

# Sistemas Distribuídos

Luís Rodrigues

# O que é um sistema distribuído?

- Conjunto de processos que se coordenada para executar uma tarefa comum
  - Partilha de informação
  - Sincronização

# Como?

- Troca de mensagens
- Partilha de memória

# Troca de mensagens

- Chamada a procedimentos remotos
- Difusão em grupo
- Edição-subscrição de eventos

# Partilha de memória

- Memória partilhada distribuída
- Espaços de tuplos distribuídos

# Quais os desafios?

- Concorrência
- Latência
- Coerência
- Tolerância a faltas
- Capacidade de escala
- Segurança

# Concorrência

- Más notícias:
  - Os sistemas distribuídos são concorrentes

# Concorrência

- Más notícias:
  - Os sistemas distribuídos são concorrentes
- Boas notícias
  - Os sistemas distribuídos são concorrentes



# Concorrência

- Más notícias:
  - Os sistemas distribuídos são concorrentes
- Boas notícias
  - Os sistemas distribuídos são concorrentes
  - **São precisos bons profissionais para fazer bons sistemas distribuídos!**

# Latência

- Consta que existe uma constante chamada velocidade da luz
- Isto significa que:
  - Por muito potente que seja a máquina
  - Por muita largura de banda que tenha a rede
  - Um processo em Portugal a coordenar-se com um processo no Japão será sempre uma tarefa lenta.

# Coerência

- Quando um processo faz uma alteração ao estado do sistema...
- ...esta alteração não fica instaneamente visível nos restantes processos
- Isto pode ser um problema:
  - Que comportamentos podemos considerar correctos?
  - Podemos ter diversas definições de coerência

# Tolerância a faltas

- Faltas nos processos
  - Paragem
  - Bizantinas
- Faltas na rede
  - Perda de mensagens
  - Reordenação de mensagens
  - Partições na rede

# Detecção de falhas

- Pode ser impossível ter um detector de falhas perfeito!
- Um processo pode ter falhado, pode estar simplesmente lento, ou com problemas na rede.

# Capacidade de escala

- Conceber sistemas que continuam a funcionar à medida que acrescentamos mais processos ao sistema

# Segurança

- Não é um problema específico dos sistemas distribuídos
- Mas é amplificado pelos sistemas distribuídos
  - Mais processos = maior exposição

# Dois exemplos

- Semântica das chamadas a procedimentos remotos
- DNS



# Semântica do RPC

- Idealmente:
  - Exatamente uma vez!
- O que é preciso para conseguir isto?
  - Falhas na rede
  - Falhas nos processos

# Manter estado para detectar duplicados!

- Por quanto tempo se tem de manter o estado?
- Quais os problemas que isto levanta?

# Alternativas

- Pelo menos uma vez
- Melhor esforço (aka, “se deus quiser”)

# Estratégia para esquivar o problema

- Programar o serviço de forma que o mesmo pedido possa ser executado múltiplas vezes sem que o resultado se altere
  - Idempotente
- **Levantar dinheiro não é idempotente**

# DNS

- Serviços de directório
  - Procurar impressoras a cores que façam frente e verso e que tirem cafés
- Serviços de nomes
  - Dado um nome saber um endereço

# DNS

```
$nslookup www.gsd.inesc-id.pt  
Server:      212.113.177.241  
Address:     212.113.177.241#53
```

```
Non-authoritative answer:  
Name: www.gsd.inesc-id.pt  
Address: 146.193.41.139
```

# DNS

```
$nslookup mail.gsd.inesc-id.pt
```

```
Server:      212.113.177.241
```

```
Address:     212.113.177.241#53
```

```
Non-authoritative answer:
```

```
mail.gsd.inesc-id.pt    canonical name = gsd.inesc-id.pt.
```

```
Name: gsd.inesc-id.pt
```

```
Address: 146.193.41.130
```

# DNS

```
$nslookup www.publico.pt
```

```
Server:          212.113.177.241
```

```
Address:         212.113.177.241#53
```

```
Non-authoritative answer:
```

```
www.publico.ptcanonical name = d2myb36x6mulj5.cloudfront.net.
```

```
Name:   d2myb36x6mulj5.cloudfront.net
```

```
Address: 52.85.187.2
```

```
Name:   d2myb36x6mulj5.cloudfront.net
```

```
Address: 52.85.187.9
```

```
Name:   d2myb36x6mulj5.cloudfront.net
```

```
Address: 52.85.187.75
```

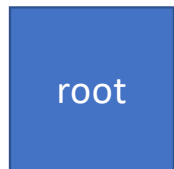
```
Name:   d2myb36x6mulj5.cloudfront.net
```

```
Address: 52.85.187.122
```



# DNS

- Um servidor central



# DNS

- Um servidor central replicado



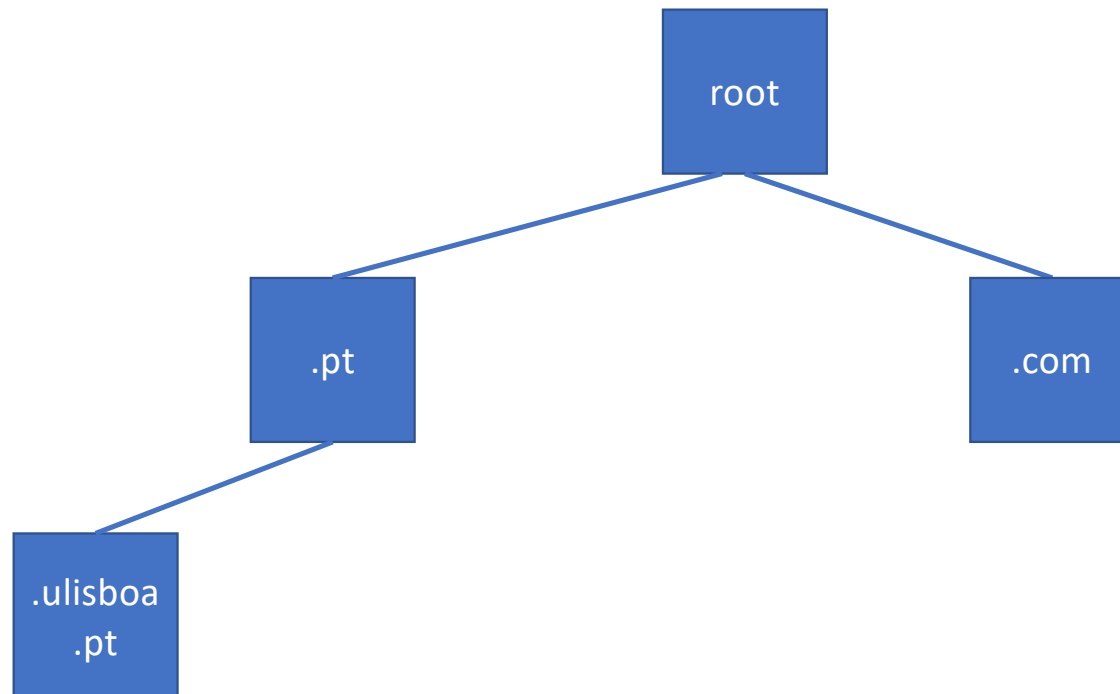
# DNS

- Um servidor central replicado e geo-replicado



# DNS

- Uma hierarquia de servidores cooperativos



# DNS

- Resolução de um nome
  - Através de um componente designado por “resolver”
- Duas estratégias possíveis
  - Resolução iterativa
  - Resolução recursiva

# DNS

- Cache
  - Em cada organização existe um servidor que faz a resolução e guarda os resultados
  - Cache pode gerar problema de coerência!

# Algumas notas

- DNS
  - Os requisitos de coerência do DNS são relativamente fracos
- O DSN usa nomes hierárquicos

# Nomes

- Como dar nomes a entidades é um problema interessante por si só
  - Que não abordaremos em pormenor nas aulas teoricas
- /Users/ler/...
- /Volumes/Google Drive/...
- C:\DOSTEMP
- FA19 B295 C86D 993C 8432 DEFD 11AA 88E6



# Nomes puros

- Como encontrar o endereço de uma entidade a partir de um nome puro?

FA19 B295 C86D 993C 8432 DEFD 11AA 88E6