

Replicação Não Uniforme com Consistência Eventual

Gonçalo Cabrita e Nuno Preguiça

NOVA LINCS

9º INForum, Aveiro, Portugal,

13 de Outubro de 2017

- O aumento de utilizadores tem forçado serviços a encontrarem novas formas de escalar
- Muitos serviços guardam os seus dados em key-value stores geo-replicadas
- Estas base de dados sacrificam a consistência dos dados em troca de alta-disponibilidade

- O aumento de utilizadores tem forçado serviços a encontrarem novas formas de escalar
- Muitos serviços guardam os seus dados em key-value stores geo-replicadas
- Estas base de dados sacrificam a consistência dos dados em troca de alta-disponibilidade

- O aumento de utilizadores tem forçado serviços a encontrarem novas formas de escalar
- Muitos serviços guardam os seus dados em key-value stores geo-replicadas
- Estas base de dados sacrificam a consistência dos dados em troca de alta-disponibilidade

- A informação guardada nestas base de dados aumenta rapidamente
- É tipicamente impossível manter todos os dados em todas as réplicas
- Alguns sistemas adoptam um modelo de replicação parcial
 - Pode ser necessário contactar várias réplicas para responder a uma query

- A informação guardada nestas base de dados aumenta rapidamente
- É tipicamente impossível manter todos os dados em todas as réplicas
- Alguns sistemas adoptam um modelo de replicação parcial
 - Pode ser necessário contactar várias réplicas para responder a uma query

- A informação guardada nestas base de dados aumenta rapidamente
- É tipicamente impossível manter todos os dados em todas as réplicas
- Alguns sistemas adoptam um modelo de replicação parcial
 - Pode ser necessário contactar várias réplicas para responder a uma query

- A informação guardada nestas base de dados aumenta rapidamente
- É tipicamente impossível manter todos os dados em todas as réplicas
- Alguns sistemas adoptam um modelo de replicação parcial
 - Pode ser necessário contactar várias réplicas para responder a uma query

Conseguimos criar um modelo de replicação onde qualquer réplica consiga responder a todas as operações de leitura sem guardar todos os dados?

Exemplo: máximo

$$\left\{ \begin{array}{c} 9 \\ \\ \end{array} \right\} \quad \left\{ \begin{array}{c} 3 \\ \\ \end{array} \right\} \quad \left\{ \begin{array}{c} 2 \\ \\ \end{array} \right\}$$

Exemplo: máximo

$$\left\{ \begin{array}{c} \mathbf{9} \\ 3 \\ 2 \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{9} \\ 3 \\ 2 \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{9} \\ 3 \\ 2 \end{array} \right\}$$

Exemplo: máximo

$$\left\{ \begin{array}{c} \mathbf{9} \\ 5 \\ 3 \\ 2 \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{9} \\ 3 \\ 2 \\ \mathbf{1} \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{9} \\ 3 \\ 2 \end{array} \right\}$$

Exemplo: máximo

$$\left\{ \begin{array}{c} \cancel{9} \\ 5 \\ 3 \\ 2 \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{9} \\ 3 \\ 2 \\ 1 \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{9} \\ 3 \\ 2 \end{array} \right\}$$

Exemplo: máximo

$$\left\{ \begin{array}{c} \cancel{9} \\ 5 \\ 3 \\ 2 \end{array} \right\} \quad \left\{ \begin{array}{c} \cancel{9} \\ 3 \\ 2 \\ 1 \end{array} \right\} \quad \left\{ \begin{array}{c} \cancel{9} \\ 3 \\ 2 \end{array} \right\}$$

Exemplo: máximo

$$\left\{ \begin{array}{c} \mathbf{5} \\ 3 \\ 2 \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{5} \\ 3 \\ 2 \\ 1 \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{5} \\ 3 \\ 2 \end{array} \right\}$$

- **Replicação Não Uniforme**
- CRDTs Não Uniformes
- Avaliação
- Conclusão e trabalho futuro

- Um modelo de replicação onde todas as réplicas podem responder a todas as queries, mantendo apenas um subconjunto dos dados
- Réplicas do mesmo objeto não precisam de ter estados **equivalentes**, apenas necessitam de ter estados **observavelmente equivalentes**
- Para dois estados serem **observavelmente equivalentes** uma operação de leitura tem de devolver o mesmo resultado para ambos os estados

- Um modelo de replicação onde todas as réplicas podem responder a todas as queries, mantendo apenas um subconjunto dos dados
- Réplicas do mesmo objeto não precisam de ter estados **equivalentes**, apenas necessitam de ter estados **observavelmente equivalentes**
- Para dois estados serem **observavelmente equivalentes** uma operação de leitura tem de devolver o mesmo resultado para ambos os estados

- Um modelo de replicação onde todas as réplicas podem responder a todas as queries, mantendo apenas um subconjunto dos dados
- Réplicas do mesmo objeto não precisam de ter estados **equivalentes**, apenas necessitam de ter estados **observavelmente equivalentes**
- Para dois estados serem **observavelmente equivalentes** uma operação de leitura tem de devolver o mesmo resultado para ambos os estados

$$\left\{ \begin{array}{c} \mathbf{9} \\ 5 \\ 3 \\ 2 \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{9} \\ 3 \\ 2 \\ 1 \end{array} \right\} \quad \left\{ \begin{array}{c} \mathbf{9} \\ 3 \\ 2 \end{array} \right\}$$

Um sistema replicado fornece **consistência eventual** num estado quiescente sse:

1. Cada réplica tiver executado **todas** as operações
2. O estado de quaisquer duas réplicas seja **equivalente**

Um sistema replicado fornece **consistência eventual** num estado quiescente sse:

1. Cada réplica tiver executado **todas** as operações
2. O estado de quaisquer duas réplicas seja **equivalente**

Um sistema replicado fornece **consistência eventual** num estado quiescente sse:

1. Cada réplica tiver executado **todas** as operações
2. O estado de quaisquer duas réplicas seja **equivalente**

Consistência Eventual Não Uniforme (NuEC)

Um sistema replicado fornece **consistência eventual não uniforme** num estado quiescente sse:

1. Cada réplica tiver executado um conjunto de operações que tenham impacto no **estado observável** final
2. O estado de quaisquer duas réplicas seja **observavelmente equivalente**

Consistência Eventual Não Uniforme (NuEC)

Um sistema replicado fornece **consistência eventual não uniforme** num estado quiescente sse:

1. Cada réplica tiver executado um conjunto de operações que tenham impacto no **estado observável** final
2. O estado de quaisquer duas réplicas seja **observavelmente equivalente**

Consistência Eventual Não Uniforme (NuEC)

Um sistema replicado fornece **consistência eventual não uniforme** num estado quiescente sse:

1. Cada réplica tiver executado um conjunto de operações que tenham impacto no **estado observável** final
2. O estado de quaisquer duas réplicas seja **observavelmente equivalente**

O objetivo principal é dividir as operações em quatro grupos:

1. Operações que são **core**
2. Operações que são **forever masked**
3. Operações que são **masked** mas podem tornar-se **core**
4. Operações que são **masked** mas tendo em conta todas as operações do sistema seriam consideradas **core**

O objetivo principal é dividir as operações em quatro grupos:

1. Operações que são **core**
2. Operações que são **forever masked**
3. Operações que são **masked** mas podem tornar-se **core**
4. Operações que são **masked** mas tendo em conta todas as operações do sistema seriam consideradas **core**

O objetivo principal é dividir as operações em quatro grupos:

1. Operações que são **core**
2. Operações que são **forever masked**
3. Operações que são **masked** mas podem tornar-se **core**
4. Operações que são **masked** mas tendo em conta todas as operações do sistema seriam consideradas **core**

O objetivo principal é dividir as operações em quatro grupos:

1. Operações que são **core**
2. Operações que são **forever masked**
3. Operações que são **masked** mas podem tornar-se **core**
4. Operações que são **masked** mas tendo em conta todas as operações do sistema seriam consideradas **core**

O objetivo principal é dividir as operações em quatro grupos:

1. Operações que são **core**
2. Operações que são **forever masked**
3. Operações que são **masked** mas podem tornar-se **core**
4. Operações que são **masked** mas tendo em conta todas as operações do sistema seriam consideradas **core**

- Replicação Não Uniforme
- **CRDTs Não Uniformes**
- Avaliação
- Conclusão e trabalho futuro

- Conjunto de tuplos (identificador, pontuação)
- Duas operações de escrita
 - `ADD(id, pontuação)`: associa uma nova pontuação com `id`
 - `RMV(id)`: remove todas as pontuações associadas com `id`

- Conjunto de tuplos (identificador, pontuação)
- Duas operações de escrita
 - `ADD(id, pontuação)`: associa uma nova pontuação com `id`
 - `RMV(id)`: remove todas as pontuações associadas com `id`

- Conjunto de tuplos (identificador, pontuação)
- Duas operações de escrita
 - `ADD(id, pontuação)`: associa uma nova pontuação com `id`
 - `RMV(id)`: remove todas as pontuações associadas com `id`

- Conjunto de tuplos (identificador, pontuação)
- Duas operações de escrita
 - `ADD(id, pontuação)`: associa uma nova pontuação com `id`
 - `RMV(id)`: remove todas as pontuações associadas com `id`

{ } { }

{ Maria, 90 } { }

{ Maria, 90 } { Maria, 90 }

{ Maria, 90
João, 80 }

{ Maria, 90 }

$\left\{ \begin{array}{l} \text{Maria, 90} \\ \text{João, 80} \end{array} \right\}$ $\left\{ \begin{array}{l} \text{Maria, 90} \end{array} \right\}$

$\left\{ \begin{array}{l} \text{Maria, 90} \\ \text{João, 85} \\ \text{João, 80} \end{array} \right\}$ $\left\{ \begin{array}{l} \text{Maria, 90} \end{array} \right\}$

$\left\{ \begin{array}{l} \text{Maria, 90} \\ \text{João, 85} \\ \text{João, 80} \end{array} \right\}$ $\left\{ \begin{array}{l} \text{Maria, 90} \\ \\ \end{array} \right\}$

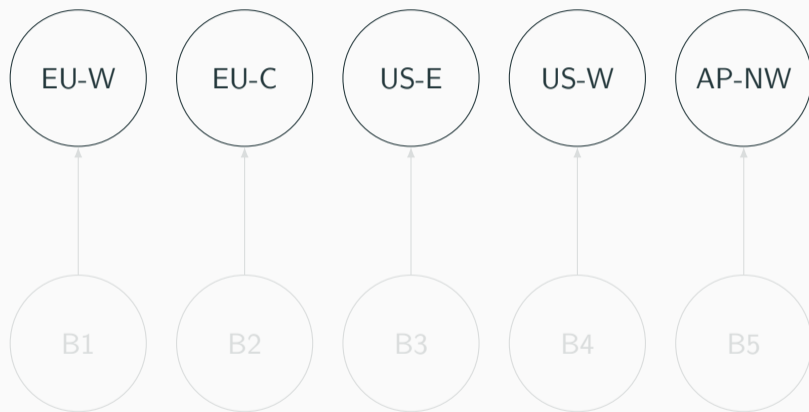
$\left\{ \begin{array}{l} \text{Maria, 90} \\ \text{João, 85} \end{array} \right\}$ $\left\{ \begin{array}{l} \text{Maria, 90} \end{array} \right\}$

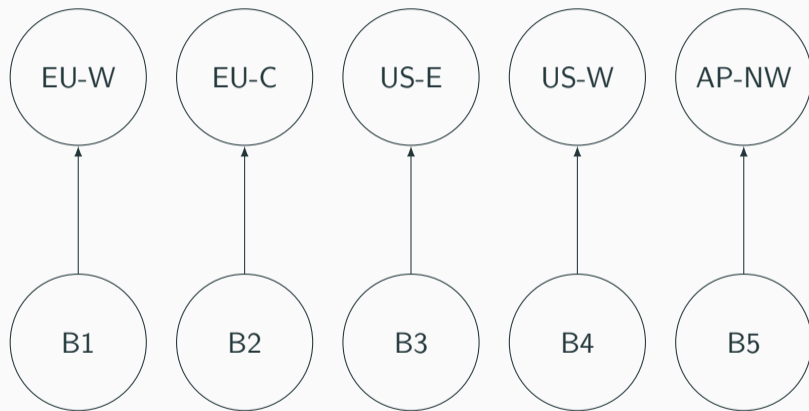
$\left\{ \begin{array}{l} \text{Maria, 90} \\ \text{João, 85} \end{array} \right\}$ $\left\{ \begin{array}{l} \text{Maria, 90} \end{array} \right\}$

$\left\{ \begin{array}{l} \text{Maria, 90} \\ \text{João, 85} \end{array} \right\}$ $\left\{ \begin{array}{l} \text{Maria, 90} \\ \end{array} \right\}$

{ João, 85 } { João, 85 }

- Replicação Não Uniforme
- CRDTs Não Uniformes
- **Avaliação**
- Conclusão e trabalho futuro





Top-K com remoções: custo de disseminação

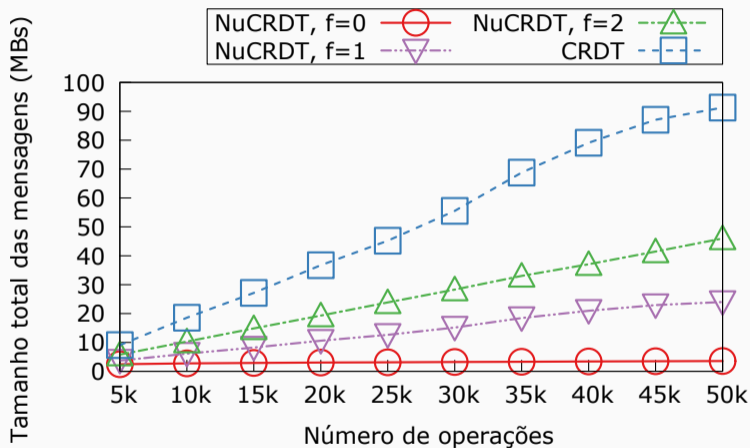


Figura 1: Tamanho total de mensagens

Top-K com remoções: tamanho médio das réplicas

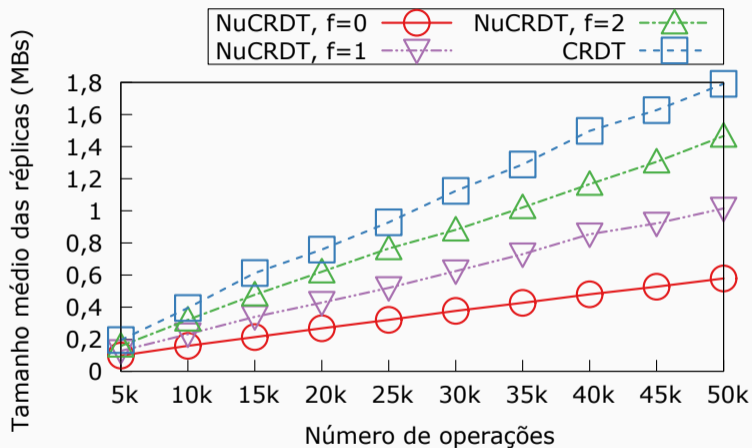


Figura 2: Tamanho médio das réplicas

Top-K com remoções: escalabilidade

Carga: 95% adições, 5% remoções

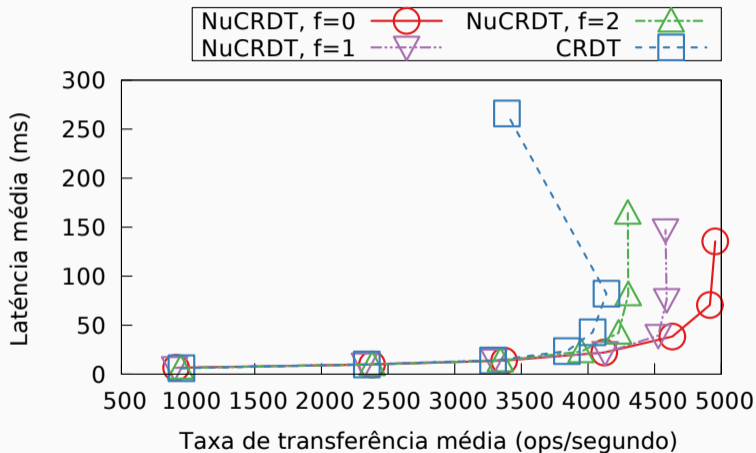


Figura 3: Escalabilidade

- Replicação Não Uniforme
- CRDTs Não Uniformes
- Avaliação
- **Conclusão e trabalho futuro**

- Introduzimos o modelo de replicação não uniforme e formalizamos a sua semântica para um sistema de consistência eventual
- Mostrámos como o modelo pode ser aplicado a CRDTs
- Implementámos um CRDT não uniforme numa base de dados geo-replicada, e avaliámos o seu desempenho

- Introduzimos o modelo de replicação não uniforme e formalizamos a sua semântica para um sistema de consistência eventual
- Mostrámos como o modelo pode ser aplicado a CRDTs
- Implementámos um CRDT não uniforme numa base de dados geo-replicada, e avaliámos o seu desempenho

- Introduzimos o modelo de replicação não uniforme e formalizamos a sua semântica para um sistema de consistência eventual
- Mostrámos como o modelo pode ser aplicado a CRDTs
- Implementámos um CRDT não uniforme numa base de dados geo-replicada, e avaliámos o seu desempenho

- Investigar a aplicabilidade deste modelo de replicação a outros modelos de consistência mais fortes
- Criar tipos de dados uteis para ambientes de Big Data e Machine Learning

- Investigar a aplicabilidade deste modelo de replicação a outros modelos de consistência mais fortes
- Criar tipos de dados uteis para ambientes de Big Data e Machine Learning

Questões?