabilidade: replicação

Contudo...

- Se pudermos tolerar inconsistências, podemos reduzir a necessidade de sincronização global
- No entanto, **tolerar inconsistências é algo que depende da aplicação.**

Armadilhas

- Muitos sistemas distribuídos se tornam desnecessariamente complexos por causa de “consertos” ao longo do tempo.
- Em geral, são feitas muitas **suposições falsas**
 - Frequentemente implícitas, escondidas dos usuários e programadores

Suposições falsas

- A rede é confiável

Suposições falsas

- A rede é confiável
- A rede é segura

Suposições falsas

- A rede é confiável
- A rede é segura
- A rede é homogênea

Suposições falsas

- A rede é confiável
- A rede é segura
- A rede é homogênea
- A topologia da rede não muda

Suposições falsas

- A rede é confiável
- A rede é segura
- A rede é homogênea
- A topologia da rede não muda
- A latência é zero

Suposições falsas

- A rede é confiável
- A rede é segura
- A rede é homogênea
- A topologia da rede não muda
- A latência é zero
- Largura de banda é infinita

Suposições falsas

- A rede é confiável
- A rede é segura
- A rede é homogênea
- A topologia da rede não muda
- A latência é zero
- Largura de banda é infinita
- O custo de transporte é zero

Suposições falsas

- A rede é confiável
- A rede é segura
- A rede é homogênea
- A topologia da rede não muda
- A latência é zero
- Largura de banda é infinita
- O custo de transporte é zero
- A rede possui um administrador